Humboldt-Universität zu Berlin

Dissertation

Absolutes und nichtabsolutes Hören
Einflussfaktoren auf das Erinnern von Tonarten

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat)
Mathematisch-Naturwissenschaftliche Fakultät II

Dipl.-Psych. Kathrin B. Schlemmer
geb. am 18. März 1973 in Berlin

Dekan: Prof. Dr. Uwe Küchler

Gutachter: 1. Frau Prof. Dr. Elke van der Meer
2. Herr Prof. Dr. Wolfgang Auhagen
3. Herr Prof. Dr. Oliver Vitouch

eingereicht: 15. April 2005
Datum der Promotion: 6. Juli 2005

Zusammenfassung

In der vorliegenden Arbeit wurde mit einer Reihe von Experimenten geprüft, ob sich die Tonarterinnerung von Nichtabsoluthörern durch aus der Gedächtnisforschung abgeleitete Einflussfaktoren erklären lässt. Zunächst erfolgte eine theoretische Betrachtung des Tonartgedächtnisses sowohl aus musikpsychologischer als auch aus gedächtnispsychologischer Perspektive. Die Analyse von Befunden zum „latenten“ und „echten“ absoluten Gehör zeigte, dass eine Reihe von potenziellen Einflussfaktoren auf die Tonarterinnerung betrachtet werden muss, um herauszufinden, ob es sich bei diesen beiden Phänomenen um unterschiedliche Ausprägungen derselben Fähigkeit handelt.

Um den Einfluss von Faktoren der Melodien, der Melodie-Lernenden und der Art des Melodie-Lernens auf die Tonarterinnerung zu prüfen, wurden insgesamt 268 Probanden gebeten, vertraute Melodien aus dem Gedächtnis zu singen. Unabhängige Variablen waren die musikalische Expertise der Probanden, ihre Fähigkeit Töne zu benennen, die Form und die Intensität des dem Experiment vorangegangenen Melodie-Lernens sowie verschiedene Charakteristika der Melodien. Abhängige Variable war die Genauigkeit, mit der die Originaltonarten der Melodien produziert wurden. Es konnten Effekte der Hör-Häufigkeit, der musikalischen Expertise, der Tonbenennung, der Melodie-Eingängigkeit sowie ein Effekt motorischer Kontextinformationen auf die Genauigkeit der Tonarterinnerung nachgewiesen werden.

Um den Häufigkeitseffekt mit einer weiteren Anforderung zu untersuchen, wurde in einem weiteren Experiment die Tonbenennungsleistung von Absoluthörern und Nichtabsoluthörern verglichen. Dabei kam die Methode der Pupillometrie zum Einsatz, um Unterschiede in der mentalen Beanspruchung beim Benennen von Tönen unterschiedlicher Klangfarbe und Tonklasse nachweisen zu können. Die Ergebnisse stützen die Annahme, dass das häufige Hören bestimmter Töne sowohl bei Absoluthörern als auch bei Nichtabsoluthörern die Tonbenennung erleichtert. Dies verweist darauf, dass auch bei der musikspezifischen Aufgabe der Tonbenennung ein so grundlegendes Prinzip des menschlichen Gedächtnisses wie die Stabilisierung von Gedächtnisinhalten durch Wiederholung zum Tragen kommt.

Insgesamt weisen die Ergebnisse darauf hin, dass Tonarterinnerung ein komplexes Phänomen ist, für das eine alleinige Erklärung als „latentes absolutes Gehör“ zu kurz greift. Statt einer schwachen Ausprägung einer hochspezialisierten Fähigkeit scheint es sich eher um eine eigene Form des Erinnerns, die auf allgemeingültigen Gedächtnisprinzipien beruht, zu handeln.

Schlagworte: Melodiegedächtnis, absolutes Gehör, Tonartgedächtnis, Langzeitgedächtnis, Arbeitsgedächtnis

Abstract

In this thesis, memory for musical keys among absolute pitch nonpossessors, which is often referred to as “latent” absolute pitch, is examined. A theoretical analysis focused on existing research about “latent” and “manifest” absolute pitch. Evidence from music-psychological and general memory research as well as neuropsychological evidence was considered. The review of existing research revealed that several factors are potentially relevant for the memory of musical keys and should be considered in trying to determine whether “latent” and “manifest” absolute pitch can be described as different levels of the same ability on an “absolute pitch continuum”.

To examine whether characteristics of learned melodies, of melody-learners, and of melody-learning influence memory for musical keys among absolute pitch nonpossessors, 268 participants were asked in a series of experiments to sing familiar melodies from memory. Independent variables were the musical expertise of participants, their ability to label pitches, type and intensity of melody-learning, and characteristics of the learned melodies. The accuracy with which learned melodies could be produced in the original key was the dependent variable. Results revealed that frequency of melody-learning as well as participants’ musical expertise and ability to label pitches influence the accuracy of key production. Whether or not a melody is catchy as well as the existence of different types of motor imagery are further influencing factors for the accuracy of key production.

To examine the frequency-of-hearing effect in more detail, another experiment compared the pitch labeling performance of absolute pitch possessors and nonpossessors. Pupillary responses were measured in order to show differences in mental resource allocation when labeling pitches of different key colors or timbres. Results support the assumption that frequent exposure to pitches of certain key colors or timbres facilitate their labeling among both absolute pitch possessors and nonpossessors. This suggests that basic principles of human memory such as learning by frequency of exposure affect also very specific tasks such as pitch labeling.

Taken together, the results suggest that memory for musical keys is a complex phenomenon which can not adequately be described as being simply a “latent” or weak form of absolute pitch. Instead, memory for musical keys can be described as a “normal” memory mechanism, influenced by factors known to influence numerous other forms of human memory.

Keywords: Tune memory, absolute pitch, memory for musical keys, long term memory, working memory

für Mathilda

Danksagung

Die vorliegende Arbeit, insbesondere die Durchführung der Experimente, wäre ohne die Hilfe zahlreicher Personen nicht möglich gewesen, bei denen ich mich an dieser Stelle bedanken möchte.

Mein Dank gilt an erster Stelle meiner Betreuerin, Prof. Dr. Elke van der Meer. Bei ihr fand ich jederzeit ein offenes Ohr für alle theoretischen und praktischen Fragen, die im Laufe der Arbeit entstanden sind. Ich habe mich bei ihr und am Lehrstuhl für Kognitive Psychologie der Humboldt Universität wissenschaftlich und menschlich sehr gut aufgehoben gefühlt und möchte mich für die vielen fruchtbaren Diskussionen im Forschungskolloquium bedanken. Ebenso möchte ich auch Prof. Dr. Oliver Vitouch, Prof. Dr. Wolfgang Auhagen und Prof. Dr. Christian Kaernbach für zahlreiche Tipps und hilfreiche Diskussionen in allen Phasen dieser Arbeit sehr herzlich danken.

Für die Unterstützung bei der Datenerhebung möchte ich mich bei Franziska Kulke, Maria Deutschmann, Melanie Radalewski, Kathleen Nepp, Bettina Wagner, Gigliola Danko, Marie Melzer, Vivien Melcher, Charlotte Krug, Anja Horn, Elke Sellnau, Anne Reißmann, Dorothea Ullwer, Falk Nindel, Johanna Schattkowsky, Stefanie Cuff und Tina Gooren bedanken.

An meinen Experimenten zur Ton- und Tonarterinnerung haben zahlreiche Chorsänger, Musiker und Freunde ohne eine finanzielle Entschädigung teilgenommen. Für ihre Geduld und ihr Interesse möchte ich mich sehr herzlich bedanken. Ich danke auch Ralf Sochaczewsky, Frank Markowitsch, Ud Joffe und Herbert Hildebrandt für die Hilfe beim Rekrutieren von Versuchsteilnehmern in ihren Chören. Katrin Kobin möchte ich für ihren unermüdlichen Eifer bei dem Versuch, die Schüler des Droste-Hülshoff-Gymnasiums für die Versuchsteilnahme zu begeistern, danken. Ich danke auch Herrn Zajonc von der Anna-Freud Oberschule. Für die geduldige Beurteilung von über 100 Melodie-Ausschnitten bedanke ich mich sehr herzlich bei Prof. Hartmut Fladt, Mirjam Schlemmer und Clemens Schlemmer.

Für die Unterstützung beim Herstellen der Musikbeispiele möchte ich Ulrich Naudé, Ralf Sochaczewsky, Maike Bühle, Hans-Joachim Maempel und Peter Sonntag danken. Dem MPI für Bildungsforschung danke ich für die großzügige Überlassung ihres Clavinovas. Für ihre Beratung in Fragen der statistischen Auswertung danke ich Prof. Dr. Jürgen Bortz, Dr. Stefan Klein und besonders Lars Kuchinke. Für die unschätzbare Hilfe beim Programmieren der Experimente in Matlab danke ich Dirk Neumann.

Für die Förderung dieser Arbeit möchte ich dem evangelischen Studienwerk Villigst sehr herzlich danken. Außerdem möchte ich mich bei Sabine Schulz für ihre organisatorische Unterstützung und ihr offenes Ohr während meiner Arbeitspausen bedanken. Schließlich möchte ich mich bei meinen Eltern für die ideelle und tatkräftige Unterstützung während des Entstehens dieser Arbeit bedanken. Mein besonderer Dank gilt meinem Mann Clemens Schlemmer, der mir in allen Phasen der Arbeit und besonders im letzten Jahr ein guter Begleiter war, mir den Rücken freigehalten und mich darin bestärkt hat, dass dies der richtige Weg ist!

Inhaltsverzeichnis

Zusammenfassung ..... **II**

Abstract ..... **III**

Danksagung ..... **V**

Einleitung **1**

1 Vom Ton zur Melodie **2**

2 Charakterisierung des absoluten Gehörs **9**

2.1 Definitionen 9
2.2 Erfassungsmethoden und typische Befunde 10
2.2.1 Tonidentifikation 10
2.2.2 Tonproduktion 11
2.2.3 Gedächtnisaufgaben 12
2.2.4 Implizite Aufgaben 13
2.3 Neuropsychologische Befunde 15
2.3.1 Morphologische Besonderheiten bei Absoluthörern 16
2.3.2 Funktionelle Besonderheiten bei Absoluthörern 17
2.4 Was macht absolutes Hören aus und wie entsteht es? 19

3 Absolutes Gehör als Kontinuumsleistung **23**

3.1 Latentes absolutes Gehör 25
3.1.1 Wiedererkennung von Tonarten 26
3.1.2 Produktion von Tonarten 29
3.1.3 Produktion gesprochener Sprache 33
3.1.4 Das Tritonus-Paradox als Hinweis auf latentes absolutes Gehör 35
3.2 Absolutes Gehör als Kontinuumsleistung? 36

4 Musik in der Gedächtnisforschung **40**

4.1 Kategorisierung von Gedächtnissystemen 40
4.2 Deklaratives Gedächtnis: episodisches und semantisches Wissen 41
4.3 Gedächtnisprozesse beim deklarativen Erinnern 42
4.3.1 Enkodierung 43
4.3.2 Speicherung 43
4.3.3 Abruf 45
4.3.4 Vergessen 47
4.4 Formen nichtdeklarativen Erinnerns 48
4.4.1 Priming 49
4.4.2 Wahrnehmungslernen 49
4.4.3 Emotionales Lernen 51
4.4.4 Fertigkeiten und Gewohnheiten 51
4.4.5 Motorische Gedächtnisinhalte 52

5 Fragestellung & Hypothesen der Untersuchungen 57
5.1 Hypothesen zur Tonarterinnerung von Nichtabsoluthörern 58
5.2 Hypothesen zur Tonerinnerung von Absolut- und Nichtabsoluthörern 59
5.3 Überblick über die Experimente 60

6 Verbreitung von Tonartrepräsentationen 62
6.1 Methodik der Produktions-Experimente 62
6.2 Ausgangspunkt Levitin (1994): Erinnerung an persönliche Lieblingslieder 65
6.2.1 Stichprobe 65
6.2.2 Produzierte Melodien 65
6.2.3 Besonderheiten bei der Versuchsdurchführung 66
6.2.4 Besonderheiten bei der Datenauswertung 66
6.2.5 Ergebnis und Diskussion 67
6.3 Kurzzeitig gelernte Melodien: Schüler 68
6.3.1 Stichprobe 68
6.3.2 Produzierte Melodien 68
6.3.3 Besonderheiten bei der Datenauswertung 69
6.3.4 Ergebnis und Diskussion 70
6.4 Kurzzeitig gelernte Melodien: Chorsänger 71
6.4.1 Stichprobe 71
6.4.2 Produzierte Melodien 72
6.4.3 Besonderheiten bei der Versuchsdurchführung 72
6.4.4 Besonderheiten bei der Datenauswertung 72
6.4.5 Ergebnis und Diskussion 73
6.5 Erinnerung an über längere Zeit gelernte Melodien 74
6.5.1 Stichprobe 74
6.5.2 Produzierte Melodien 74
6.5.3 Besonderheiten bei der Datenauswertung 75
6.5.4 Ergebnis und Diskussion 75
6.6 Erinnerung an einen Werbejingle 76
6.6.1 Stichprobe 76
6.6.2 Produzierte Melodie 77
6.6.3 Besonderheiten bei der Datenauswertung 77
6.6.4 Ergebnis und Diskussion 77
6.7 Erinnerung an die eigene Handymelodie 78
6.7.1 Stichprobe 78
6.7.2 Produzierte Melodien 78
6.7.3 Besonderheiten bei der Datenauswertung 79
6.7.4 Ergebnis und Diskussion 79
6.8 Erinnerung an ein Instrumental-Konzert 80
6.8.1 Stichprobe 80
6.8.2 Produzierte Melodien 81
6.8.3 Besonderheiten bei der Datenauswertung 81
6.8.4 Ergebnis und Diskussion 82
6.9 Zusammenfassende Diskussion der Produktions-Ergebnisse 83
6.10 Methodik der zusammenfassenden Analyse 85
6.10.1 Charakterisierung der Versuchspersonen 85
6.10.2 Charakterisierung der Lern-Faktoren 87
6.10.3 Charakterisierung der Melodien 87

6.11 Ergebnisse der zusammenfassenden Analyse 90
6.11.1 Expertise- und Tonbenennungseffekte 93
6.11.2 Einfluss der Lernfaktoren 94
6.11.3 Einfluss der Melodie-Merkmale 95
6.11.4 Interaktionen zwischen den Einflussfaktoren 97
6.12 Diskussion der zusammenfassenden Analyse 98

7 Zur Vertrautheit beim intentionalen Melodielernen 101
7.1 Methode 101
7.1.1 Charakterisierung der Stichprobe 101
7.1.2 Gelernte Melodien 102
7.1.3 Lernphase 102
7.1.4 Abrufphase: Experiment 103
7.1.5 Herstellung des Versuchsmaterials 104
7.1.6 Besonderheiten bei der Versuchsdurchführung 105
7.1.7 Datenanalyse 105
7.2 Ergebnisse 106
7.2.1 Produktionsaufgabe 106
7.2.2 Wiedererkennungsaufgabe 107
7.2.3 Tonarterinnerung, Tonbenennung & musikalische Expertise 111
7.3 Diskussion der Ergebnisse 113

8 Zur Vertrautheit bei der Tonbenennung 116
8.1 Methode 120
8.1.1 Charakterisierung der Stichprobe 120
8.1.2 Reizmaterial 120
8.1.3 Ablauf des Experiments 121
8.1.4 Geräte 123
8.1.5 Analyse der Pupillenrohwerte 124
8.2 Ergebnisse 125
8.2.1 Klassifikation der Vpn anhand der Verhaltensdaten 125
8.2.2 Hypothesenbezogene Auswertung der Verhaltensdaten 127
8.2.3 Hypothesenbezogene Auswertung der Pupillendaten 128
8.3 Diskussion 131

9 Motorische Gedächtnisinhalte bei der Tonarterinnerung 136
9.1 Abruf von Instrumentalmusik 137
9.1.1 Methode 137
9.1.2 Ergebnisse 139
9.1.3 Diskussion 146
9.2 Produktion mit Interferenzaufgaben bei Chorsängern und Laien 149
9.2.1 Methode 149
9.2.2 Spezifische Hypothesen für die beiden Stichproben 152
9.2.3 Ergebnisse 153
9.2.4 Diskussion 161
9.3 Zusammenfassende Diskussion der Ergebnisse beider Experimente 163

10 Zusammenfassung und Diskussion der Hauptergebnisse 166

Literatur 173

Einleitung

In der vorliegenden Arbeit geht es um einen besonderen Aspekt des Musikgedächtnisses: die Erinnerung an absolute Tonhöhen durch nicht absolut hörende Personen. Von vielen aktiv Musizierenden wird das Vorhandensein eines impliziten Tonart-Gedächtnisses schon seit längerer Zeit vermutet, und in jüngerer Zeit wurde es experimentell auch an Nichtmusikern beobachtet. Ein Beispiel für das Vorhandensein eines impliziten Tonart-Gedächtnisses ist die Antizipation des Anfangs vertrauter Melodien oder Tonfolgen, wie z.B. der Begrüssungs"melodie" kurz vor dem morgendlichen Einschalten des eigenen Computers. Achtet man auf die Übereinstimmung der Anfangstöne solcher Melodie-Antizipationen und der kurz darauf erklingenden Originalfassung, so lässt sich häufig eine verblüffende Übereinstimmung feststellen.

Vor rund 10 Jahren erbrachte Levitin (1994) mit einem seitdem oft zitierten Experiment den wissenschaftlichen Nachweis, dass sich auch der musikalische Laie an Tonarten hoch vertrauter Melodien erinnern kann. Obgleich es einige Folgestudien mit meist ähnlichen Ergebnissen gab, fand bislang kaum eine Suche nach Erklärungsmöglichkeiten für dieses Phänomen statt. Beispielsweise wurde es bald als „latentes absolutes Gehör" bezeichnet, ohne dass hinterfragt wurde, inwieweit abgesehen von der Tatsache, dass absolute Tonhöhen abgerufen werden können, Ähnlichkeiten mit absolutem Hören vorliegen.

Die vorliegende Arbeit widmet sich sowohl der Frage nach der Verbreitung von Tonartrepräsentationen in bisher nicht untersuchten Populationen beim Abruf verschiedener Arten von Melodien als auch der Suche nach Erklärungsfaktoren für die sowohl intra- als auch interindividuell beobachtete unterschiedliche Genauigkeit der Tonarterinnerung.

Bei der Theoriebildung stößt man zunächst aufgrund der Bezeichnung „latentes absolutes Gehör" auf die Literatur zum absoluten Gehör. Als die Fähigkeit, einzelne Töne ohne Kontext kategorisieren zu können, ist dieses Phänomen zwar von der – häufig impliziten – Erinnerung musikalischer Laien an Tonarten „echter" Musik zu unterscheiden. Trotzdem ist es auch für Gedächtnisforscher ein interessantes Phänomen, weil eine isolierte Fähigkeit einiger weniger Menschen vorliegt, die womöglich in abgeschwächter Form weiter verbreitet ist als lange vermutet wurde. Interessant und der Betrachtung würdig ist jedoch auch die Untersuchung des Tonart-Gedächtnisses aus der Perspektive der Gedächtnisforschung. Von der allgemeinen gedächtnispsychologischen Literatur wurde die Erinnerung an Tonarten, wie auch allgemeiner die Erinnerung an Musik wesentlich seltener behandelt als z.B. die Erinnerung an visuelle oder sprachliche Inhalte. Daher wird in dieser Arbeit versucht, die Übertragbarkeit gedächtnispsychologischer Befunde auf musikalische Gedächtnisinhalte zu prüfen und Hypothesen für die eigenen Untersuchungen zu generieren.

Die theoretische Annäherung an das Phänomen der Tonarterinnerung erfolgt also von zwei Seiten aus: aus der Perspektive der Forschung zum absoluten Gehör und aus der Perspektive der Gedächtnispsychologie. Dabei soll nach Schnittstellen gesucht werden, die eine Formulierung von Hypothesen erlauben, die mit beiden Perspektiven kompatibel sind.

1 Vom Ton zur Melodie

Zur Einordnung des in der vorliegenden Arbeit thematisierten Phänomens, der Erinnerung an die Tonarten gelernter Melodien, soll zunächst definiert werden, was musikalische Töne ausmacht, wie sie wahrgenommen und verarbeitet werden und wie aus Tönen Melodien entstehen.

Prinzipiell lässt sich Schall physikalisch in zwei Kategorien unterteilen, nämlich Töne und Geräusche. Geräusche sind dadurch gekennzeichnet, dass keine geordneten Schwingungen, sondern ungeordnete Luftbewegungen vorliegen. Aus Geräuschen ist demnach auch keine einzelne Tonhöhe heraushörbar. Töne dagegen bestehen im einfachsten Fall aus einer einzigen Sinusschwingung, die sich durch ihre Frequenz (Anzahl der Schwingungen pro Sekunde) und die Amplitude (Größe des Ausschlags) kennzeichnen lässt. Musikalische Töne setzen sich aus mehreren Sinusschwingungen zusammen, wobei die Tonhöhe sich im Wesentlichen durch die Grundfrequenz definiert, während die Obertöne beim Hörer zum Erleben einer bestimmten Klangfarbe beitragen. In der vorliegenden Arbeit geht es nur um musikalische Töne, die im folgenden vereinfacht als Töne bezeichnet werden. Musikalische Töne haben eine bestimmte Eigenschaft, die als Oktaväquivalenz bezeichnet wird. Das bedeutet, dass Töne im Oktavabstand (Frequenzverhältnis 2:1) als sehr ähnlich wahrgenommen und von musikalischen Laien sogar verwechselt werden. Da die Oktaväquivalenz in sehr vielen musikalischen Kulturen auftritt (Nettl, 1956), scheint es sich hierbei um eine perzeptuelle Universalie zu handeln. Im Tonsystem wird die Oktaväquivalenz dadurch sichtbar gemacht, dass Töne im Oktavabstand den gleichen Namen teilen (z.B. *c*', *c*'', etc.). Aufgrund der Oktaväquivalenz wurde vorgeschlagen, dass Tonhöhen zweidimensional dargestellt werden können: Die Dimension der Tonklasse („tone chroma", auch: Tonigkeit) definiert die Position des Tons innerhalb einer Oktave, während die Dimension der Tonhöhe („pitch height", auch: Helligkeit) die genaue Lage im Tonraum angibt (Révész, 1913, 1946; Shepard, 1964; Zusammenfassung bei Deutsch, 1999). Die Zweidimensionalität von Tönen wurde bereits von Drobisch (1855) in Form einer Helix dargestellt, bei der Töne im Oktavabstand jeweils auf der gleichen Position im Tonklassen-Kreis liegen (vgl. Abb. 1.1). Die Unterscheidung von Tonklasse und Tonhöhe spielt bei der Definition des absoluten Gehörs eine Rolle und wird in Kapitel 2 wieder aufgegriffen.

Töne werden im menschlichen Innenohr frequenzspezifisch verarbeitet, d.h. je nach der Grundfrequenz eintreffender Töne werden die Haarzellen an einem bestimmten Ort auf der Basilarmembran maximal ausgelenkt, der bei hohen Frequenzen näher an der Basis und bei tiefen Frequenzen näher am Apex der Cochlea liegt (vgl. Spitzer, 2002, S. 58 ff). Die frequenzspezifische Verarbeitung der Töne setzt sich über das gesamte Hörsystem fort, und für fast alle Zentren der Hörbahn wurden sogenannte Frequenzkarten nachgewiesen (King, Schnupp & Doubell, 2001). Experimente an Tieren und Menschen haben ergeben, dass Töne im primären auditorischen Kortex in Form einer Tonlandkarte repräsentiert sind (vgl. Weinberger, 1999).

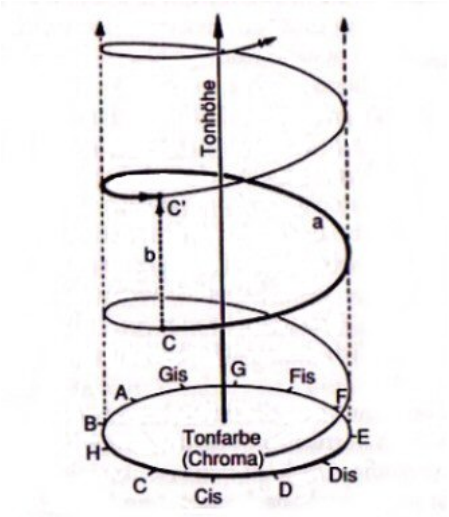

Abb. 1.1: Repräsentation von Tönen als Helix, wobei die Tonklasse durch den horizontalen Tonklassen-Kreis und die Tonhöhe durch die vertikale Dimension dargestellt wird (aus Vitouch, 2005, S. 726).

Die frequenzspezifische Verarbeitung und Projektion von Tönen bis in den primären auditorischen Kortex legen die Erwartung nahe, dass die Frequenzinformation gehörter Töne mindestens bei Aufgaben für das Kurzzeitgedächtnis (KZG, z.B. Vergleich von Tönen) und möglicherweise auch beim Abruf aus dem Langzeitgedächtnis (LZG) zur Verfügung stehe. Dabei muss jedoch berücksichtigt werden, dass beim Hören von Musik einzelne Töne für sich genommen kaum eine musikalische Bedeutung haben. Bereits die Wahrnehmung einer einstimmigen Melodie, gewissermaßen das Grundgerüst der westlichen Musik, setzt das Inbeziehungsetzen aktuell gehörter Töne mit vergangenen und folgenden Tönen, also das Hören im Kontext, voraus. Aufgrund dieser Tatsache kann man bei der Untersuchung von Ton-Repräsentationen zwischen *absoluter* und *relationaler* Information unterscheiden. Betrachtet man z.B. die zwei Töne *a*' und *c*'', so ist die absolute Information durch die beiden Grundfrequenzen definiert, die in der Normstimmung mit 440 Hz und 523 Hz angegeben werden. Mit relationaler Information (die oft auch als relativ bezeichnet wird) ist dagegen das Verhältnis der beiden Töne zueinander gemeint, das hier eine kleine Terz ist. Die absoluten Informationen sind spezifisch für bestimmte Tonklassen, während Tonrelationen, sogenannte Intervalle, sich im Tonraum nach oben und unten verschieben (transponieren) lassen, ohne ihre Identität zu verlieren. Die kleinen Terzen zwischen *fis*' und *a*', sowie zwischen *f*' und *as*' oder *cis*'' und *e*'' sind hinsichtlich des Abstands der Töne also identisch, zumindest in der heute gebräuchlichen gleichtemperierten Stimmung.

Klammert man nun bei der Definition einer Melodie Aspekte der Tondauer (Rhythmus, Tempo) sowie klangliche Aspekte (Lautstärke, Klangfarben) vorerst aus und reduziert die Definition auf Aspekte der Tonhöhe, so folgt aus den Ausführungen des vorigen Abschnitts, dass sich eine Melodie sowohl als Folge absoluter Tonhöhen (aber bezogen auf einen Stimmton wie z.B. *a* = 440 Hz) als auch als Serie von Tonrelationen definieren lässt. Anschaulich wird dies in der folgenden Abbildung am Beispiel der Melodie „Freude schöner Götterfunken" dargestellt, wobei der jeweils vorrangige Aspekt der Tonfolge (relational vs. absolut) rot hervorgehoben ist.

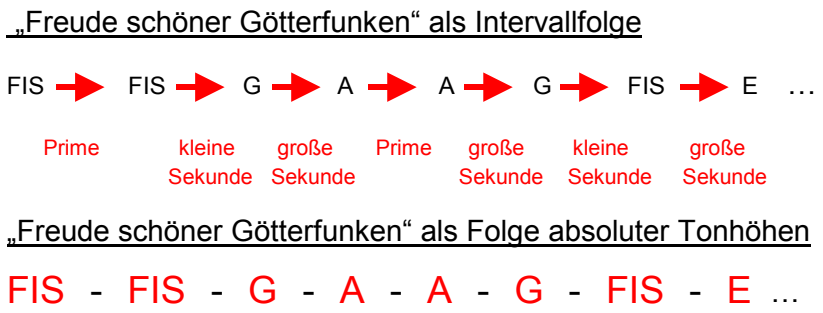

Abb. 1.2: Schematische Darstellung des Anfangs von „Freude schöner Götterfunken“ als Folge von Tonrelationen (oben) und als Folge von (durch ihre Relation zu einem Stimmton definierten) absoluten Tonhöhen (unten).

Nachfolgend sollen einige Befunde zur Wahrnehmung und kurzzeitigen Erinnerung von Tönen und Melodien referiert werden, die darauf hindeuten, dass der kurzzeitigen Erinnerung sowohl absolute als auch relationale Toninformationen zur Verfügung stehen.

In einer Serie von Experimenten konnte Deutsch (Zusammenfassung in Deutsch, 1999) zeigen, dass auch in einer Umgebung von zufälligen Tonfolgen eine Kontextabhängigkeit der Einzelton-Wahrnehmung und -Erinnerung vorliegt. Die Zielstellung von Deutschs Experimenten war der Nachweis eines separaten, also vom verbalen Gedächtnis unterscheidbaren Tonhöhengedächtnisses. Daher bediente sie sich in ihren Experimenten der Methodik der Interferenz- oder Doppelaufgabe, die in der Gedächtnispsychologie intensiv zur Erforschung des Arbeitsgedächtnisses (Baddeley & Hitch, 1976) eingesetzt wurde. Der Begriff des Arbeitsgedächtnisses bezeichnet ein System von Komponenten, die Informationen aufrecht erhalten und manipulieren können. Dabei gelangt Information zunächst in die zentrale Exekutive (eine Art Kontrollinstanz), wo sie enkodiert und in eines der beiden Subsysteme überführt wird. Dort wird die Information modalitätsspezifisch im räumlich-visuellen Notizblock und in der phonologischen Schleife verarbeitet.

Da im Modell des Arbeitsgedächtnisses angenommen wird, dass sich Aufgaben, die verschiedene Komponenten beanspruchen, weniger stören als Aufgaben, die die gleiche Komponente belasten, kann man seine Gültigkeit mit Hilfe von Doppelaufgaben prüfen. Dabei wird eine Primäraufgabe bearbeitet (z.B. das Memorieren einer Reihe von Silben), während gleichzeitig eine Sekundäraufgabe vorliegt, z.B. das Hören irrelevanter (= nicht zu beachtender) Töne oder das leise Sprechen irrelevanter Silben (z.B. „ta ta ta“). Sinkt die Leistung in der Primäraufgabe bei Vorliegen der Sekundäraufgabe im Vergleich zu einer ungestörten Bedingung, so wird dies als Hinweis dafür angesehen, dass Primär- und Sekundäraufgabe das gleiche Gedächtnissystem beanspruchen. Dieser Beweis liegt allerdings nur dann vor, wenn die Interferenz spezifisch ist, d.h., wenn sich andere Sekundäraufgaben (von denen man annimmt, dass sie nicht das gleiche Gedächtnissystem beanspruchen) nachweisen lassen, die die Leistung in der Primäraufgabe nicht beeinträchtigen. Kann man dies nicht, so ist statt von spezifischer Interferenz eher von einer zentralen Ressourcenüberlastung auszugehen.

Mithilfe von spezifischer Interferenz ließ sich die von Baddeley und Hitch (1976) postulierte Unterteilung des Arbeitsgedächtnisses in eine phonologische und eine räumlich-visuelle Speicherkomponente stützen (vgl. Baddeley, 1986). Innerhalb der phonologischen Speicherkomponente lassen sich wiederum eine passive Speicher- und eine aktive Rehearsalkomponente unterscheiden, die in Kapitel 4.4.5 im Kontext der Diskussion motorischer Gedächtnisinhalte aufgegriffen werden. Welche Subsysteme des Arbeitsgedächtnisses an der Verarbeitung von Musik beteiligt sein könnten, wird bereits seit einiger Zeit kontrovers diskutiert (vgl. z.B. Berz, 1995; Carroll-Phelan & Hampson, 1996). Während es einerseits Hinweise auf eine gemeinsame Verarbeitung von Musik und Sprache im gleichen Gedächtnissystem gibt (z.B. Salamé & Baddeley, 1989), deuten die im Folgenden berichteten Ergebnisse von Deutsch auf eine Verarbeitung in getrennten Subsystemen hin.

In einem ersten Experiment spielte Deutsch (1970) ihren musikalisch untrainierten Versuchspersonen (Vpn) einen Ton vor, dem ein Retentionsintervall von 5 Sekunden und dann ein zweiter Ton folgte. Die Vpn sollten beurteilen, ob die beiden Töne identisch oder unterschiedlich waren (in 50 % der Fälle waren sie identisch, in 50 % der Fälle unterschieden sie sich um einen Halbton). In einer ungestörten Bedingung waren die Vpn dazu uneingeschränkt in der Lage (100 % richtige Antworten), während die Leistung in einer zweiten Bedingung, in der während des Retentionsintervalls 6 zu ignorierende Töne erklangen, auf 67,7 % richtige Antworten absank. Erklangen dagegen 6 Zahlen im Retentionsintervall, so wurde die Leistung nicht entscheidend beeinträchtigt, weder bei zu ignorierenden Zahlen (97,6 % richtige Antworten), noch bei zu erinnernden Zahlen (94,4 % richtige Antworten).

Eine Reihe von weiteren Untersuchungen mit gleicher Methodik von Deutsch und anderen Forschern hat gezeigt, dass Töne im Retentionsintervall das Einzeltongedächtnis am meisten stören, wenn sie in der gleichen Oktave wie der erste Ton erklingen (Deutsch, 1974), unabhängig davon, ob sie in einer ähnlichen oder unähnlichen Klangfarbe (Semal & Demany, 1991) oder Lautstärke (Semal & Demany, 1993) dargeboten werden. Deutsch interpretiert diese Datenlage als Hinweis auf ein separates Tongedächtnis, wobei die Erinnerungsleistung weder auf verbaler Kodierung (dagegen spricht die fehlende Interferenz durch Sprache) noch auf einer rein klanglichen Kodierung (dagegen spricht die Unabhängigkeit der Interferenz von Klangfarbe und Lautstärke) basieren kann. Statt dessen scheint die Tonhöhe im Sinne der Grundfrequenz (also inklusive Oktavlage) aus dem dargebotenen Klang abstrahiert gespeichert zu werden. Dieser Prozess scheint zudem von der musikalischen Expertise abhängig zu sein, wie Pechmann & Mohr (1992) zeigen konnten. In einem Experiment mit sehr ähnlicher Methodik hing die Interferenzanfälligkeit der Vpn mit ihrer musikalischen Expertise zusammen, denn während sich bei Musikern die Leistungen nur durch interpolierte Töne verschlechterte, geschah dies in geringerem Maße bei musikalischen Laien auch durch interpolierte Sprache und visuelles Material. Interessante Ergebnisse erbrachten weitere Experimente, in denen die Beziehung zwischen den beiden Testtönen und den sechs interpolierten Tönen variiert wurde (Deutsch, 1978; Olsen & Hansen, 1977; Deutsch & Boulanger, 1984). Es zeigte sich, dass Zufallstonfolgen mit einem großen Tonumfang stärker störten als geordnete Tonfolgen (Tonleitern) mit kleinem Tonumfang. Diese Ergebnisse ver-

weisen darauf, dass Tonfolgen, die gewissermaßen zu einer Melodie integriert werden können und als solche im Tonraum von den Testtönen unterscheidbar sind, leichter ignoriert werden können als solche Tonfolgen, die mangels erkennbarer Zusammenhänge nur als Serie von Einzeltönen erinnert werden können.

Aus den berichteten Ergebnissen wird deutlich, dass einzelne Tonhöhen besser unter bestimmten (experimentellen) Bedingungen kurzfristig im Gedächtnis behalten werden können, nämlich bei weitgehender Absenz anderen tonalen Materials. Unter solchen Bedingungen haben auch Rakowski & Morawska-Büngeler (1987) die grundsätzliche Erinnerungsfähigkeit an Einzeltöne über Zeitintervalle im Minutenbereich nachweisen können. Unter im musikalischen Sinne normalen Bedingungen dagegen, also beim Vorhandensein mehrerer aufeinander bezogener Töne, scheint das Gedächtnis für einzelne Töne stark kontextabhängig zu sein. „Thus, we must conclude that even when there is no obvious melody, individual tones are not heard independently of a tonal context. Subjects' implicit knowledge of the conventional tonal relationships in Western music supplies this framework, or schema", kommentiert Crowder (1993, S. 126) die vorliegende Befundlage.

Zusammenfassend ist festzuhalten, dass die Beziehungen zwischen gehörten Tönen, also relationale Informationen, im KZG eine wichtige Rolle spielen. Trotz Interferenz blieb jedoch in allen berichteten Experimenten eine überzufällige Antwortleistung beim Tonvergleich erhalten. Man kann also von einer kurzfristigen Repräsentation auch der absoluten Tonhöhen ausgehen. Dass die Repräsentation einzelner absoluter Tonhöhen (ohne Interferenz) mit zunehmender Zeit schwindet, wird aus den Daten zahlreicher Experimente, die hier nicht näher ausgeführt werden sollen, deutlich (Bachem, 1954; Harris, 1952; Koester, 1945; Rakowski, 1994; Rakowski & Morawska-Büngeler, 1987; Wickelgren, 1966, 1969).

Welche Art von relationalen Informationen im KZG repräsentiert werden, zeigt sich in Untersuchungen mit unterschiedlichen Methoden in etwas unterschiedlicher Weise. Obgleich die Untersuchung des kurzfristigen Melodiegedächtnisses mit *Produktionsmethoden* aufgrund stark variierender Fähigkeiten der Vpn zu singen, eine Klaviertastatur zu bedienen oder vorgestellte Melodien gar in Notenschrift wiederzugeben, sehr schwierig ist, stellten Sloboda & Parker (1985) ihren Vpn eine solche Produktionsaufgabe. Dabei sollten vorgespielte (den Vpn unbekannte) russische Volkslieder mehrmals nachgesungen werden. Es zeigte sich, dass die unbekannten Melodien nur recht ungenau wiedergegeben werden konnten. Offensichtlich war die Gedächtniskapazität mit der Wiedergabe der genauen Melodie überfordert, so dass einige Melodieaspekte (z.B. das Metrum) genauer wiedergegeben werden konnten als andere (z.B. der genaue Rhythmus und die Intervalle). *Wiedererkennungsexperimente* mit neu gelernten Melodien beschreiben Dowling & Harwood (1986). Kennzeichen dieser Methode ist, dass Vpn zunächst eine unbekannte Kurzmelodie vorgespielt wird und dann mehrere leicht abgeänderte Versionen dieser Melodie (Transposition, Transposition mit chromatisch veränderten Intervallen, diatonische Intervallveränderungen, konturverletzende Intervallveränderungen). Bei jeder abgeänderten Version soll beurteilt werden, ob sie mit der ersten Melodie übereinstimme oder nicht. Dowling und seine Mitarbeiter fanden heraus, dass

kurz nach dem ersten Hören einer Melodie eine exakte Transposition leicht mit einer Version mit veränderten Intervallen verwechselt wird, wohingegen Konturveränderungen (Veränderungen der Intervallrichtung) bemerkt werden (Dowling & Fujitani, 1971; Dowling, 1978; Bartlett & Dowling, 1980). Wird die Darbietung einer neuen Melodie dagegen von einem Retentionsintervall von einigen Sekunden (gefüllt mit anderen Tönen) gefolgt, so verschlechtert sich zwar die Wiedererkennungsleistung insgesamt, aber relativ betrachtet verbessert sich die Erinnerung an exakte Intervalle (Dowling & Bartlett, 1981). Zusammengefasst deuten diese Experimente darauf hin, dass beim kurzzeitigen Erinnern an unbekannte Musik die Kontur einer Melodie eine herausragende Rolle spielt und dass der Abruf der genauen Intervalle von der Anforderung abhängig ist. Leider werden in sämtlichen berichteten Experimenten die intervallveränderten Melodieversionen nur mit der exakten Transposition verglichen, nicht aber die Transposition mit dem Original (keine triviale Aufgabe, wenn zwischen den beiden Vergleichsmelodien mehrere andere Versionen in unterschiedlichen Tonarten vorgespielt werden), so dass man keine Schlussfolgerungen für die kurzfristige absolute Ton-Repräsentation neu gelernter Kurzmelodien ziehen kann.

Implizite Aufgaben für die kurzfristige Erinnerung an neue Melodien deuten anders als die berichteten expliziten Methoden auf eine sehr genaue Repräsentation bereits nach einmaligem Hören hin. So fanden Crowder und Mitarbeiter (Serafine, Davidson, Crowder & Repp, 1986, Crowder, Serafine & Repp, 1990, Zusammenfassung in Crowder, 1993) nach dem einmaligen Vorspielen von Melodien mit nonsense-Texten, dass Wörter besser wiedererkannt werden, wenn sie mit der gleichen Melodie kombiniert gehört werden als in Kombination mit einer anderen (aber auch bereits gehörten) Melodie („original“ vs. „mismatch“). Die explizite Erinnerung an die Wörter und Melodien war nur zufällig. Diese Art von Test sehen die Autoren als ökologisch valider an als explizite Testmethoden, weil es dem Alltag der meisten Menschen entsprechen dürfte, Melodien nur einmal zu hören und spätere Wiederholungen, z.B. die Wiederholung eines Leitmotivs, implizit zu erinnern. Auf den Beitrag impliziter Methoden zur Messung absoluter Ton-Repräsentationen wird in den folgenden Kapiteln nochmals eingegangen.

Für die Frage nach der langfristigen Repräsentation oft gehörter Melodien sind Befunde von Deutsch (1972) mit Melodien, bei denen die Oktavlagen der Töne verändert wurden, aufschlussreich. Bei solchen Melodien bleiben die originalen Tonklassen erhalten, aber die Intervalle werden dadurch verändert, dass z.B. der zweite Ton von „Hänschen klein“ nicht eine kleine Terz tiefer als der erste Ton liegt, sondern beispielsweise eine Oktave + kleine Terz (= kleine Dezime) tiefer oder eine Oktave – kleine Terz (= große Sexte) höher. Deutsch (1972) konnte zeigen, dass die Wiedererkennung gut bekannter Melodien unter solchen Bedingungen sehr schwierig ist. Dowling & Hollombe (1977) sowie Idson & Massaro (1978) konnten nachweisen, dass bei Beibehaltung der Kontur (also der Intervallrichtungen „aufwärts“ und „abwärts“) eine Wiedererkennung eher möglich ist als bei Verletzung der originalen Kontur. Dies deutet darauf hin, dass die langfristige Repräsentation der Töne einer Melodie nicht aus einer Serie von Tonklassen besteht, sondern dass die genauen Intervalle, ebenso wie die Kontur der Melodie, repräsentiert sind. Einen Hinweis auf die Repräsentation genauer Inter-

valle bei vertrauten Melodien liefern auch Befunde von Attneave & Olson (1971), die bei musikalischen Laien die Fähigkeit, einen Nachrichten-Jingle von verschiedenen Starttönen aus korrekt zu singen, nachwiesen.

Welche Bedeutung aber hat die Tonart, in der eine Melodie gelernt wurde, also die Folge absoluter Tonhöhen, für die langfristige Repräsentation der Melodie? Der Gestaltpsychologe von Ehrenfels (1890) betonte, dass eine transponierte Melodie ihre Identität im Sinne einer Gestaltqualität behalte, solange alle Ton-Relationen unverändert bleiben. Melodien seien in diesem Sinne visuellen Formen ähnlich, die ihre Identität ebenfalls beibehalten, wenn sie gedreht oder im Wahrnehmungsfeld verschoben werden (vgl. Deese & Grindley, 1947). Für die Erhaltung der Identität transponierter Melodien gibt es Hinweise aus der musikalischen Praxis, vor allem aus dem Bereich der Vokalmusik. So ist es zum Einen bei Volksliedern gängige Praxis, ein Lied ohne Zuhilfenahme eines Tongebers auf einer bequemen Tonhöhe anzustimmen. Dies führt dazu, dass z.B. „Hänschen klein“ in vielen verschiedenen Tonarten gesungen wird, ohne dass dies der Wiedererkennung der Melodie abträglich ist. Zum Anderen ist es auch bei Kunstliedern üblich, dass sie in verschiedenen Stimmlagen singbar sind. Dies wird an verschiedenen Ausgaben für „hohe“, „mittlere“ und „tiefe“ Stimmen ebenso deutlich wie an der von professionellen Liedbegleitern intensiv geübten Fähigkeit, „vom Blatt“ zu transponieren, also eine andere Tonart zu spielen als notiert, um den Bedürfnissen des jeweiligen Liedsängers gerecht zu werden. So lange bei Transpositionen keine extremen Lagen jenseits der menschlichen Tonunterscheidungsfähigkeit verwandt werden, werden transponierte Melodien also problemlos wiedererkannt. Diese Tatsache schlägt sich in der Theorie für musikalische Gedächtnisinhalte von Carroll-Phelan & Hampson (1996) insofern nieder, als sie für Melodien eine hinsichtlich Tonhöhe, Tempo, Klangfarbe und Lautstärke flexible Repräsentation postulieren. Statt einer Reihe absoluter Tonhöhen sei im Gedächtnis ein „pitch string“ im Sinne von Tonrelationen gespeichert, der mit einem „rhythm plan“ assoziiert sei.

Die Frage, die im Zentrum der vorliegenden Arbeit steht, ist nun, inwieweit im LZG neben den zweifellos sowohl bei Musikern als auch bei musikalischen Laien vorhandenen relationalen Tonhöhen auch absolute Tonhöhen gelernter Melodien repräsentiert sind. Da die Definition des absoluten Gehörs (vgl. Kap. 2) im wesentlichen auf ein „LZG für absolute Tonhöhen“ hinausläuft, impliziert sie zugleich, dass die langfristige Absolutrepräsentation von Tonhöhen der nicht absolut hörenden Allgemeinbevölkerung nicht möglich sei. Trotzdem stellt sich die Frage, ob es bestimmte Umstände gibt, die in der Person (z.B. musikalische Expertise), in den Melodien (z.B. Eingängigkeit) oder in der Art des Lernens (z.B. Lernhäufigkeit) liegen und die eine Ausbildung absoluter Ton-Repräsentationen begünstigen. Da vorhandene Befunde zum langfristigen absoluten Ton-Erinnern bislang häufig unter der Bezeichnung „latentes“ absolutes Gehör und somit im Kontext der Literatur zum absoluten Gehör beschrieben werden, wird im folgenden Kapitel zunächst das absolute Gehör dargestellt, bevor in Kapitel 3 auf die latente Form eingegangen wird.

2 Charakterisierung des absoluten Gehörs

Das absolute Gehör als die auch unter Musikern seltene Fähigkeit, Töne ohne eine externe Referenz zu benennen, fasziniert nicht nur Musiker und Musikwissenschaftler, sondern wird auch in der Psychologie bereits seit über 100 Jahren empirisch untersucht (z.B. Stumpf, 1883; Kries, 1892; Meyer, 1899; Abraham, 1902). Dabei befassen sich die frühen Schriften im Prinzip mit den Themen, die auch heute noch, z.T. mit verfeinerter Methodik, untersucht werden. Dazu zählen beispielsweise die Anlage-Umwelt Kontroverse der Entstehung absoluten Gehörs, der Leistungsvergleich zwischen Mensch und Tier, die Suche nach gedächtnispsychologischen Erklärungsmöglichkeiten der Fähigkeit sowie die Verbreitung unterschiedlicher Ausprägungsgrade absoluten Gehörs. In diesem Kapitel sollen für die vorliegende Arbeit wesentliche Aspekte des absoluten Gehörs zusammenfassend dargestellt werden. Die Konzentration liegt dabei auf messmethodischen Aspekten sowie auf typischen Befunden, die mit unterschiedlichen Methoden erhalten werden und die aus gedächtnispsychologischer Perspektive interessant sind. Außerdem werden neuere Befunde aus dem Bereich der Neuropsychologie berichtet. Eine ausführliche Zusammenfassung der frühen Literatur bis in die 1980er Jahre findet sich bei Heyde (1987), neuere Überblicksartikel stammen von Takeuchi & Hulse (1993), Ward (1999) und Vitouch (2005).

2.1 Definitionen

Absolutes Gehör wird als die Fähigkeit definiert, einzelne Töne zu benennen („passives“ absolutes Gehör) oder sehr genau zu produzieren („aktives“ absolutes Gehör), ohne dabei auf einen Referenzton (z.B. das A einer Stimmgabel) zurückgreifen zu müssen. Die Faszination, die diese so simpel anmutende Fähigkeit ausübt, liegt wesentlich an ihrer Seltenheit, aber auch an dem vor allem von der frühen Forschungsliteratur postulierten, heute jedoch kritisch hinterfragten Zusammenhang mit musikalischer (Hoch-) Begabung. In der westlichen Allgemeinbevölkerung wird die Inzidenz absoluten Gehörs mit 1:10.000 bis 1:1.000 angegeben, während sie für professionelle Musiker bei 1:100 bis 1:5 liegt (z.B. Bachem, 1940, 1955).

In der Sprache der Gedächtnispsychologie wird absolutes Gehör als LZG für absolute Tonhöhen beschrieben, wohingegen das relationale Tonhöhenurteil auf Arbeitsgedächtnis-Mechanismen beruhe (Ward, 1999). Da sie keine internen Ton-Standards im LZG gespeichert haben, müssen Nichtabsoluthörer Vergleiche mit externen Referenztönen durchführen (also die zu vergleichenden Töne durch Rehearsalmechanismen aufrecht erhalten, bis sie das Intervall bestimmt haben), um einen Tonnamen zu finden. Empirisch gestützt wird diese Unterscheidung durch EEG-Befunde, die in Abschnitt 2.3 näher beschrieben werden.

In der Sprache der Messtheorie lässt sich absolutes Gehör als die Fähigkeit, rationalskalierte Tonhöhenurteile zu fällen, definieren, wohingegen Nichtabsoluthörer nur intervallskaliert urteilen können (Gigerenzer, 1985). Aus Sicht der Informationstheorie klingt die Leistung von Absoluthörern besonders eindrucksvoll: Nach Zatorre (2003, S. 692) stellt absolutes Gehör „a notable exception to the usual cognitive limit on the number of perceptual categories for

stimuli that differ in a single dimension along a continuum“ dar. Die meisten Menschen haben nur 6-8 Kategorien für die Tonhöhe (Pollack, 1952; Miller, 1956), bei Absoluthörern dagegen sind bis zu 70 Kategorien nachweisbar (12 Tonklassen über mehr als 5 Oktaven).

Anschaulich lässt sich das absolute Gehör anhand des in Kapitel 1 erläuterten Helixmodells für Töne (vgl. Abb. 1.1) beschreiben: Absoluthörer haben im Gedächtnis Ankerpunkte für die Ton-Helix gespeichert, so dass diese sozusagen fest verankert ist und jederzeit einzelne Töne auf der Helix lokalisiert und benannt werden können. Dass dabei gelegentlich Oktavverwechslungen auftreten, wird als Hinweis dafür angesehen, dass sich Absoluthörer stärker am Chroma-Aspekt von Tönen orientieren als am Helligkeits-Aspekt. Nichtabsoluthörer dagegen verfügen nicht über einen Ankerpunkt im Sinne eines „absoluten Nullpunkts“, daher brauchen sie einen Referenzton, um die Helix vorübergehend zu fixieren und von dort aus Vergleichsurteile zu fällen. Einen Nichtabsoluthörer kann man hinsichtlich des Ankertons „täuschen“ und ihm z.B. ein *b* als *a* angeben. Viele nicht absolut hörende Chorsänger z.B. haben so lange kein Problem damit, in einer anderen Tonart zu singen, als in der Partitur notiert, bis diese Transposition an die Grenzen ihres Stimmumfangs führt.

2.2 Erfassungsmethoden und typische Befunde

Ein Standard-Verfahren zur einheitlichen Erfassung absoluter Hörleistungen liegt bisher nicht vor. Grundsätzlich werden drei Methoden angewandt, um absolute Hörleistungen zu erfassen: Tonidentifikations-, Tonproduktions- und Gedächtnisaufgaben. Daneben gibt es implizite Testmethoden, die bei der Untersuchung von Kleinkindern und Tieren zur Anwendung kommen. Jede Methode bringt unterschiedliche Aspekte absoluten Gehörs zum Vorschein, außerdem trennen die Methoden auch unterschiedlich gut zwischen Absolut- und Nichtabsoluthörern. Daher fordern Takeuchi & Hulse (1993) die Verwendung mehrerer Methoden, um auch das weniger perfekt ausgebildete absolute Gehör erfassen zu können. In den meisten Experimenten mit Absoluthörern wird allerdings nur eine Methode genutzt.

2.2.1 Tonidentifikation

Tonidentifikationsaufgaben sind in Experimenten mit Absoluthörern am weitesten verbreitet. Sie bestehen grundsätzlich darin, dass den Vpn eine Reihe von Tönen vorgespielt wird, die diese benennen sollen. Um wirklich *absolute* Tonurteile zu erfassen, ist es bei dieser Methode unerlässlich, relationale Tonurteile zu verhindern, denn musikalisch versierte Nichtabsoluthörer können von einem einzigen Ankerton aus alle folgenden Töne anhand von Intervallbestimmungen benennen. Maßnahmen zur Ausschaltung relationaler Tonurteile sind das Vermeiden von Feedback, die Beschränkung der Antwortzeit (relationale Tonurteile dauern länger als absolute), das Einschieben von Störreizen (z.B. Akkordfolgen, nachzusprechende Zahlen- oder Wortfolgen, Imitation einer Sirene) zwischen die Testtöne sowie – effektiv, aber wenig praktikabel – die Nutzung sehr langer Inter-Stimulus-Intervalle (ISIs). Experimente mit Tonidentifikationsaufgaben unterscheiden sich in der Auswahl der Töne und Klangfarben, in der Erfassung von Antwortlatenzen sowie in der Fehlertoleranz bei der Auswertung.

Experimente mit Tonidentifikationsaufgaben, bei denen Töne in unterschiedlichen Oktavlagen und Klangfarben dargeboten wurden, haben gezeigt, dass die Identifikationsrate und das Identifikationstempo vieler Absoluthörer von der Tonklasse (Marvin & Brinkman, 2000; Miyazaki, 1990; Takeuchi & Hulse, 1991), der Klangfarbe (Marvin & Brinkman, 2000; Miyazaki, 1989) und der Oktavlage (Miyazaki, 1989; Heyde, 1987) abhängig ist. Das bedeutet, dass die Töne, die auf der Klaviertastatur den weißen Tasten entsprechen (z.B. *c, d, a*), schneller und häufiger korrekt benannt werden können als Töne, die den schwarzen Tasten entsprechen (z.B. *cis, des, ais*). Das gleiche gilt für musikalische Töne (z.B. Klaviertöne) im Vergleich zu Sinustönen sowie für Töne mittlerer Lage im Vergleich zu Tönen in extrem hoher oder tiefer Oktavlage. Wodurch diese tonspezifischen Leistungsunterschiede entstehen könnten, wird in Kap. 8 im Zusammenhang mit einer eigenen Untersuchung diskutiert.

Festzuhalten ist hier, dass die bessere Identifikation von Tönen, die häufiger in der gängigen Musikliteratur auftreten (weiße Tasten, Klaviertöne, Töne mittlerer Lage) einen Hinweis auf die Lernabhängigkeit absoluter Hörleistungen liefert. Darüber hinaus hat die große interindividuelle Variabilität der Tonbenennungsleistungen von Absoluthörern, die teilweise sogar mit denen von Nichtabsoluthörern überlappen, zu der Frage geführt, ab welchem Leistungsniveau man von einem absoluten Gehör spricht. Für diese Frage gibt es bislang keine allgemeingültige Antwort. Statt dessen wird sie in Experimenten, in denen sowohl Absolut- als auch Nichtabsoluthörer untersucht werden, je nach Datenlage entschieden. Einige Experimente wählen als Cut-off Punkt für die Definition einer Vp als Absoluthörer die 80 %-Marke (z.B. Barkowsky, 1992; Tervaniemi, Alho, Paavilainen, Sams & Näätänen, 1993), wohingegen Wilson, Keefe & Reutens (2002) noch etwas strenger sind und erst ab einer Rate von 90 % Richtigurteilen von absolutem Gehör sprechen. Fragt man nach der subjektiven Einschätzung der eigenen Tonidentifikationsleistung, so schließen Terhardt & Seewann (1983) aus ihren Experimenten (vgl. Kap. 3), dass sich Absoluthörer erst ab einer kritischen Schwelle von 70-80 % richtigen Urteilen als solche wahrnehmen. Allerdings gibt es insgesamt nicht sehr viele Studien, in denen Absoluthörer und Nichtabsoluthörer mit den gleichen Anforderungen untersucht werden. Takeuchi & Hulse (1993) berichten in ihrem Übersichtsartikel bei verschiedenen Experimenten mit Absoluthörern einen mittleren Leistungsbereich von 58 bis100 %, relativieren diese große Variabilität allerdings durch den Hinweis, dass z.B. Halbton- und Oktavfehler nicht durchgängig als Fehler bewertet werden. Sie ziehen sehr nachvollziehbar die Schlussfolgerung, dass ohne einheitliche Test- und Auswertungsmethoden keine untere Leistungsgrenze absoluten Gehörs definiert werden kann. Eine verwandte Frage ist die danach, ob sich absolutes Gehör statt als dichotom ausgeprägte Fähigkeit adäquater als Leistungskontinuum beschreiben lässt. Diese Frage wird in Kapitel 3 ausführlich diskutiert.

2.2.2 Tonproduktion

Tonproduktionsaufgaben bestehen darin, dass Vpn gebeten werden, einzelne Töne durch Einstellen eines Tongenerators (seltener durch Singen oder Summen) zu produzieren. Takeuchi & Hulse (1993) schlagen vor, statt der Produktionsgenauigkeit die Konsistenz bei der

Produktion gleicher Töne auszuwerten, da diese unabhängig von der Normstimmung von A = 440 Hz sei. Die Produktion von Tönen ist schwieriger als die Identifikation, und nicht alle Absoluthörer sind zur Tonproduktion fähig (z.B. Petran, 1932; Révész, 1946). Mithilfe von Produktionsaufgaben wurde herausgefunden, dass die Genauigkeit der Tonproduktion Tages- und sogar Monatsschwankungen unterliegen kann (Wynn, 1972, 1973). Dies könnte ein Hinweis auf die Abhängigkeit beteiligter Hirnfunktionen von hormonellen - und Stoffwechselprozessen sein. Außerdem konnte die obere Leistungsgrenze absoluter Hörleistungen auf 0,1 Halbtöne (Bachem, 1937, 1940) bis 0,2 Halbtöne (Rakowski, 1978; van Krevelen, 1951) festgelegt werden. Da Produktionsaufgaben für Absoluthörer schwieriger sind als Wiedererkennungsaufgaben, ist es nicht überraschend, dass erstere schärfer zwischen Absolut- und Nichtabsoluthörern trennen können als letztere (Petran, 1932; Rakowski & Morawska-Büngeler, 1987; Siegel, 1972).

2.2.3 Gedächtnisaufgaben

Per definitionem handelt es sich bei absolutem Gehör um die Fähigkeit, Töne im LZG zu behalten, während Nichtabsoluthörer Töne nach einer gewissen Zeit vergessen. Dieser Unterschied wurde mithilfe von Gedächtnisaufgaben untersucht, bei denen Vpn entweder ein Vergleichsurteil über zwei Töne (gleich vs. unterschiedlich) fällen oder einen gehörten Ton nach bestimmten Zeitintervallen reproduzieren sollen. Dabei variiert das Intervall zwischen den beiden Tönen bzw. nach dem ersten Ton von einigen Sekunden bis eine Woche. Beide Aufgaben führten zu ähnlichen Ergebnissen: Für Retentionsintervalle unter einer Minute findet nur sehr wenig Vergessen statt, und es gibt dabei keine Unterschiede zwischen Absolut- und Nichtabsoluthörern. Bei Zeitintervallen ab einer Minute aufwärts dagegen unterscheiden sich die beiden Gruppen: Die Leistung von Absoluthörern bleibt auf gleichem Niveau konstant, während die von Nichtabsoluthörern auf das Zufallsniveau absinkt (Bachem, 1940, 1954; Rakowski, 1972; Rakowski & Morawska-Büngeler, 1987). Diese Ergebnisse werden von Takeuchi & Hulse (1993) als Hinweis darauf angesehen, dass bei kurzen Retentionsintervallen sowohl Absoluthörer als auch Nichtabsoluthörer das Echogedächtnis nutzen. Gleichzeitig scheinen Absoluthörer den Namen gehörter Töne abzuspeichern, den sie auch dann noch abrufen können, wenn die Echogedächtnis-Spur zerfallen ist. Da die verbale Kodierung von Tonnamen Nichtabsoluthörern nicht möglich ist, sinkt ihre Leistung bei längeren Retentionsintervallen stark ab.

Die Interpretation, dass eine verbale Kodierung gehörter Töne die Leistungsfähigkeit des LZG von Absoluthörern ausmacht, wird durch Befunde von Siegel (1974) erhärtet. Siegel zeigte mit einer an Deutsch (1970) erinnernden Methodik, dass Absoluthörer beim Vergleich zweier Töne (gleich vs. unterschiedlich) mit „interpolierten“ Tönen im Retentionsintervall nur dann einen Leistungsvorteil gegenüber Nichtabsoluthörern haben, wenn die Vergleichstöne sich um mindestens 0,75 Halbtöne unterscheiden und das Retentionsintervall länger als eine Minute ist. Bei kürzeren Intervallen konnten beide Gruppen das Echogedächtnis nutzen, während bei kleineren Tonabständen die verbale Kodierung nicht mehr „griff“ (offensichtlich konnte verbal nicht zwischen „etwas zu hohem *c*“ und „etwas zu niedrigem *c*“ unterschieden

werden). Auch Bachems (1954) Ergebnis, dass Töne oberhalb von 5000 Hz, die keine definierbare Tonklasse mehr zu haben scheinen („chroma fixation“), von Absoluthörern nicht besser erinnert werden konnten als von Nichtabsoluthörern, verweist auf die Bedeutung der verbalen Kodierung von Tönen bei Absoluthörern. Relativierend weisen allerdings Zatorre & Beckett (1989) auf das Vorhandensein multipler Kodierungen bei Absoluthörern hin. Ihre Vpn wurden während des Behaltens von Buchstaben- und Tonfolgen über bis zu 27 Sekunden mit verbalen und musikalischen Interferenzaufgaben gestört. Es zeigte sich, dass verbale Interferenzaufgaben das Behalten der Buchstaben-, aber nicht der Tonfolgen beeinträchtigten. Musikalische Interferenzaufgaben hatten weder auf das Behalten der Buchstaben- noch der Tonfolgen einen Einfluss. Als weitere Kodierungen neben der verbalen postulieren die Autoren einen auditiven Toneindruck, kinästhetische sowie visuelle Toneigenschaften.

2.2.4 Implizite Aufgaben

Implizite Aufgaben zur Erfassung absoluter Hörleistungen wurden bisher vor allem dort angewandt, wo explizite Aufgaben nicht oder nur schwer nutzbar waren: bei der Untersuchung von Tieren und sehr jungen Kindern (vor dem Spracherwerb). Die Untersuchung solcher Probanden erfolgte vor dem Forschungshintergrund der „perceptual invariances“ – Töne im Oktavabstand oder transponierte Melodien beispielsweise werden von erwachsenen Menschen als perzeptuell invariant (= äquivalent, aber unterscheidbar) angesehen. Kann man bei Tieren oder jungen Kindern Unterschiede bei der Wahrnehmung von Invarianzen nachweisen, so wäre das ein Hinweis auf eine mögliche Wurzel menschlicher Musikwahrnehmung (vgl. ausführlicher: Hulse, Takeuchi & Braaten, 1992).

Eine Methode, die bisher vor allem bei Singvögeln zur Anwendung kam, besteht darin, dass die Tiere lernen, auf eine Scheibe zu picken, wenn sie eine bestimmte Klangsequenz, z.B. eine aufsteigende Tonfolge, hören und bei einer anderen Klangsequenz, z.B. einer absteigenden Tonfolge, nicht zu picken. Ausgewertet wird die Latenz zwischen Klangdarbietung und Beginn des Pickens. Die Tonsequenzen bestehen meist aus computergenerierten künstlichen Klangfolgen. Eine Reihe von Untersuchungen, die seit den 1980er Jahren von Stewart Hulse und Mitarbeitern publiziert wurde (Braaten, Hulse & Page, 1990; Cynx, Hulse & Polyzois, 1986; Hulse & Cynx, 1985, 1986; Hulse, Cynx & Humpal, 1984; Hulse, Page & Braaten, 1990; Page, Hulse & Cynx, 1989) hat gezeigt, dass Sperlinge lernen können, fallende und steigende Tonfolgen zu unterscheiden und dabei unterschiedliche Intervalle als äquivalent zu betrachten. Allerdings gelingt die Transferleistung z.B. von kleinen Sekundschritten (= ein Halbton) auf große Sekundschritte (= zwei Halbtöne) nur dann, wenn Trainings- und Testsequenzen in der gleichen Oktavlage dargeboten werden („frequency range constraint“). Durch eine Variante des Experiments, in der nur ein Ton der Trainingssequenzen wiederholt dargeboten wurde, worauf die Vögel ähnlich reagierten wie auf die Tonfolgen, konnte man feststellen, dass sie vorrangig auf der Basis der absoluten Tonhöhen (tiefer bei steigenden Sequenzen, höher bei fallenden Sequenzen) reagiert haben (Hulse & Cynx, 1985). Hulse et al. (1992) folgern aus ihren Ergebnissen, dass Vögel eine Hierarchie der Informationsverarbeitung haben, nämlich die Verarbeitung absoluter Tonhöhen bevorzugen. Nur wenn sie mit

Reizen konfrontiert werden, bei denen absolute *und* relationale Informationen relevant sind, werden auch Tonrelationen verarbeitet. Diese Ergebnisse werden durch Experimente mit „natürlicheren“ Vogelgeräuschen gestützt (Ratcliffe & Weisman, 1985; Weisman, Ratcliffe, Johnsrude & Hurly, 1990; Hill & Lein, 1987). Unter Nutzung der gleichen Tonfolgen wie Hulse und Mitarbeiter fand D'Amato (1988) bei Cebus-Affen ebenfalls einen frequency range constraint. Auch Moody & Stebbins (1989, 1990) konnten mit ähnlicher Methodik bei Makaken eine Bevorzugung der Verarbeitung absoluter Tonhöhen feststellen.

Während die bisher beschriebene Versuchsanordnung es kaum erlaubt, erwachsene Menschen unter ähnlichen Bedingungen zu testen, um einen echten Vergleich der bevorzugten Informationsverarbeitungsebene zu erhalten, haben Saffran & Griepentrog (2001) eine Methode entwickelt, die einen solchen Vergleich erlaubt. Sie nutzten eine statistische Segmentationsaufgabe, bei der Tonfolgen zunächst in einer dreiminütigen Lernphase dargeboten wurden. Die Tonfolgen bestanden aus einer Serie von „tone words“, wobei die Bezeichnung „Wort“ hier im statistischen Sinn genutzt wird und eine Tonfolge definiert, bei der die Übergangswahrscheinlichkeiten zwischen den Tönen *innerhalb* des Wortes höher ist als zwischen Tönen an den *Wortgrenzen*. Beispielsweise folgt bei einem Wort „*c fis g*“ auf ein *c* immer ein *fis* und auf das *fis* immer ein *g*, wohingegen auf das *g* verschiedene neue Wörter (also Töne) folgen können. In der Testphase wurden dann einzelne tone words dargeboten, die teilweise den tone words aus der Lernphase entsprachen und teilweise aus Teilen zwei verschiedener tone words zusammengesetzt waren („part words“). Die Testwörter wurden so konstruiert, dass die Tonrelationen bei tone words und part words gleich waren, so dass sie sich nur hinsichtlich der absoluten Tonhöhen unterschieden. Acht Monate alte Babies, bei denen die Hörpräferenz über die Dauer des Hinblickens gemessen wurde, zeigten mehr Interesse für die neuen part words als für bekannte tone words. Dies interpretieren die Autoren als Hinweis darauf, dass sie die absolute Information der Reize nutzen konnten, da keine aufgabenrelevante relationale Information vorlag. In einer weiteren Aufgabe, die (genau entgegengesetzt) *nur* auf der Basis der Nutzung relationaler Information gelöst werden konnte, ergaben sich keine Unterschiede in den Blickdauern bei tone words und part words. Offensichtlich konnten die Babies die relationale Information nicht in gleicher Weise nutzen wie die absolute. Erwachsene Vpn, die die gleichen Lernsequenzen hörten und dann mit einer forced-choice Aufgabe zwischen tone words und part words unterscheiden sollten, zeigten ein genau entgegengesetztes Antwortmuster: Sie zeigten überzufällige Leistungen, wenn sie relationale Informationen nutzen konnten; stand dagegen nur absolute Information zur Verfügung, antworteten sie auf dem Zufallsniveau.

Interessanterweise änderte sich das Antwortverhalten sowohl der Babies als auch der Erwachsenen, wenn tonale (statt bisher atonale) Tonsequenzen verwendet wurden (Saffran, 2003): Babies schauten in der Absolut-Bedingung nun länger auf die bekannten tone words als auf part words, während sich in der Relativ-Bedingung wieder keine Unterschiede in den Blickzeiten ergaben. Erwachsene dagegen konnten besser die tone words aus der Relativ-Bedingung wiedererkennen, aber auch in der Absolut-Bedingung war ihre Leistung nun überzufällig. Diese Daten liefern einen Hinweis auf entgegengesetzte Präferenzen zwischen

Babies und Erwachsenen bei der Nutzung absoluter und relationaler Information. Ähnlich wie die Ergebnisse aus den Tierstudien deutet sich eine Präferenz der Nutzung absoluter Information auch bei Babies an. Trotzdem ist methodisch anzumerken, dass ein noch überzeugenderer Nachweis gegenläufiger Informationsverarbeitungstendenzen bei Erwachsenen und Babies dann erbracht wäre, wenn auch die Erwachsenen mit einer impliziten Methode getestet worden wären. Eine Untersuchung von 3- und 5-jährigen Kindern mit gleicher Testmethode haben White, Dale & Carlsen (1990) durchgeführt: Bei der Aufgabe, unter drei Dreitonsequenzen diejenige zu benennen, die nicht zu den anderen beiden passt, wählten die 3jährigen Kinder eher diejenige mit abweichendem Anfangston oder Ambitus, während 5jährige Kinder bei der Entscheidung die Melodierichtung (fallend vs. steigend) mit einbezogen. Allerdings bemerken Hulse et al. (1992) zu recht, dass dreitönige „Melodien“ nur sehr wenig relationale Information zur Verfügung stellen und dass Untersuchungen mit längeren Tonsequenzen aufschlussreich wären.

Eine noch wichtigere Kritik an Saffrans Methodik (die im Prinzip auch Whites Untersuchung betrifft) stammt von Trehub (2003) und betrifft die Erinnerungsspanne. In allen besprochenen Aufgaben handelt es sich weniger um Aufgaben für das LZG als vielmehr um arbeitsgedächtnisbasierte Aufgaben. Auf die langfristige Repräsentation absoluter Tonhöhen kann man aus diesen Ergebnissen nur schwer schließen. Hinweise auf die frühe Bevorzugung absoluter Informationen auch für das langfristige Erinnern von Melodien stammen bisher vor allem aus Experimenten mit expliziten Methoden. Sergeant & Roche (1973) lehrten 3-6jährige Kinder verschiedene Lieder und analysierten die Tonarten, in denen die Kinder die Lieder spontan wiedergaben. Es zeigte sich, dass die Genauigkeit, mit der die absoluten Tonhöhen produziert wurden, mit zunehmendem Alter abnahm, während die Genauigkeit, mit der relationale Merkmale (Kontur, Intervalle, Tonalität) produziert wurden, zunahm.

Insgesamt weisen die referierten Ergebnisse also darauf hin, dass im Alter von ca. 6-8 Jahren ein Wechsel der Verarbeitungsstrategie für Melodien stattfindet, nämlich eine zunehmende Bevorzugung relationaler statt absoluter Information (vgl. Hulse et al., 1992; Takeuchi & Hulse, 1993). Die Verarbeitung relationaler Information „stört“ die Orientierung an absoluten Tonhöhen, und da erstere für die Melodie- und Sprachwahrnehmung die maßgebliche Information darstellt, geht die absolute Strategie allmählich verloren, es sei denn, es wird ein „echtes“ absolutes Gehör erworben. Diese Argumentation (die sich auf die Verlerntheorie absoluten Gehörs von Abraham, 1901, gründet) besticht insofern, als die genannte Altersgrenze ungefähr mit der für den Erwerb absoluten Gehörs, wie auch für den akzentfreien Erwerb von Fremdsprachen, übereinstimmt. Die im folgenden Kapitel dargestellten Formen „latenten“ absoluten Gehörs könnten demnach Verarbeitungsresiduen aus der frühen Kindheit darstellen.

2.3 Neuropsychologische Befunde

In jüngster Zeit wurde das absolute Gehör besonders in der kognitiven Neurowissenschaft intensiver untersucht. Dies hat nach Zatorre (2003) zwei Gründe: Erstens könne man am

absoluten Gehör untersuchen, wie eine hochspezialisierte Fähigkeit (die keine Zusammenhänge zu anderen kognitiven Funktionen zu haben scheint) mit Gehirnfunktionen zusammenhänge, und zweitens sei diese Fähigkeit ein gutes Beispiel für die Interaktion von genetischen und Umweltfaktoren während der kognitiven Entwicklung. Untersucht wurden sowohl funktionale als auch morphologische Aspekte, die absolutem Gehör zugrunde liegen könnten.

2.3.1 Morphologische Besonderheiten bei Absoluthörern

Eine vielbeachtete morphologische Besonderheit im Gehirn von Absoluthörern wurde von Schlaug, Jäncke, Huang & Steinmetz (1995) beschrieben und von Zatorre, Perry, Beckett, Westbury & Evans (1998) im wesentlichen bestätigt. Beide Teams fanden mithilfe der Magnetresonanztomographie (MRT) eine Vergrößerung der typischen Linksasymmetrie des planum temporale bei Absoluthörern sowohl im Vergleich mit Nichtabsoluthörern als auch im Vergleich mit Musikern. Ein Befund von Keenan, Thangaraj, Halpern & Schlaug (2001) deutet darauf hin, dass die stärkere Linksasymmetrie nicht auf einer linksseitigen Vergrößerung, sondern auf einem rechtsseitig geringeren Volumen beruht, was die Interpretation erschwert. Die Funktion des planum temporale, das in der Nähe des primären auditorischen Kortex liegt, wird mit der Verarbeitung sprachlicher Informationen angegeben, insofern liegt eine Interpretation der stärkeren Linksasymmetrie als anatomischer „marker“ für die Benennungskomponente absoluten Gehörs nahe. Ob eine stärkere Linksasymmetrie des planum temporale allerdings Ursache oder Folge der Entwicklung eines absoluten Gehörs ist, ließe sich nur durch Längsschnittstudien, die es bislang noch nicht gibt, endgültig klären. Einerseits wurden Asymmetrien in dem hier relevanten Bereich des Gehirns schon bei Neugeborenen und sogar vor der Geburt nachgewiesen (Witelson & Pallie, 1973). Andererseits zeigen Vermessungen der Gehirne nicht nur von Musikern, welch große Veränderungen im Gehirn durch Erfahrungen (wie z.B. tägliches Musizieren) hervorgerufen werden können (vgl. im Überblick: Spitzer, 1996, 2002; Pantev, Engelien, Candia & Elbert, 2003).

Erschwert wird die Interpretation der morphologischen Befundlage außerdem durch eine klinische Einzelfallstudie von Zatorre (1989), der einen Absoluthörer mit epileptischen Anfällen nach Entfernung eines Teils des linken Temporallappens untersucht hat. Die Operation war insofern erfolgreich, als die Anfälle nachließen, das absolute Gehör aber erhalten blieb. Das bedeutet, dass entweder nur Gehirnregionen entfernt wurden, die für das absolute Gehör nicht relevant sind (anteriorer, nicht aber posteriorer Temporallappen) oder dass der linke Temporallappen nicht der alleinige „Sitz“ des absoluten Gehörs ist. Vitouch (2005) schlägt als integrative Interpretation der Ergebnisse von Schlaug et al. (1995) und Zatorre (1989) vor, die berichtete verstärkte Linksasymmetrie bei Absoluthörern als stärkere Sprachlateralisierung (nach links) zu betrachten, die eine stärkere Spezialisierung der rechten Seite für Tonhöhenanalysen ermöglichen könnte, da dort sozusagen noch Kapazitäten frei sind. Diese Interpretation verdeutlicht, dass es zum besseren Verständnis zugrundeliegender Prozesse sehr hilfreich wäre, gleichzeitig morphologische Vermessungen und funktionelle Analysen aktiver Areale beim Musikhören (oder spezieller: beim Tönebenennen) vorzunehmen.

2.3.2 Funktionelle Besonderheiten bei Absoluthörern

Mithilfe der Elektroenzephalographie (EEG) ist es möglich, elektrische Hirnpotenziale, die mit bestimmten Reizen oder Ereignissen korreliert auftreten (ereigniskorrelierte Potenziale, kurz EKPs), zu identifizieren. Dabei treten je nach Aufgabenstellung verschiedene EKPs auf, die nach ihrer Auslenkung und der Auftrittslatenz benannt werden. Eine Komponente der basalen Reizverarbeitung ohne bewusste Aufmerksamkeitszuwendung ist die Mismatch Negativity (MMN), die detailliert von Näätänen und Kollegen beschrieben wurde (Zusammenfassung bei Näätänen, 1992). Die MMN tritt auf, wenn nach einer Folge gleicher Reize („standards", z.B. Töne einer bestimmten Tonhöhe, Lautstärke, Klangfarbe oder Intensität) ein unterschiedlicher Reiz („deviant") erklingt (das „oddball"-Paradigma). Der „deviant" bewirkt eine Negativierung, die 100-200 ms nach seinem Auftritt maximal ist. Diese Komponente ist insofern automatisch, da sie auch auftritt, wenn der Proband ein Buch liest oder gleichzeitig eine visuelle Aufgabe bearbeitet. Aufschlussreich für die Reizverarbeitung von Absoluthörern ist eine MMN-Untersuchung von Tervaniemi, Alho, Paavilainen, Sams & Näätänen (1993). Dort konnte bei der Untersuchung von Absolut- und Nichtabsoluthörern gezeigt werden, dass „deviants", die um einen Viertelton oder Halbton vom „standard" abweichen, eine MMN auslösen und dass sich diese zwischen Absolut- und Nichtabsoluthörern nicht unterschied. Dieser Befund korrespondiert mit Verhaltensdaten aus anderen Experimenten, die zeigten, dass Absoluthörer Töne kurzfristig nicht feiner kategorisieren können als Nichtabsoluthörer (vgl. Siegel, 1972, Oakes, 1955). Das „Besondere" des absoluten Gehörs scheint demnach nicht in der Phase der initialen Reizverarbeitung zu liegen.

Die Besonderheit der Reizverarbeitung von Absoluthörern zeigt sich stärker in Aufgaben, die eine Aufmerksamkeitszuwendung erfordern und das Arbeitsgedächtnis beanspruchen. Solche Aufgaben wurden ebenfalls mithilfe des oddball-Paradigmas untersucht. Eine Reihe von Experimenten widmet sich der Untersuchung der P300, die in Zusammenhang mit Kontext-updating Prozessen im Arbeitsgedächtnis gebracht wird (Donchin, 1981; Donchin & Coles, 1988). Wird eine Vp aufgefordert, beim Hören von Tönen auf „deviants" mit einem Tastendruck zu reagieren, so erscheint ca. 300 ms nach Auftritt des „deviants" eine Positivierung. Eine Untersuchung von Klein, Coles & Donchin (1984) fand bei Absoluthörern eine kleinere P300 als bei Nichtabsoluthörern, jedoch nur bei einer tonalen, nicht bei einer visuellen Aufgabe. Sie fassen dieses Ergebnis als Beleg dafür auf, dass Absoluthörer die tonale Aufgabe mit einer absoluten Strategie lösen und daher kein Kontext-updating notwendig sei, während es bei Nichtabsoluthörern, die eine relationale Strategie anwenden, zu Aktualisierungsprozessen komme. Diese Interpretation wurde durch zum Teil nicht geglückte Replikationsversuche (zusammengefasst in Bischoff, 2000) in Frage gestellt. Um zur Klärung dieser Frage beizutragen, stellte Bischoff (2000) ihren Vpn (Absoluthörer und Nichtabsoluthörer) zwei auditive und eine visuelle oddball-Aufgabe, wobei eine auditive Aufgabe eher durch eine absolute Strategie (Entdeckung nicht diatonischer Töne in tonalem Kontext) und die andere eher durch eine relationale Strategie (Unterscheidung auf- und absteigender Intervalle) zu lösen war. Sie fand bei Absoluthörern, die auch gute relationale Hörfähigkeiten hatten und nach eigener Angabe eine relationale Strategie anwandten, eine normal ausgeprägte P300, bei

Absoluthörern mit schwächeren relationalen Hörfähigkeiten dagegen eine reduzierte P300. Die P300 ist diesen Ergebnissen zufolge als „marker“ für die angewandte Strategie aufzufassen, die unter Absoluthörern durchaus nicht einheitlich sein muss.

Die Nutzung des Arbeitsgedächtnisses bei Absoluthörern wurde auch mithilfe der Positronen-Emissions-Tomographie (PET) untersucht. Zatorre et al. (1998) fanden, dass ein Areal im rechten inferioren Frontalkortex (Area 47 / 11), das mit dem Erhalt von Tonhöheninformation im Arbeitsgedächtnis in Zusammenhang gebracht wird, bei Absoluthörern beim Hören von Tönen weniger aktiv ist als bei Nichtabsoluthörern. Statt dessen war bei den Absoluthörern ein Areal im linken posterioren dorsolateralen Frontalkortex (Area 8 / 6), das ansonsten mit dem Aufbau konditionierter Assoziationen in Zusammenhang gebracht wird, beim Tönehören aktiver als bei Nichtabsoluthörern. Dies könnte demnach das Gebiet sein, in dem die verbalen Marken für die Tonkategorien gespeichert sind. Bei Nichtabsoluthörern ist das gleiche Gebiet aktiv, wenn sie gebeten werden, Intervalle zu benennen, eine für sie lösbare Aufgabe (Bermudez & Zatorre, 2001). Zatorre (2003) folgert aus diesen Befunden, dass das absolute Gehör auf Hirnstrukturen aufbaut, die aus anderen Kontexten bekannt und erforscht sind. Trotzdem sei bislang nicht erklärbar, wie es zu der festen Assoziation zwischen Tonkategorien und Tonnamen kommt, die letztlich zur geringeren Beanspruchung des Arbeitsgedächtnisses von Absoluthörern bei der Erinnerung an Einzeltöne führt. Zatorre (2003) vermutet, dass die Antwort teilweise in der Funktion subkortikaler Kerne, beispielsweise des inferioren Colliculus, die Periodizität und Gleichzeitigkeit kodieren, liegen könnte, die jedoch noch weiterer Untersuchung bedarf (vgl. Langner, Albert & Briede, 2002).

Eine etwas andere Methodik als Zatorre et al. (1998), die auch zu einem anderen Aktivitätsmuster führte, wurde in einer PET-Studie von Wilson, Keefe & Reutens (2002) angewandt: Subtrahiert man von der Aktivität während einer Tonbenennungs-Bedingung die Aktivität einer Bedingung, in der Vpn Geräusche hörten und darauf mit „C“ antworten sollten, so war nicht der linke, sondern der rechte Frontalkortex (Area 10) aktiv. Subtrahiert man wiederum von der Aktivität einer Bedingung, in der ein Ton als relativ zu einem Akkord als „tonal“ oder „atonal“ klassifiziert werden sollte (relationale Aufgabe), die Aktivität während der Tonbenennungs-Bedingung, so ist auch keine (relative) Aktivität des rechten Frontalkortex’ mehr nachweisbar. Bei Nichtabsoluthörern war in der gleichen Region weder in der Tonbenennungs- noch in der Akkordklassifikations-Bedingung eine (relative) Aktivität nachweisbar. Wilson et al. (2002) schließen aus diesem Ergebnis, dass der rechte Frontallappen, der sich in anderen Studien als wichtig für den Abruf von episodischen und semantischen Gedächtnisinhalten erwiesen habe, bei Absoluthörern während der Tonbenennung aktiv sei, weil sie erfolgreich einen Tonnamen aus dem Gedächtnis abrufen. Dies zeige sich allerdings nur, wenn Effekte der verbalen Benennung durch eine geeignete Baseline-Bedingung eliminiert werden. Die besondere Bedeutung, die der rechten Hirnhälfte zugemessen wird, erinnert an die Interpretation der morphologischen Befundlage von Vitouch (2005, s.o.).

Die Frage, welche Rolle das planum temporale für die Entwicklung des absoluten Gehörs spielt, bleibt unbeantwortet, da sich dort weder bei Zatorre et al. (1998) noch bei Wilson et al.

(2002) Aktivierungen gezeigt haben. Wilson et al. postulieren, dass die planum temporale Asymmetrie bei Absoluthörern ein Effekt des frühen Musizierens und somit nicht spezifisch für absolutes Gehör sein könnte. Um diesen Aspekt zu untersuchen, wäre ein Experiment mit Absoluthörern und Nichtabsoluthörern, die sich hinsichtlich des Ausbildungsbeginns nicht unterscheiden, erforderlich.

2.4 Was macht absolutes Hören aus und wie entsteht es?

Die neuropsychologischen Befunde verdeutlichen, dass eine Trennung von Repräsentations- und Benennungskomponenten des absoluten Gehörs gerechtfertigt ist, da durch beide Komponenten unterschiedliche Hirnregionen beansprucht werden. Während die Repräsentationskomponente sich durch den erfolgreichen Abruf von Tonkategorien aus dem LZG und dadurch geringere Beanspruchung des Arbeitsgedächtnisses bei Tonvergleichsaufgaben charakterisieren lässt, besteht die Benennungskomponente in der Assoziation der Tonkategorien mit Tonnamen. Dabei ist zu bedenken, dass die Benennung der Tonkategorien wichtig, aber nicht allein entscheidend für die Gedächtnisleistung von Absoluthörern ist, denn die Befunde von Zatorre & Beckett (1989) verweisen auf eine multiple Kodierung von Tönen bei Absoluthörern. Zahlreiche Absoluthörer beschreiben dies dadurch, dass verschiedene Töne für sie eine ganz bestimmte Klangqualität haben, die sie automatisch beim Hören eines Tons empfinden.

Wie nun entsteht eine Repräsentation von Tonkategorien, Tonnamen und Toncharakteren? Anhand der beschriebenen Tonbenennungsunterschiede (z.B. Effekt der schwarzen und weißen Klaviertasten) wird sichtbar, dass Lernfaktoren bei der Entwicklung absoluten Gehörs eine Rolle spielen müssen. Es zeichnet sich die Tendenz ab, dass Töne, die entweder insgesamt in der Musikliteratur oder in der persönlichen Hörerfahrung häufiger auftreten, besser bzw. schneller benannt werden können (Simpson & Huron, 1996). Ein weiteres Argument für die Bedeutung von Lernfaktoren ist die Beobachtung, dass Absoluthörer meist sehr früh zu musizieren begonnen haben, während selten von Absoluthörern ohne irgend eine Form musikalischer Ausbildung berichtet wird. Befragungen von Musikern ergaben eine hohe Korrelation zwischen Beginn der Instrumentalausbildung und der Wahrscheinlichkeit, ein absolutes Gehör zu haben (Sergeant, 1969; Miyazaki, 1988; Baharloo, etc, 1998). Außerdem zeigt sich eine erhöhte Auftrittsrate absoluten Gehörs bei Geburtsblinden, für die eine Orientierung an absoluten Tonhöhen funktional ist (Welch, 1988; Hamilton, Pascual-Leone, Rodrigues & Schlaug, 2000) und die Klänge auch besser lokalisieren können (Rauschecker, 2001; Röder, Teder-Sälejärvi, Sterr, Rösler, Hillyard & Neville, 1999). Andererseits betonen viele Absoluthörer, dass sie niemals explizit Tonnamen gelernt haben, sondern ihr absolutes Gehör „ganz natürlich“ zu ihnen kam (Corliss, 1972). Trotz des Ausprobierens verschiedenster Methoden zum Erwerb absoluten Gehörs, auf die hier nicht näher eingegangen werden kann (vgl. Vitouch, 2005, für einen Überblick), ist noch immer unklar, *was* Absoluthörer durch frühes Musizieren genau lernen. Möglicherweise werden lediglich durch häufiges Spielen nach Noten gehörte Töne mit ihren Namen (der beim Hören in der Partitur gelesen wird) assoziiert und umgekehrt Tonnamen mit einem charakteristischen Klang. In jedem Fall scheint ein musika-

lisches Training, das auf relationale Tonaspekte fokussiert, z.B. das Spielen einer bekannten Melodie aus dem Gedächtnis in verschiedenen Tonlagen, dem Erwerb absoluten Gehörs eher abträglich zu sein.

Dass Lernfaktoren nicht alleine entscheidend sind, zeigt sich zum einen an der Tatsache, dass mit Ausnahme von Brady (1970) keine Fälle berichtet werden, in denen erwachsene Menschen erfolgreich ein absolutes Gehör erlernten. Dies verweist auf das Vorhandensein eines „Lernfensters“ für den Erwerb absoluten Gehörs, das etwa mit Erreichen des 6. Lebensjahres zu Ende geht (Prägungstheorie absoluten Gehörs, zuerst von Copp, 1916, formuliert). Zum anderen zeigt sich, dass auch nicht alle trainierten Kinder ein absolutes Gehör erlernen, was vermutlich einerseits an den Trainingsmethoden liegt, andererseits aber auch ein Hinweis für die Notwendigkeit vererbter Voraussetzungen sein könnte. Die Annahme, dass genetische Einflüsse auf die Ausprägung absoluten Gehörs vorliegen, stammt zum einen aus familiären Häufigkeitsschätzungen der Fähigkeit. Eine Befragung von Baharloo, Service, Risch, Gitschier & Freimer (2000) ergab eine vierfach erhöhte Auftrittswahrscheinlichkeit absoluten Gehörs bei Familienangehörigen von Absoluthörern. Allerdings könnte dies ebenso an einer die Entwicklung absoluten Gehörs begünstigenden Entwicklungsumgebung liegen (vgl. Levitin, 1999; Vitouch, 2005). Einen weiteren Hinweis auf genetische Einflüsse liefert die erhöhte Auftrittsrate absoluten Gehörs in mehreren asiatischen Kulturen, die nicht nur auf das Sprechen von Tonsprachen (Japanisch und Koreanisch sind z.B. keine Tonsprachen) und nicht nur auf ein (in asiatischen Ländern typisches) frühes musikalisches bzw. spezielles Absoluthör-Training zurückzuführen ist (Zatorre, 2003). Allerdings liegt ein wirklicher Beweis für genetische Einflüsse in Form von kulturvergleichenden Absoluthör-Trainings mit vergleichbaren Trainings- und Erfassungsmethoden bislang nicht vor.

Die Frage, *was* genau vererbt werden könnte, das die Entwicklung eines absoluten Gehörs fördere, wird von Chin (2003) diskutiert. Sie nimmt an, dass vererbte kognitive Stile für den individuellen Trainingserfolg hinsichtlich der Entwicklung absoluten Gehörs mit verantwortlich seien. Eine eher analytische Informationsverarbeitung sowie ein eher schmaler Aufmerksamkeitsfokus bedingen die Konzentration auf Einzeltöne statt einer ganzheitlichen Verarbeitung von Melodien. Erhalten Kinder mit einem solchen kognitiven Stil früh genug Musikunterricht, der absoluten Hörstrategien zumindest nicht entgegenläuft, so sei die Entwicklung eines absoluten Gehörs wahrscheinlich. Zur Stützung dieser Theorie werden klinische Befunde herangezogen. Beispielsweise zeichnen sich autistische Kinder und Erwachsene durch eine extrem analytische Informationsverarbeitung mit teilweise verblüffendem Gedächtnis für Details aus. In Übereinstimmung mit Chins Hypothese zeigten Heaton, Hermelin & Pring (1998), dass autistische Kinder die Kombination von Tierbildern mit Einzeltönen (= eine dem absoluten Gehör analoge Leistung) signifikant besser erlernten als eine gesunde Kontrollgruppe. In einer Gruppe autistischer Erwachsener konnten Mottron, Peretz & Menard (2000) eine bessere Orientierung an absoluten Tonhöhen nachweisen als bei einer gesunden Kontrollgruppe. Eine Gruppe gesunder Absoluthörer (N = 13) wurde von Brown, Cammuso, Sachs, Winklosky, Mullane, Bernier, Svenson, Arin, Rosen-Sheidley & Folstein (2003) hinsichtlich kognitiver und sozialer Aspekte untersucht und mit Nichtabsoluthörern verglichen.

Sie fanden anhand standardisierter klinischer Interviews eine höhere Rate „sozial exzentrischer“ Absoluthörer (46 %) als Nichtabsoluthörer (15 %). Außerdem schnitten die Absoluthörer in einem Block-Design Test, der eine einzelheitliche Informationsverarbeitung misst, relativ zu einigen anderen Intelligenztestaufgaben besser ab, die Nichtabsoluthörer aber nicht. Da dieses Befundmuster typisch für autistische Verhaltensweisen im weitesten (subklinischen) Sinne ist, folgern die Autoren, dass die einzelheitliche Informationsverarbeitung im Sinne einer starken Orientierung an Detailinformationen ein Kennzeichen absoluten Gehörs sein könnte.

Die Annahme, dass bestimmte kognitive Stile vererbt werden, ist auch deshalb überzeugender als die Suche nach einem einzelnen Gen für absolutes Gehör, da die kognitiven Defizite, die mit einer erhöhten Auftrittsrate absoluten Gehörs einhergehen, nicht alle die gleichen (genetischen) Ursachen haben. Beim Williams-Syndrom, für das ebenfalls eine erhöhte Auftrittsrate absoluten Gehörs nachgewiesen wurde, liegt als Erklärung für den in höherem Alter möglichen Erwerb absoluten Gehörs der im Vergleich zu gesunden Kindern spätere oder sogar ausbleibende Wechsel von lokaler (präoperationaler: Piaget, 1950) zu globaler (operationaler) Informationsverarbeitung nahe (Case, Okamoto, Griffin, McKeough, Bleiker, Henderson & Stephenson, 1996; Sameroff & Haith, 1996; Siegler, 1996). Eine begrenzte Arbeitsgedächtniskapazität bei Kindern mit Williams-Syndrom könnte der Grund dafür sein, dass sie sich bei der Melodiewahrnehmung auch in höherem Alter an den absoluten Tonhöhen orientieren (Lenhoff, Perales & Hickok, 2001). Ob die von Brown et al. (2003) beobachteten sozialen und kognitiven Besonderheiten von Absoluthörern sich auch in einer größeren Stichprobe zeigen, und erst recht der Nachweis, dass sie vererbt und die Ursache der Entwicklung absoluten Gehörs sind, erfordert weitere Untersuchungen, auch solche mit Längsschnittmethodik.

Aus der erhöhten Auftrittsrate absoluten Gehörs bei bestimmten kognitiven Defiziten wird deutlich, dass sich absolutes Gehör auch als Generalisierungsdefizit beschreiben lässt. Es tritt vermehrt bei solchen Patientengruppen auf, die nicht von der absoluten Tonhöhe einer Melodie abstrahieren *können*. In einem nichtklinischen Zusammenhang konnte Miyazaki (1995, 2004; Miyazaki & Rakowski, 2002) auch bei gesunden Personen zeigen, dass absolutes Gehör nicht in allen Situationen von Nutzen ist. Bei der Darbietung von Tönen, die im Vergleich zur Normstimmung von A = 440 Hz um bis zu einen Viertelton verstimmt waren, und der Aufgabe der Intervallbestimmung zwischen zwei Tönen, urteilten Absoluthörer im Vergleich zu Nichtabsoluthörern langsamer und fehlerhafter, während es bei nicht verstimmten Tönen keine Leistungsunterschiede gab. Dies führte Miyazaki (1995) auf eine der Aufgabe nicht angemessene absolute Strategie, bei der zunächst die Tonnamen bestimmt werden und dann das Intervall dazwischen berechnet wird, zurück. Bei einer weiteren Aufgabe sollten notierte und gehörte Melodien hinsichtlich des Intervallverlaufs verglichen werden (Miyazaki & Rakowski, 2002). Hier zeigten Absoluthörer bessere Leistungen bei nicht transponierten Melodien als bei transponierten Melodien, während die Leistungen von Nichtabsoluthörern von der (hier irrelevanten) Tonart unabhängig war. Ein ähnliches Ergebnis erhielt Miyazaki (2004) für den Vergleich zweier gehörter Melodien. Er schlussfolgert, dass Absoluthörer,

die nicht zwischen absoluter und relationaler Strategie „umschalten“ können, sowohl beim Hören von Musik als auch beim Musizieren benachteiligt sind. Anekdotische Berichte von absolut hörenden Chorsängern, die Probleme bekommen, sobald der Chor absinkt oder in alter Stimmung musiziert, bestätigen diese Annahme. Jedoch sei hier an das Ergebnis von Bischoff (2000) erinnert, die eine Reihe von Absoluthörern untersuchte, denen ein „Strategiewechsel“ durchaus möglich war.

Die anfangs beschriebene Unterscheidung von Repräsentations- und Benennungskomponenten des absoluten Gehörs veranlasste in jüngerer Zeit einige Forscher dazu, die Repräsentationskomponente alleine experimentell zu untersuchen. Im Prinzip läuft dies auf eine Untersuchung der Unabhängigkeit beider Komponenten hinaus, denn es wird gefragt, ob absolute Tonerinnerung ohne Tonbenennung (sogenanntes „latentes“ absolutes Gehör) möglich ist. Während die oben beschriebenen expliziten Aufgaben mit Einzeltönen dies bei Nichtabsoluthörern eher nicht nahe legen, wiesen Saffrans (2003) Ergebnisse auf eine überzufällige implizite Erinnerung an absolute Tonhöhen hin. Um auch die explizite Erinnerung an absolute Tonhöhen zu testen, wurde die Repräsentation bekannter Melodien untersucht. Welche Methoden dabei zur Anwendung kamen und zu welchen Ergebnissen führten, wird im folgenden Kapitel beschrieben. Da „latente“ Absolutrepräsentationen im Kontext der Frage diskutiert werden, ob sich absolutes Gehör statt als dichotom eher als kontinuierlich ausgeprägte Fähigkeit beschreiben lässt, wird diese Frage zuerst aufgegriffen.

3 Absolutes Gehör als Kontinuumsleistung

Bislang gibt es keine umfangreichen und repräsentativen empirischen Häufigkeitsschätzungen des absoluten Gehörs, die eine Antwort auf die Frage erlauben, ob eine dichotome Klassifikation in Absoluthörer vs. Nichtabsoluthörer eher zutrifft als eine kontinuierliche oder zumindest mehrstufige Beschreibung absoluter Hörleistungen. Da das absolute Gehör keinen ausreichend hohen prognostischen Wert beispielsweise zur Vorhersage musikalischer Höchstleistungen besitzt und andererseits die bloße Häufigkeitsschätzung absoluten Gehörs in der Allgemeinbevölkerung wissenschaftlich nicht interessant genug ist, hat bislang kein Forscher den hohen Aufwand einer größeren Repräsentativbefragung auf sich genommen. Kleinere Repräsentativbefragungen an einer musikalisch unausgelesenen Schüler-Stichprobe (n = 451, Altersbereich 14-17 Jahre) sowie einer Stichprobe von 80 Studierenden des Berner Musikkonservatoriums stammen von Hurni-Schlegel & Lang (1978). Unter den Schülern fanden sie eine annähernde Normalverteilung der Tonidentifikationsleistungen (bei Toleranz von Halbtonfehlern), deren Mittelpunkt nahe dem Niveau des Zufalls lag, so dass man dieses Ergebnis auch als Rateverteilung interpretieren kann (vgl. Vitouch, 2005). Das Leistungsniveau der Musikstudierenden lag insgesamt höher und ließ eine annähernde Gleichverteilung zwischen 40 und 100 % Trefferquote erkennen, ein Befund, der eher auf ein Kontinuumsmodell absoluten Gehörs als auf die dichotome Leistungsverteilung hinweist. Jedoch ist bei Hurni-Schlegel & Langs Methode die Nutzung relationaler Informationen nicht auszuschließen (verwandt wurden ausschließlich Töne einer Oktave). Eine deutlich zweigipflige Verteilung und somit klare Unterschiede zwischen Absoluthörern und Nichtabsoluthörern fand Barkowsky (1992), der 120 Musikstudenten und 10 Absoluthörer mit Wiedererkennungsaufgaben in verschiedenen Varianten testete und dabei auch die Antwortzeiten erfasste. Bimodale Verteilungen der Tonidentifikationsleistungen fanden auch Carroll (1975), Miyazaki (1988) und Riker (1946).

Die teilweise überlappenden oder zumindest nur graduell unterschiedlichen Leistungen von Absoluthörern und Nichtabsoluthörern treten nur bei *einer* der zur Messung absoluten Gehörs verwendeten Methode auf, nämlich bei der Tonidentifikation. Bei der Tonproduktion dagegen zeigen Absoluthörer deutlich konsistentere Leistungen als Nichtabsoluthörer (vgl. Kap. 2.2.2). Auch bei Gedächtnisaufgaben zeigen sich bei Behaltensintervallen von über einer Minute klare Leistungsvorteile der Absoluthörer gegenüber Nichtabsoluthörern: Letztere urteilen bei so langen Behaltensintervallen nur noch auf dem Zufallsniveau, während bei Absoluthörern kaum schlechtere Leistungen auftreten als bei kürzeren Behaltensintervallen (vgl. Kap. 2.2.3). Diese Befunde lassen sich zum einen dadurch erklären, dass es bei der Tonidentifikation, nicht aber bei der Tonproduktion oder bei Gedächtnisaufgaben, nur eine geringe Anzahl an Antwortalternativen gibt (bei Toleranz von Oktavverwechslungen: 12), so dass die Wahrscheinlichkeit, einen Ton durch Raten richtig zu benennen, bei Tonidentifikationsaufgaben wesentlich höher ist als bei den beiden anderen Aufgaben. Zum zweiten gibt es viele Absoluthörer, die Töne benennen, aber nicht produzieren können (Petran, 1932; Révész, 1946), wohingegen der umgekehrte Fall bisher nicht berichtet wurde. Produktions-

aufgaben scheinen somit für Absoluthörer schwieriger zu bewältigen zu sein als Tonidentifikationsaufgaben und damit auch trennschärfer bei der Klassifikation in Absolut- und Nichtabsoluthörer. Takeuchi & Hulse (1993) schließen aus dieser Datenlage, dass die Klassifikation in zwei Populationen gerechtfertigt scheint, weil substanzielle Unterschiede zwischen Absoluthörern und Nichtabsoluthörern bei der Tonproduktion, Tonerinnerung und auch im Tonbenennungstempo (Barkowsky, 1992) vorliegen. Trotzdem gibt es bei der Tonidentifikation Leistungsüberlappungen zwischen den Populationen, die darauf zurückzuführen sind, dass auch *innerhalb* der Population der Absoluthörer große Leistungsunterschiede vorliegen, namentlich hinsichtlich der Fehleranfälligkeit bei Tönen verschiedener Klangfarben, Oktavlagen oder schwarzer Klaviertasten. Takeuchi & Hulse schlagen eine Art Hierarchie absoluten Gehörs vor: Ein niedriges Absoluthör-Level liege vor bei Abhängigkeit der Tonidentifikationsleistungen von Klangfarbe, Oktavlage und Tonklasse sowie schlechten Produktionsleistungen, während bei hohem Absoluthör-Level geringere tonspezifische Effekte auftreten und gute Produktionsleistungen vorliegen.

Aufschlussreich für die Frage der Dimensionalität absoluten Gehörs und vor allem für die Frage nach den zugrunde liegenden kognitiven Prozessen könnte die Untersuchung des phänomenologischen Wahrnehmungsaspektes bei der Tonidentifikation sein (vgl. Vitouch, 2005). Absoluthörer können Töne vermutlich deshalb so schnell identifizieren, weil jeder Ton einen charakteristischen Klang zu haben scheint, der fest mit dem entsprechenden Tonnamen assoziiert ist. Hinsichtlich des Wahrnehmungsaspektes unterscheiden sich Absoluthörer deutlich von Nichtabsoluthörern, denn von letzteren wurde bisher keine unterschiedliche Wahrnehmungsqualität verschiedener Töne berichtet. Interessanterweise gelang es Brady (1970) zwar, im Erwachsenenalter ein sehr genaues absolutes Gehör zu erlernen, dass sich hinsichtlich der Fehlerrate und Reaktionszeiten nicht von dem „echter“ Absoluthörer unterschied, jedoch ohne eine Änderung in der Wahrnehmung von Tönen zu bewirken. Brady hatte offensichtlich eine stabile mentale Vergleichsskala etabliert, die ihm absolute Tonurteile ermöglichte. Allerdings hat er nicht für jeden Ton die Wahrnehmung einer bestimmten Tonqualität erlernt. Aufschlussreich wäre nun die Untersuchung dieses Wahrnehmungsaspektes bei „weniger perfekten“ Absoluthörern (die starke tonspezifische Unterschiede zeigen), und zwar sowohl bei falsch als auch bei richtig benannten Tönen. Empfinden diese für jeden vorgespielten Ton z.B. die „f-igkeit“ eines *f*'s, oder z.B. nur für Töne einiger vertrauter Klangfarben, nur für weiße Töne? Nachgewiesene längere Reaktionszeiten für Töne fremder Klangfarben weisen auf Vergleichsprozesse hin, sind aber gewissermaßen ein „indirektes“ Maß für das eigentlich interessierende subjektive Erleben bei der Tonidentifikation. Eine Abnahme an empfundener „Tonqualität“ mit abnehmender „Güte“ absoluten Gehörs wäre ein Hinweis auf ein „Empfindungskontinuum“ innerhalb der Population der Absoluthörer. Denkbar ist aber auch eine starke Empfindungsqualität sogar bei falsch benannten Tönen, die ein Hinweis dafür wäre, dass nicht unbedingt Vergleichsprozesse stattfinden, sondern eher größere (und somit weniger) Kategorien vorliegen, z.B. nur mit Ganzton- statt mit Halbtonauflösung. Vitouch (2005) schlägt als Grundlage der Tonklassen-Empfindung von Absoluthörern ein „Hebbianisches“ Modell vor, nach dem für die absolute Wahrnehmung und Erkennung be-

stimmter Töne Netzwerke existieren, die bei passendem Input aktiviert werden und einen passenden Bewusstseinseindruck erzeugen. Es wäre durchaus vorstellbar, dass bei Absoluthörern mit unterschiedlicher „Auflösung“ absoluten Gehörs eine unterschiedliche Anzahl solcher Netzwerke vorliegt. Bei einem Absoluthörer, der regelmäßig Benennungsfehler bei schwarzen Tasten macht, könnte ein absolutes Gehör mit Ganztongenauigkeit vorliegen, bei dem ein gehörtes *fis* kein eigenes *fis*-Netzwerk aktiviert, sondern entweder das *f*- oder das *g*-Netzwerk.

Zusammenfassend lässt sich feststellen, dass Befunde aus Tonproduktions- und Tonerinnerungsaufgaben sowie zur Geschwindigkeit der Tonidentifikation eine dichotome Einteilung in Absoluthörer und Nichtabsoluthörer nahe legen, dass es aber innerhalb der Absoluthörer-Population ein Leistungskontinuum bei der Genauigkeit der Tonidentifikation gibt. Aufschlussreich für die weitere Untersuchung der Dimensionalität absoluten Gehörs wären grössere Repräsentativbefragungen, bei denen absolutes Gehör mehrdimensional gemessen wird, d.h. unter Verwendung von Tonproduktion und Tonidentifikation sowie unter Berücksichtigung des Wahrnehmungsaspektes beim Hören oder Produzieren von Tönen.

3.1 Latentes absolutes Gehör

Als Argument für die Dimensionalität absoluten Gehörs wurden auch Befunde zum sogenannten latenten oder rudimentären absoluten Gehör angeführt (Vitouch, 2005). Damit werden relativ genaue Tonartproduktions- oder Wiedererkennungsleistungen bezeichnet, die den befragten (nicht absolut hörenden) Personen meist nicht bewusst sind. Es wird im folgenden häufiger von Ton*art*erinnerung die Rede sein, weil die Erinnerung an ganze Melodien untersucht wird. Durch die Angabe der (Anfangs-) Tonart sind die absoluten Tonhöhen der ganzen Melodie festgelegt, daher ist Tonarterinnerung gleichbedeutend mit absoluter Tonrepräsentation einer Melodie. Abzugrenzen ist die Tonarterinnerung jedoch von dem Erleben sogenannter Tonartencharaktere. Jene stammen aus einer musikalischen Epoche, in der Musikinstrumente noch nicht gleichstufig temperiert gestimmt waren, so dass sich die verschiedenen Tonarten hinsichtlich der Reinheit der Intervalle unterschieden. Typischerweise traten in Tonarten mit vielen Vorzeichen etwas verstimmtere Intervalle auf, was von Komponisten als Ausdrucksmittel genutzt wurde. So gilt beispielsweise D-Dur als hell und strahlend, während E-Dur oft mit Begräbnisszenen assoziiert wurde (für eine ausführliche Darstellung vgl. Auhagen, 1983). In der gleichstufig temperierten Stimmung dagegen sind alle Halbtöne exakt gleich groß, so dass sich Tonarten hinsichtlich der relationalen Tonhöhen nicht mehr unterscheiden. Dies bedeutet, dass die Wiedererkennung einer bestimmten Tonart anhand relationaler Merkmale nicht möglich ist, da Unterschiede nur in den absoluten Tonhöhen (d.h., Cis-Dur liegt insgesamt einen Halbton höher als C-Dur) bestehen. Eine Tonartencharakteristik, die eine Grundlage zur Tonarterkennung bildet, kann es folglich in einer gleichstufig temperierten Stimmung nicht geben (Jorgenson, 1991).

Um sicherzugehen, dass Tonarterkennung im Experiment tatsächlich auf der Beurteilung der absoluten Tonhöhen beruht, ist es daher wichtig, dass Transpositionen auf gleichstufig tem-

perierten Instrumenten und mit exakt gleicher Spielweise wie das Original hergestellt werden. Dies ist z.B. bei Einspielungen auf dem Klavier nicht selbstverständlich gegeben, da erstens nicht alle Klavierstimmer konsequent gleichstufig temperiert stimmen und zweitens beispielsweise Cis-Dur schwieriger zu spielen ist als C-Dur. Noch größere Probleme ergeben sich bei der Nutzung von Orchesterinstrumenten, die in verschiedenen Lagen (d.h. bei grösseren Transpositionen) etwas unterschiedliche Klangfarben haben. Außerdem ist auch dort die gleichstufig temperierte Stimmung nicht immer vollständig realisiert. Vermutlich ist diese Schwierigkeit ein Grund dafür, dass bisher keine Wiedererkennungsexperimente mit von Orchesterinstrumenten eingespielten Vergleichsmelodien vorliegen. Die Weiterentwicklung der Tontechnik lässt allerdings vermuten, dass digitale Transpositionen von eingespielten Melodien, die die genannten Probleme umgehen, bald verfügbar sein werden.

3.1.1 Wiedererkennung von Tonarten

Eine frühe Untersuchung zur Tonart-Erkennung stammt von Corso (1957), der interessanterweise das Ziel hatte, nachzuweisen, dass *keine* Tonarterinnerung vorliege und somit die klassische Lehre von den Tonartencharakteristika in einer konsequent gleichstufig temperierten Stimmung keine psychologische Realität habe. Alle folgenden Untersuchungen hatten das gegenteilige Forschungsanliegen, indem sie nachweisen wollten, dass eine Tonarterinnerung vorliege. Corso ließ alle 12 diatonischen Tonarten sowohl in Dur als auch in Moll in jeweils drei „Skalen“ auf einem frisch gestimmten Flügel einspielen: erstens als aufsteigende Tonleiter, zweitens als „zufällige“ Tonfolge der sieben diatonischen Stufen und drittens als Akkordfolge Tonika – Subdominante – Dominante – Tonika. Die Trials wurden vollständig randomisiert dargeboten, allerdings für alle Vpn in der gleichen Reihenfolge und ohne Störbedingungen zwischen den Trials. Obgleich es kein Feedback über die Richtigkeit der Tonartidentifikation gab, ist die Nutzung relationaler Strategien in diesem Design nicht ausgeschlossen, zumal das durchschnittliche ISI nur ca. 6 s dauerte und somit viel zu kurz für ein sicheres „Vergessen“ vorher gehörter Töne war. Corsos Stichprobe umfasste 27 Musikstudierende mit einer mittleren Instrumentalerfahrung von 8 Jahren. Es zeigte sich, dass die Tonarten am besten als Tonleiter, dann als Akkordfolge und am schlechtesten als Zufalls-Tonfolge benannt werden konnten. Bei der Genauigkeit der Tonbenennung gab es große interindividuelle Unterschiede sowie Unterschiede zwischen den verschiedenen Tonarten. C-Dur wurde in allen „Skalen“ am besten erkannt (als Tonleiter: in 45,7% der Fälle), wohingegen andere Tonarten (als Tonleiter: Des-Dur und As-Dur, als Akkordfolge: A-, Ges- und Es-Dur) nur auf dem Zufallsniveau erkannt wurden. Obgleich bei der Darbietung als Tonleiter oder Akkordfolge fast alle Tonarten sowohl in Dur als auch in Moll überzufällig gut benannt werden konnten, interpretiert Corso dieses Ergebnis als Hinweis dafür, dass *keine* Tonartencharakteristika erkannt werden können. Diese Interpretation beruht vor allem auf der Tatsache, dass die Zufalls-Tonfolgen, die nach Corso am ehesten realer Musik entsprechen, kaum überzufällig gut benannt werden konnten. Ob Zufalls-Tonfolgen wirklich eher der musikalischen Realität entsprechen als die omnipräsente Akkordfolge T-S-D-T, kann kritisch hinterfragt werden, aber auch methodisch ist an der Studie zu bemängeln, dass erstens durch die

separate Einspielung der Trials auf einem Flügel weder eine gleichtemperierte Stimmung noch eine völlig gleichartige Spielweise gesichert werden konnte und zweitens die Nutzung relationaler Antwortstrategien nicht ausgeschlossen werden kann. Außerdem handelt es sich bei allen von Corso genutzten „Skalen“ um relativ künstliche Stimuli, die nicht unbedingt aufschlussreich für die Tonarterkennung beim Hören realer Musik sind. Diese wurde von Terhardt und Mitarbeitern untersucht.

Terhardt & Ward (1982) ließen jeweils die ersten 5 s der 12 Präludien in Dur-Tonarten des Wohltemperierten Klaviers von J.S. Bach in verschiedenen Tonarten auf einem Flügel einspielen: Neben der Originaltonart nutzten sie Transpositionen um einen, 4, 6 oder 7 Halbtöne nach oben und nach unten. Die Wahl fiel gerade auf dieses Werk, weil dort als gesichert gelten kann, dass es nur in einer einzigen Version (nicht etwa in anderen Tonarten) gehört und somit gelernt wurde. Der Fokus von Terhardts Experimenten liegt damit weniger auf dem Versuch, das Vorliegen von Tonartencharakteren in relativ abstrakten Tonfolgen nachzuweisen, sondern vielmehr auf dem Versuch, das Lernen von absoluter Toninformation auch im Erwachsenenalter nachzuweisen (kaum ein „normaler“ Klavierschüler bewältigt vor dem als Ende der sensiblen Phase angegebenen 7. Geburtstag sämtliche Präludien des Wohltemperierten Klaviers!). Folglich verwendete er als Versuchspersonen (fast) nur solche mit musikalischer Grundbildung, denen das Werk zumindest vom Hören vertraut war. Um den Vpn eine Urteilsbasis zu geben, wurde die Originaltonart vor der Präsentation jeder Version angesagt, außerdem sahen die Vpn einen vereinfachten Partiturausschnitt, so dass sie sich die Originaltonart vor dem Hören jeder Version innerlich vorstellen konnten. Ihre Aufgabe bestand darin, für jeden gehörten Musikausschnitt anzugeben, ob er in der Originaltonart, zu hoch oder zu tief erklungen war. Die Ergebnisse wurden anhand eines „Recognition Score“ R dargestellt, der die Anzahl korrekt erkannter Originalversionen mit der Anzahl fälschlicherweise als original bezeichneter Transpositionen in Beziehung setzt und von 0 (= keine Wiedererkennung der Originaltonarten) bis 1 (= perfekte Wiedererkennung der Originaltonarten) variiert. Terhardt & Ward (1982) konnten zeigen, dass bei der Gegenüberstellung der Originalversion mit um 6 Halbtöne nach unten und nach oben transponierten Versionen R = 0,606 (n = 47 Vpn, ohne Berücksichtigung der 5 getesteten Absoluthörer: R = 0,561) beträgt, während beim Vergleich aller 9 Versionen „nur“ ein R = 0,484 (n = 22 Vpn, ohne Absoluthörer: R = 0,406) erzielt wurde. Beim Vergleich von Originalversionen nur mit Halbtontranspositionen sank R auf 0,248 (n = 10 Vpn). Aus dem Ergebnis, dass bei einer Abweichung um +/- 1 Halbton die relative Anzahl an „ok“-Antworten nur noch halb so groß war wie bei Darbietung der Originalversion, schließen die Autoren, dass die Schwelle zur Erkennung von Transpositionen bei Nichtabsoluthörern bei ca. 1,5 Halbtönen liegt, während sie bei Absoluthörern etwas niedriger, also bei ca. einem Halbton liegt.

Die berichteten Ergebnisse gelten jedoch zunächst nur für Klaviermusik. Um festzustellen, ob die Tonarterkennung tatsächlich auf der Erkennung absoluter Tonhöhen und nicht etwa auf der Wahrnehmung geringfügiger Unterschiede in Spieltechnik oder Klangfarbe beruhen, wäre eine Replikation des Experiments mit künstlichen Klangfarben und möglichst digital hergestellten Transpositionen erforderlich. Zumindest den Einfluss der Spieltechnik und der

Klangfarbe kontrollierten Terhardt & Seewann (1983) in einer weiteren Experimentalreihe, indem sie modifizierte Klaviertöne sowie Sinustöne nutzten. Die modifizierten Klaviertöne wurden hergestellt, indem jede Version einen Halbton zu tief und 6 % zu langsam eingespielt und das Tonband dann um 6 % beschleunigt wurde, um eine Erhöhung um einen Halbton zu erzielen. Somit wurde z.B. ein klingendes C-Dur in H-Dur eingespielt. Die Sinustöne wurden als Klangfarbe auf einem Digitalklavier gewählt, auf dem die Testexzerpte ein weiteres Mal eingespielt wurden. Sollten spieltechnische und klavierspezifische klangliche Aspekte für die Tonarterkennung maßgeblich sein, so sollte sich die Wiedererkennungsleistung mit diesem Tonmaterial verschlechtern. Dies war nicht der Fall: Der Vergleich der Leistung von 11 Vpn bei „normal“ eingespielten Klavierversionen (R = 0,28 bei der Beurteilung von Original, Halbton- und Großterztranspositionen) und modifizierten Klavierversionen (R = 0,34) sowie Sinuston-Versionen (R = 0,31) ergab keine signifikanten Unterschiede, was gegen einen Einfluss von Klangfarbe und Spieltechnik spricht.

In den R-Verteilungen von insgesamt 135 Vpn des Hauptexperiments (nur „normal“ eingespielte Klavierversionen) wurde eine Bimodalität der Leistungen sichtbar: Eine kleine Gruppe von Vpn, die vor allem aus Absoluthörern bestand, zeigte nahezu perfekte Wiedererkennungsleistungen, während eine größere Gruppe von Vpn, die vor allem aus Nichtabsoluthörern bestand, schlechtere, aber durchaus überzufällige Leistungen zeigte. Außerdem unterschieden sich die Fehlertendenzen zwischen Absolut- und Nichtabsoluthörern hinsichtlich der Abhängigkeit von der Originaltonart des Stückes (nur bei Nichtabsoluthörern vorhanden). Diese Ergebnisse interpretieren die Autoren als Hinweis auf unterschiedliche Strategien zur Aufgabenbewältigung. Absoluthörer können einzelne Töne bewusst kategorisieren und dadurch die Tonart erkennen, während Nichtabsoluthörer unbewusst eine größere Anzahl an Tönen nutzen, um ein Gefühl für die Tonart zu bekommen. Die Strategie der Nichtabsoluthörer ist fehleranfälliger, aber trotzdem zeigt sich insgesamt, dass es auch Nichtabsoluthörern möglich ist, überzufällig gut Transpositionen einer bekannten Melodie von der Originalversion zu unterscheiden. Als Ursache dieser Fähigkeit schließen die Autoren das Vorliegen von Tonartencharakteristika explizit aus, da die Wiedererkennung von der linearen Entfernung der Tonarten abhängig ist und nicht von der Verwandtschaft und somit Unterschiedlichkeit der Tonarten. Obgleich der Befund, dass die Wiedererkennungsgenauigkeit unabhängig von der genutzten Klangfarbe und Einspiel-Technik ist, interessant ist, bleiben als methodische Schwachstellen der Terhardt-Studien das Fehlen von Störbedingungen zwischen den Trials, durch die in Verbindung mit dem Vorhandensein der notierten Originalversionen relative Antwortstrategien begünstigt oder zumindest nicht zuverlässig verhindert werden konnten.

Einen methodisch verfeinerten Beweis für die auch Nichtabsoluthörern mögliche Unterscheidung zwischen Original und Transposition bei einem Präludium aus dem Wohltemperierten Klavier legten Vitouch & Gaugusch (2000) vor. Sie fertigten mithilfe des Transpositionsmodus eines Digitalklaviers eine mit Ausnahme der Tonart identische Replikation des C-Dur Präludiums in Cis-Dur an und wählten die Zeit als Störbedingung zwischen den Trials, die mit einem ISI von 24 Stunden stattfanden. Bei ihren 52 nicht absolut hörenden Vpn lagen insgesamt zu 59 % richtige Antworten vor, ein kleiner, aber signifikanter Effekt. Zusammen

mit den Ergebnissen von Terhardt und Mitarbeitern deutet dieser Befund darauf hin, dass bei Vpn mit musikalischer Vorbildung, die das untersuchte Werk zumindest vom Hören kennen, eine rudimentäre und den Vpn in den meisten Fällen nicht bewusste Tonarterinnerung vorliegt.

Eine Generalisierung der bisherigen Befundlage auf musikalisch untrainierte Hörer legten Schellenberg & Trehub (2003) vor. Sie wählten die Soundtracks von sechs bekannten TV-Serien als musikalisches Material sowie nur solche Vpn (48 überwiegend musikalisch untrainierte College-Studierende), denen alle sechs Serien vertraut waren. Musikausschnitte von 5 s Dauer wurden digital um einen und zwei Halbtöne nach oben und unten transponiert und den Vpn in zwei Sitzungen im Paarvergleich vorgespielt (Original vs. Transposition nach oben *oder* unten, wobei jede Vp von jeder Serie pro Sitzung nur eine Transpositionsrichtung und -größe hörte). Bei der Aufgabe, unter den beiden Versionen das Original zu benennen, erzielten die Vpn bei den Halbtontranspositionen 58 % und bei den Ganztontranspositionen 70 % richtige Urteile. Es werden keine Unterschiede bezüglich der Transpositions-Richtung berichtet, wohingegen es (innerhalb jeder Serie berechnet) immerhin einen kleinen Effekt der „lifetime viewing estimates“, also der Seh-Häufigkeit der Serie, auf die Wiedererkennungsrate dieser Serie zu geben schien. Da in einem Kontrollexperiment, bei dem 48 neue Vpn unter Original und Ganztontranspositionen von sechs unbekannten Musikstücken die richtige Version benennen sollten, nur Ergebnisse auf dem Zufallsniveau (49 % Richtigurteile) erzielt wurden, wird die überzufällige Erkennung der Originalversionen bei den TV-Serien auf deren Bekanntheit bei der getesteten Stichprobe zurückgeführt. Die Autoren interpretieren diesen Befund als Hinweis darauf, dass nicht absolut hörende und musikalisch untrainierte Erwachsene durchaus in der Lage sind, auch langfristig die absoluten Tonhöhen häufig gehörter Musik im Gedächtnis zu behalten. Nur mit einem Test wie dem hier vorgestellten, nicht aber mit einer für musikalische Laien bedeutungslosen Tonbenennungsaufgabe, werde diese den meisten Vpn unbewusste Fähigkeit sichtbar.

3.1.2 Produktion von Tonarten

Eine zweite Testmethode besteht darin, die Vpn vorgestellte Stückanfänge singen, summen, oder durch ein Musikinstrument produzieren zu lassen. Die oft zitierte Studie von Levitin (1994) lieferte bislang den überzeugendsten Nachweis für das durch Produktion gemessene Vorhandensein absoluter Tonrepräsentationen bei musikalischen Laien. Levitin ging von der Zwei-Komponenten-Hypothese absoluten Gehörs aus. Um die Verbreitung der Repräsentations-Komponente in der allgemeinen, musikalisch nicht vorgebildeten Bevölkerung zu untersuchen, ließ Levitin 46 musikalisch wenig trainierte College-Studierende aus einer Reihe von CDs zwei Lieder auswählen, die ihnen vertraut waren, und diese nacheinander ansingen. Um sicherzustellen, dass für alle genutzten Lieder nur eine einzige Originalversion existiert, wurden erstens im Rahmen einer Normierungsstudie nur solche Lieder ausgewählt, die den Status von Evergreens haben und zweitens solche Lieder aus der Auswahl ausgeschlossen, von denen Cover-Versionen in vom Original abweichenden Tonarten veröffentlicht wurden. Die Produktionen wurden hinsichtlich der Anfangstöne mit den Originalen verglichen. Es

zeigte sich, dass im ersten Trial 26 % der Vpn die Originaltonart sangen und 57 % der Vpn höchstens einen Halbton daneben lagen, während im zweiten Trial 23 % der Vpn die Originaltonart sangen und 51 % der Vpn höchstens einen Halbton daneben lagen. Während Levitin die Definition eines „Treffers“ auf eine Abweichung von bis zu 2 Halbtönen nach oben oder unten verbreitert und somit feststellen kann, dass 81 % der Vpn zumindest bei einem Trial einen „Treffer“ landen konnte, betonen Schellenberg & Trehub (2003, S. 265), dass man an den Daten auch ablesen könne, dass über die Hälfte von Levitins Vpn bei mindestens einem Trial die Töne nur mit der Genauigkeit des Zufallsniveaus produzierte. Die Interpretation der Daten hängt offensichtlich von der Betrachtungsseite ab. Will man wie Levitin nachweisen, dass die Repräsentations-Komponente absoluten Gehörs wesentlich weiter verbreitet ist als das „echte“ absolute Gehör, so ist dieser Nachweis gelungen: In beiden Trials produzierten jeweils über 50 % der Vpn *nicht* auf dem Zufallsniveau, sondern genauer. Dieser Prozentsatz liegt weit über der mit weniger als einem Prozent angegebenen Inzidenz „echten“ absoluten Gehörs. Andererseits lässt sich aus Levitins Daten nicht schließen, dass *alle* Vpn, oder generalisierend, alle Popmusik-Hörer, eine absolute Repräsentation aller gehörten Lieder aufweisen.

Zudem betonen Kritiker der Produktionsmethodik (Schellenberg & Trehub, 2003; Vitouch, 2005), dass singmotorische Einflüsse auf die Produktionsgenauigkeit nicht auszuschließen sind. Vitouch führt diesbezüglich ein Experiment von Wallace (1985) an, das nach einem Sensibilitätstraining für Stimmbandspannung und Kehlkopfstellung eine Ton-Identifikationsrate von 90 % vermelden kann. Levitin stellte diesem Argument (vorausschauend) Befunde von Ward & Burns (1978) und Murry (1990) entgegen (vgl. Kap. 4.4.5). Diese zeigen, dass das singmotorische Tonhöhengedächtnis auch bei Sängern nicht so genau ist, dass es die in seinem Experiment aufgetretene Produktionsgenauigkeit erklären könnte. Aufgrund der bisherigen Datenlage ist der relative Beitrag singmotorischer Gedächtnisinhalte zum Gedächtnis an absolute Tonhöhen nicht abschätzbar. Den Befund von Wallace (1985) relativierend ist allerdings anzumerken, dass die meisten von Levitin getesteten Vpn musikalische Laien waren. Kennt man den Namen des zu produzierenden Anfangstons nicht, so kann die ungefähre Lagenschätzung z.B. des Tons *c* für diese Aufgabe nicht von großem Nutzen sein. Der Beitrag motorischer Gedächtnisinhalte wird im empirischen Teil dieser Arbeit nochmals aufgegriffen und ein Untersuchungsdesign vorgestellt. Vorerst lässt sich aus Levitins Daten folgern, dass ein erheblicher Anteil seiner Vpn die Tonarten von Evergreens oberhalb der Ratewahrscheinlichkeit produzieren konnte.

In eine ähnliche Richtung deuten Daten von Halpern (1989), die bislang die einzige Untersuchung absoluten Tongedächtnisses vorlegte, in der Produktions- und Wiedererkennungsmethoden mit vier Experimenten verglichen wurden. Halpern untersuchte im Unterschied zu Levitin keine Lieder, für die es eine objektiv festgelegte Tonart gibt, sondern Volkslieder, die in vielen verschiedenen Tonlagen gesungen werden. Entsprechend konnte sie nicht die *Genauigkeit* der absoluten Repräsentationen messen, sondern die *Konsistenz* der Tonartwahl über mehrere Produktionen hinweg. In einem ersten Experiment ließ Halpern 22 musikalisch untrainierte und nicht absolut hörende Vpn (College-Studierende) in zwei Sitzungen (ISI

mindestens 48 h) sechs Volkslieder mehrmals ansingen. Dann verglich sie die Standardabweichungen der Anfangstöne der gleichen Lieder innerhalb der gleichen und zwischen zwei Sitzungen („same-within“ und „same-between“) sowie jene unterschiedlicher Lieder innerhalb der gleichen und zwischen zwei Sitzungen („different-within“ und „different-between“). Es ergaben sich signifikant kleinere Standardabweichungen bei gleichen Liedern (mittlere SD = 1,28 Halbtöne) als bei unterschiedlichen Liedern (mittlere SD = 2,52 Halbtöne), wobei es keinen Unterschied zwischen „within“ und „between“ gab; ein längeres ISI zwischen zwei Produktionen des gleichen Liedes führte also nicht zu variableren Produktionen. Dieses Ergebnis verweist auf das Vorhandensein absoluter Repräsentationen der getesteten Lieder.

In ihrem ersten Experiment hatte Halpern die Mitwirkung relationaler Hinweisreize durch Störtöne zwischen den Trials und das Kauen von Kaugummi (dessen Zuverlässigkeit bei der Verhinderung inneren Vokalisierens allerdings zu hinterfragen ist) zu minimieren versucht. Nach eigenen Angaben haben ihre Vpn jedes Trial ohne Referenz auf andere Trials bearbeitet. Um jedoch den Einfluss der Singmotorik näher zu beleuchten, führte Halpern ein weiteres Experiment durch. Dabei mussten die 39 Vpn (13 musikalisch ausgebildete und 26 musikalisch untrainierte Vpn) den imaginierten Anfangston der Volkslieder auf einem Keyboard mit verdeckter Tastatur suchen, wobei ansonsten die gleichen Maßnahmen zur Verhinderung der Nutzung relationaler Strategien getroffen wurden. Diese Prozedur führte zu insgesamt höheren Standardabweichungen. Bei beiden Gruppen wurden unterschiedliche Lieder „within“ und „between“ mit einer mittleren Standardabweichung von 4 Halbtönen produziert, während es bei gleichen Liedern zu kleineren Standardabweichungen kam (mittlere SD bei Musikern 2,36 Halbtöne, bei Nichtmusikern 2,2 Halbtöne). Dabei waren die Standardabweichungen gleicher Lieder sowohl bei Musikern als auch bei Nichtmusikern größer bei längerem ISI. Halpern folgert aus diesem Experiment, dass absolute Tonrepräsentationen der getesteten Lieder durch die gewählte Methode schwieriger zu externalisieren waren.

Um eine Aufgabe zu stellen, bei der weder gute Sänger noch gute Pianisten „im Vorteil“ bei der Externalisierung vorgestellter Tonhöhen sind, führte Halpern ein drittes Experiment durch. Hier mussten 29 Vpn (15 Musiker, 14 Nichtmusiker) zunächst vier Volkslieder ansingen. In einer zweiten Sitzung wurden die Vpn angewiesen, sich die Anfangstöne der gleichen Volkslieder vorzustellen. Daraufhin sollten sie in einer Art „probe tone“ Aufgabe (in Anlehnung an Krumhansl & Shepard, 1979) insgesamt 15 vorgespielte Anfangstöne (das Original sowie 1-7 Halbtöne nach oben und nach unten) hinsichtlich ihrer Ähnlichkeit mit ihrem imaginierten Anfangston beurteilen. Es ergaben sich höhere Ähnlichkeitsurteile für die in der ersten Sitzung gesungene Tonhöhe sowie für die kleine Terz und die Quinte aufwärts und die große Terz abwärts, wohingegen die Ähnlichkeitsurteile für alle anderen Töne niedriger waren. Halpern folgert aus diesem Ergebnis, dass die Vpn sich an ihren gewählten Lieblingston erinnern, aber auch andere Anfangstöne akzeptieren, jedoch nicht unbedingt solche in linearer Nähe zum gewählten Lieblingston, sondern eher solche im Terzabstand. Gerade die Ablehnung von Anfangstönen im Sekundabstand zum gewählten Lieblingston sei ein Hinweis auf die Absolutrepräsentation. Relativierend ist zu diesem Experiment allerdings zu

bemerken, dass die Varianz in den Ähnlichkeitsurteilen nicht sehr hoch war; auf einer Antwortskala von 1 bis 7 lag der niedrigste Mittelwert bei 3,8 und der höchste bei 4,8.

Schließlich sollte ein viertes Experiment zeigen, dass auch der Tonhöhenbereich, in dem sich Vpn ein Stück vorstellen können, systematisch mit dem bevorzugten Anfangston der Stücke zusammenhänge. Bei der Aufgabe, die höchsten und tiefsten Töne anzugeben, die die Vpn sich als Anfangstöne vorgegebener Volkslieder vorstellen können, zeigte sich allerdings, dass viele Lieder in einem ähnlichen Tonhöhenbereich präferiert werden. Es gab nur jeweils einen „Ausreißer“ nach oben („Joy to the world“) und nach unten („Row row row your boat“). Halpern (1989) interpretiert diesen Befund dahingehend, dass es keine allgemeingültige Frequenzgrenze gebe, ober- bzw. unterhalb derer man sich alle Lieder nicht mehr vorstellen könne, sondern dass diese Grenze vom imaginierten Lied abhänge. Obwohl die gefundene Ähnlichkeit der gewählten Tonhöhenbereiche sicherlich auch mit dem musikalischen Material (Volkslieder) zu erklären ist, deren Kennzeichen es ja gerade ist, dass sie in mittlerer (bequemer) Lage gesungen werden, macht das Ergebnis des letzten Experiments ein methodisches Problem der ganzen Experimentalserie deutlich. Dieses liegt in der alleinigen Analyse der Anfangstöne und nicht des Lied-Ambitus’. Ein durchaus denkbarer Mechanismus wäre es ja, dass alle Volkslieder in etwa gleicher Lage präferiert werden, die bequem innerhalb des eigenen Stimmumfangs liegt. Nicht jedes Volkslied beginnt aber auf dem gleichen Ton innerhalb des Lied-Ambitus’, sondern manche eher im oberen Bereich (z.B. „Joy to the world“), andere eher im unteren (z.B. „Row row row“). Ein bloßer Vergleich der Anfangstöne verschiedener Lieder könnte somit eine größere Tonart-Varianz suggerieren als der Vergleich des gewählten Ambitus’. Somit können die Daten von Halperns (1989) Experimenten nicht sicher als Hinweis auf absolute Tonartrepräsentationen bei Volksliedern gewertet werden.

Interessant ist in diesem Zusammenhang ein Eigenexperiment von Ward (1990), über das er in Ward (1999) berichtet. Ward hatte über ein Jahr hinweg alle spontanen Ohrwürmer in den Tonarten, die ihm in den Kopf kamen, aufgenommen und diese Tonarten mit denen eines gezielten Abrufs verglichen. Die Tonarten waren nicht zufällig verteilt, sondern konzentrierten sich in einem Bereich von 3-4 Halbtönen. Bei einer Analyse der Töne, die jeweils genau in der Mitte des Lied-Ambitus’ lagen, stellte er fest, dass die mittleren Melodietöne immer im Bereich *e* bis *g* lagen, wobei das *fis* am häufigsten vorkam. Da *fis* genau die Mitte von Wards Stimmumfang bildet, folgert er, dass sowohl motorische als auch sensorische Gedächtnisinhalte vorliegen können.

Bergeson & Trehub (2002) nutzten ebenfalls Halperns Methode, um die Tonart-Konsistenz von Wiegenliedern und Sprache bei jungen Müttern zu untersuchen. Sie fanden heraus, dass bei wiederholtem Singen des gleichen Liedes sich die Tonart (jeder einzelnen Vp) nur um weniger als einen Halbton unterscheidet, wohingegen die Unterschiede bei gesprochener Sprache mit ca. 5 Halbtönen wesentlich größer sind. Bergeson & Trehub (2002) analysieren zwar die Tonart der Lieder, statt nur der Anfangstöne, geben allerdings keine Auskünfte über

die Gesamtvarianz der Tonarten über alle Vpn. Somit ist auch hier nicht auszuschließen, dass alle Wiegenlieder von allen Vpn in sehr ähnlicher, mittlerer Lage gesungen werden.

Interessant ist der von Bergeson & Trehub (2002) vorgenommene Vergleich von gesungener und gesprochener Sprache, der zeigt, dass gesprochene Sprache viel stärker in der Tonart variiert als gesungene. Dies ist vermutlich damit zu erklären, dass die absolute Tonhöhe von Sprache im Englischen keine bedeutungstragende Rolle spielt: Ob z.B. eine Frage insgesamt etwas höher oder tiefer gestellt wird, ist weniger wichtig, als dass die Stimme sich am Ende des Satzes hebt. Die relationalen Tonhöhen sind also bedeutungstragend, und es fällt den meisten Menschen auf, wenn die Satzmelodie der Bedeutung des Satzes entgegenläuft. Zwar könnte man argumentieren, dass auch in der Musik relationale Merkmale wichtiger sind als absolute, denn Intervall- oder Rhythmusveränderungen werden auch von Laien bei ihnen bekannter Musik sofort bemerkt (z.B. Bartlett & Snelus, 1980), während Transpositionen schwieriger zu bemerken sind. Allerdings führen Bergeson & Trehub (2002) Befunde von Saffran & Griepentrog (2001) an, die zeigen, dass sich Babies beim Hören von Tonfolgen in stärkerem Maße an absoluten Tonhöhen orientieren als Erwachsene (vgl. 2.2). Wenn dies zutrifft, käme das von Bergeson & Trehub (2002) berichtete konstante Singen von Wiegenliedern in etwa gleicher Tonlage der Wahrnehmungspräferenz von Babies entgegen und wäre insofern funktional, als es der schnellen Wiedererkennung der Lieder dient. Andererseits könnte sich die im Vergleich zu Wiegenliedern in der Tonart variablere gesprochene Sprache möglicherweise auch durch die Unterschiede im Ambitus von Text und Liedern erklären. Da untrainierte Sänger meist nur in einem relativ kleinen Tonhöhenbereich bequem singen können, stehen z.B. für Lieder mit dem Tonumfang einer Oktave nur wenige Tonarten zur Verfügung. Demgegenüber können gesprochene Äußerungen mit einem kleineren Ambitus in vergleichsweise mehr Tonlagen bequem gesprochen werden. Auch Bergeson & Trehubs (2002) Daten liefern somit keinen eindeutigen Beweis dafür, dass die Sängerinnen von Wiegenliedern eine absolute Repräsentation der Lieder aufweisen.

3.1.3 Produktion gesprochener Sprache

Da nicht nur die Tonhöhenkonsistenz gesungener, sondern auch gesprochener Sprache ein Hinweis auf das Vorhandensein absoluter Tonartrepräsentationen sein könnte, sollen die vorhandenen Befunde hier berichtet werden. Allerdings gibt es für Sprecher „westlicher“ Sprachen bislang noch kaum Befunde, die auf die konsistente Nutzung ähnlicher Tonlagen beim Sprechen hinweisen. Immerhin fanden Church & Schacter (1994) mit einer Primingaufgabe, dass das implizite Gedächtnis für gesprochene Wörter von der Tonhöhe der Sprechstimme abhängig ist. Bei der Aufgabe, tiefpassgefilterte Wörter, die zur Hälfte vorher gehört und zur Hälfte neu waren, zu identifizieren, waren die Leistungen besser, wenn die Sprechstimme, die die Wörter vorlas, die gleiche war als bei unterschiedlichen Sprechstimmen. Diesen Unterschied gab es nicht bei der expliziten Wiedererkennungsaufgabe der Wörter. In einem weiteren Experiment wurde die Grundfrequenz der Sprechstimme manipuliert. Bei der Wort-Identifikation wurden Wörter, die von der gleichen Stimme mit gleicher Grundfrequenz vorgelesen wurden, besser identifiziert als solche, bei denen sich die Grundfrequenz der

gleichen Stimme zwischen Enkodierung und Primingaufgabe unterschied. Wieder gab es keinen solchen Leistungsvorteil bei einer expliziten Wiedererkennungsaufgabe der Wörter. Diese Befunde weisen darauf hin, dass die Grundfrequenz der Stimme beim Sprechen durchaus enkodiert und im „perceptual representation system“, das die Autoren für die Primingphänomene verantwortlich machen, gespeichert wird. Allerdings lassen sich aus dieser Studie keine Rückschlüsse auf die Lebensdauer solcher Repräsentationen ziehen.

Die Frage der langfristigen Absolutrepräsentation von Sprechstimmen wurde in jüngster Zeit bezogen auf Sprecher von Tonsprachen untersucht. In Tonsprachen, z.B. im Chinesischen, spielen die Tonlagen gesprochener Silben eine wichtigere Rolle als z.B. im Deutschen oder Englischen, da viele Silben in unterschiedlicher Intonation auch eine unterschiedliche Bedeutung haben. So gibt es im Chinesischen vier „Töne“, die sowohl durch ihr Register als auch durch ihre Tonhöhenkontur definiert sind. Diana Deutsch und ihre Mitarbeiter untersuchten vor diesem Hintergrund die Hypothese, dass Sprecher von Tonsprachen beim Sprechen der gleichen Silben über mehrere Sitzungen hinweg konsistent die gleichen Tonhöhen nutzen (Deutsch, Henthorn & Dolson, 2004). In einem ersten Experiment baten sie sieben musikalisch untrainierte vietnamesische Muttersprachler, eine Liste von vietnamesischen Wörtern in zwei Sitzungen (an verschiedenen Tagen) vorzulesen. Die Untersuchung der gesprochenen Tonhöhen ergab, dass diese sich zwischen den beiden Sitzungen bei vier Vpn nur um maximal einen Viertelton unterschieden und bei den restlichen drei Vpn nur um maximal 1,1 Halbtöne. In einem zweiten Experiment baten die Autoren 15 musikalisch untrainierte chinesische Muttersprachler, eine Liste von chinesischen Wörtern (in verschiedenen „Tönen“) in zwei Sitzungen (an verschiedenen Tagen) mehrmals vorzulesen. Die Untersuchung der gesprochenen Tonhöhen ergab, dass diese sich sowohl innerhalb als auch zwischen den beiden Sitzungen nur sehr wenig unterschieden. Jeweils über die Hälfte der Vpn produzierte Tonhöhen, die sich um maximal einen Viertelton unterschieden, während der maximale Tonhöhenunterschied zwischen den Sitzungen 1,5 Halbtöne betrug und innerhalb einer Sitzung 1,25 Halbtöne. Um zu untersuchen, ob auch die Sprecher von einer Intonationssprache, nämlich Englisch, zu so tonhöhenkonsistenten Lautäußerungen fähig sind, baten die Autoren in einem dritten Experiment 14 musikalisch untrainierte Amerikaner, eine Liste von englischen Wörtern in zwei Sitzungen (an verschiedenen Tagen) vorzulesen. Hier ergaben sich mit den chinesischen Vpn vergleichbare Leistungen beim Vergleich der innerhalb einer Sitzung gesprochenen Wörter. Diese unterschieden sich bei den amerikanischen Vpn um maximal 1,5 Halbtöne, wobei auch hier über die Hälfte der Vpn Unterschiede bis maximal einen Viertelton produzierte. Ein größerer Leistungsunterschied zwischen den chinesischen und amerikanischen Vpn fand sich jedoch beim Vergleich der Tonhöhen zwischen den Sitzungen; hier wurden zwar auch von den Amerikanern nur maximale Tonhöhenunterschiede von 1,5 Halbtönen erzielt, aber die durchschnittlichen Unterschiede waren signifikant größer als bei den Chinesen. Deutsch et al. (2004) interpretieren diese Datenlage als Hinweis auf ein genaueres absolutes Tongedächtnis bei den Sprechern von Tonsprachen als bei den Sprechern von Intonationssprachen, da absolute Tonhöhen für erstere eine größere Bedeutung haben als für letztere. Trotzdem bleibt festzuhalten, dass auch die Amerikaner mit maxima-

len Tonhöhenunterschieden von 1,5 Halbtönen deutlich tonhöhenkonsistenter gesprochen haben als Bergeson & Trehubs (2002) Vpn.

3.1.4 Das Tritonus-Paradox als Hinweis auf latentes absolutes Gehör

Einen gewissermaßen noch *indirekteren* Hinweis auf das langfristige Vorhandensein rudimentärer absoluter Tonrepräsentationen liefert das von Deutsch und Kollegen beschriebene Tritonus-Paradox (Deutsch, 1986; Deutsch, 1991; Deutsch, Kuyper & Fisher, 1987; Deutsch, North & Ray, 1990). Sie nutzten für ihre Experimente Shepard-Töne, also Töne mit identifizierbarer Tonklasse (*c, d, g*, etc.), aber nicht identifizierbarer Oktavlage (*c², c³,* etc.). Shepard-Töne entstehen durch die Schichtung von mehreren Sinuskomponenten im Oktavabstand, wobei die Lautstärke der mittleren Sinuskomponenten am höchsten ist und nach oben und unten abfällt. Spielt man z.B. wiederholt die gleiche aufsteigende Tonleiter, so entsteht dadurch der Eindruck einer beständig steigenden Tonfolge, die jedoch niemals die Hörschwelle erreicht. Beim Vergleich von zwei Shepard-Tönen wird die Richtung des entstehenden Intervalls (auf- oder absteigend) allein durch den Abstand der Töne im Tonklassenkreis bestimmt: *cis-d* wird z.B. immer als aufsteigende Sekunde, nie als absteigende Septime gehört (Shepard, 1964).

Interessant ist nun die Wahrnehmung zweier Shepard-Töne im Abstand eines Tritonus, da dieses Intervall sowohl aufsteigend als auch absteigend wahrgenommen werden kann (der Tritonus teilt die Oktave in der gleichtemperierten Stimmung genau in zwei Hälften). Deutsch (1986) fand aber gerade keine zufällige Zuordnung der Intervallrichtungen bei ihren Vpn, sondern statt dessen, dass die Richtungsurteile erstens systematisch von der Position der Töne im Tonklassenkreis abhängen und dass es zweitens interindividuelle Unterschiede in der Wahrnehmung der Intervallrichtung gab. In einer größeren Befragung von 29 nicht absolut hörenden Vpn fanden Deutsch et al. (1987) wiederum systematische Bevorzugungen von „aufwärts"-Intervallen bei einigen Tonklassen, während es bei anderen Tonklassen zu systematischen Bevorzugungen von „abwärts"-Intervallen kam. Dies sehen Deutsch et al. (1987) als Hinweis auf das Vorhandensein absoluter Tonrepräsentationen in der musikalisch untrainierten Allgemeinbevölkerung: Die konsistente Beurteilung z.B. des Intervalls *c-fis* als aufwärts, *f-h* dagegen als abwärts sei nur durch eine Art subjektiven Tonhöhenkreis in dem Sinne, dass z.B. *f* der höchste Ton einer subjektiven Skala sei, zu erklären.

Zusammenhänge mit dem Ausmaß musikalischer Expertise fanden Deutsch et al. (1987) nicht, dagegen wurde von Deutsch (1991) ein interessanter Zusammenhang zur Sprache festgestellt. Sie fand, dass sich die subjektiven Tonhöhenkreise von Kaliforniern und Südengländern deutlich unterschieden, denn während bei den Kaliforniern die höchste Tonklasse *d* war, war sie bei den untersuchten Südengländern das *g*, also etwa eine halbe Oktave entfernt. Da sie als Korrelat des subjektiven Tonhöhenkreises den (auch interkulturell bzw. zwischen verschiedenen Dialekten der gleichen Sprache variierenden) individuellen Stimmumfang der Vpn vermuteten, setzten Deutsch et al. (1990) das Oktavband, in das die meisten Tonhöhenwerte beim spontanen Sprechen von 15 Vpn fielen, mit der jeweiligen indivi-

duellen Positionierung des Tonklassenkreises beim Tritonus-Paradox-Test in Beziehung. Sie fanden signifikante Zusammenhänge. Genau genommen widerspricht Deutsch (1991) mit diesem letzten Befund ihrer eigenen Interpretation des Tritonus-Paradox' als Hinweis auf rudimentäres absolutes Gehör in der Allgemeinbevölkerung, denn der Zusammenhang zur Sprechstimme weist ja gerade auf einen *relationalen* Mechanismus hin. Offensichtlich hat ein Großteil der untersuchten Vpn ein sehr genaues Gedächtnis für ihre bevorzugten Stimmlage und im Verhältnis zur bevorzugten Stimmlage dann auch ein sehr genaues Gedächtnis für eine als am höchsten empfundene Tonklasse. Inwieweit diese Formen des Gedächtnisses, die vermutlich stark auf motorischen Gedächtnisinhalten basieren, tatsächlich als Hinweise auf absolute Tonhöhenrepräsentationen gewertet werden können, ist allerdings fraglich.

3.2 Absolutes Gehör als Kontinuumsleistung?

Untersuchungen mit verschiedenen Methoden haben gezeigt, dass Nichtabsoluthörern unter bestimmten Umständen der Abruf von Tonarten gelernter Lieder möglich ist. Je nach Methode ergeben sich jedoch Konfundierungsprobleme, die eine Interpretation der Befunde hinsichtlich der Dimensionalitätsfrage absoluten Gehörs erschweren.

Die methodische Herausforderung bei Wiedererkennungsaufgaben zur Messung absoluter Melodierepräsentationen liegt in der Herstellung von Transpositionen, die hinsichtlich relationaler und klanglicher Aspekte identisch sind, sowie in einer Darbietung, die die Nutzung relationaler Hinweisreize verhindert. Bei weitgehender Kontrolle dieser Faktoren zeigt sich, dass eine Erinnerung an die Originalversionen von Melodien vorliegt, die den Vpn vertraut sind. Diese Erinnerung ist jedoch weitgehend unbewusst und bei Nichtabsoluthörern weniger zuverlässig als bei Absoluthörern.

Bei der Nutzung von Produktionsmethoden scheint das Paradigma von Levitin (1994), das einen Vergleich gesungener Melodien mit Originalversionen beinhaltet, zu eindeutiger interpretierbaren Ergebnissen zu führen als die Messung der Konsistenz produzierter Lieder ohne festgelegten Tonartstandard. Der Vorteil des Levitin-Paradigmas liegt vor allem darin, dass man weiß, welche Tonart gelernt wurde und folglich repräsentiert sein könnte. Bei allen Konstanzmethoden dagegen wurde bisher eine mögliche Konfundierung der Tonartkonstanz durch eine konstant bevorzugte Sprech- und Singlage zu wenig beachtet. Dass ein solcher Zusammenhang bestehen könnte, deuten die Daten von Wards (1990) Eigenexperiment an. Auch die beschriebenen Experimente von Deutsch und Mitarbeitern verweisen darauf, dass eine bevorzugte Sprechlage relativ stabil und gut erinnerbar ist und auch zur Orientierung für tonale Aufgaben dienen kann (siehe Tritonus-Paradox). Unbedingt erforderlich erscheint daher bei Nutzung der Konstanzmethode die Erfassung des Stimmumfangs. Daneben stellt sich aber auch die Frage, inwieweit Levitins (1994) Befunde auf andere Vpn sowie andere Melodien generalisierbar sind und inwiefern auch bei dieser Methode motorische Gedächtnisinhalte eine Rolle beim Melodieabruf spielen.

Die Repräsentation von gesprochener Sprache bzw. der absoluten Tonhöhen vertrauter Sprechstimmen wurde bisher nicht sehr häufig untersucht, da diese zumindest in westlichen

Sprachen nicht relevant sind. Während es Hinweise dafür gibt, dass eine implizite Erinnerung an absolute Tonhöhen gesprochener Sprache vorliegt, ist die Befundlage bei expliziter Erinnerung uneindeutig (Deutsch et al., 2004 vs. Bergeson & Trehub, 2002). Eindeutiger dagegen ist sie bei Tonsprachen, bei denen Tonhöhen bedeutungstragend sind und überzufällig gut erinnert werden.

Vor dem Hintergrund dieser Befunde sowie der im vorigen Kapitel zusammengefassten Erkenntnisse zum absoluten Gehör stellt sich die Frage, inwieweit die durch Wiedererkennung, Produktion oder implizite Methoden gemessene Erinnerung an die Tonarten gelernter Melodien oder Tonfolgen als abgeschwächte bzw. latente Form absoluten Gehörs angesehen werden und somit eine Kontinuumshypothese absoluten Gehörs stützen kann. Anders ausgedrückt kann man fragen, ob die Erinnerung von Nichtabsoluthörern an die Tonarten gelernter Lieder mit der Repräsentationskomponente absoluten Gehörs gleichzusetzen ist. Diese Gleichsetzung ist im Prinzip gleichbedeutend mit der Annahme einer weitgehenden Unabhängigkeit von Repräsentations- und Benennungskomponenten absoluten Gehörs, da bei Nichtabsoluthörern im Allgemeinen keine gute Benennung produzierter Töne vorliegt.

Vergleicht man die Erinnerung an Einzeltöne mit der Erinnerung an Tonarten gelernter Lieder, so ergeben sich deutliche Unterschiede bei Absolut- und Nichtabsoluthörern: Absoluthörer können einzelne Töne *ohne* Kontext genau produzieren und sich an einzelne gehörte Töne *mit und ohne* Kontext (z.B. Interferenztöne) genau erinnern. Die Erinnerung an Tonarten scheint auf dem Abruf einzelner Töne zu beruhen. Nichtabsoluthörer dagegen können einzelne Töne nur *mit* Kontext, z.B. unter Zuhilfenahme relationaler Hinweisreize, einigermaßen genau produzieren und sich an einzelne gehörte Töne nur *ohne* Kontext (also ohne Interferenztöne) genau erinnern. Die Erinnerung an Tonarten scheint weniger auf dem Abruf einzelner Töne als auf der Orientierung an dem gesamten vorliegenden tonalen Kontext zu beruhen. Bei der Benennung einzelner Töne zeigen sich sowohl bei Absolut- als auch bei Nichtabsoluthörern tonspezifische Unterschiede (Marvin & Brinkman, 2000), jedoch auf unterschiedlichem Leistungsniveau. Außerdem unterscheiden sich Fehlergrößen und Antwortlatenzen bei der Tonbenennung bei Absolut- und Nichtabsoluthörern. Auch der Umgang mit Transpositionen unterscheidet sich zwischen Absolut- und Nichtabsoluthörern: Während viele Absoluthörer durch Transpositionen erheblich gestört werden, bemerken Nichtabsoluthörer Transpositionen zwar im experimentellen Setting mit einiger Wahrscheinlichkeit, fühlen sich im musikalischen Alltag durch Transpositionen bis zu einer gewissen Grenze aber nicht gestört.

Dieses Befundmuster verweist auf grundsätzlich unterschiedliche Strategien bei der Aufgabenbewältigung durch Absolut- und Nichtabsoluthörer. Absoluthörer, denen stabile interne Tonkategorien als Standards im LZG zur Verfügung stehen, können Töne unter anderem *aufgrund* der verbalen Etikettierung trotz Interferenz im Gedächtnis behalten und ohne Kontext abrufen. Das verweist darauf, dass Repräsentations- und Benennungskomponente absoluten Gehörs *nicht* als unabhängige Komponenten zu betrachten sind. Der Abruf von Tönen ist insofern absolut, als keine Vergleichsprozesse stattfinden, sondern beim Hören eine

bestimmte Tonqualität empfunden wird. Daher ist der Abruf sehr schnell möglich. Auch die häufigen Oktavfehler von Absoluthörern sind ein Hinweis darauf, dass sich Absoluthörer weniger an der Helligkeit als an der Tonqualität gehörter Töne orientieren. Nichtabsoluthörer dagegen können einzelne Töne nur anhand von relationalen Hinweisreizen (z.B. Stimmumfang) oder aufgrund eines relativ ungenauen Gefühls für die ungefähre Lage von Tönen abrufen, daher ist der Abruf langsamer als der von Absoluthörern. Dass die Fehlerhäufigkeit bei der Tonbenennung mit der Fehlergröße abnimmt, ist ebenfalls ein Hinweis darauf, dass sich Nichtabsoluthörer vor allem am Helligkeitsaspekt gehörter Töne orientieren.

Das Vorhandensein unterschiedlicher Strategien bei Absolut- und Nichtabsoluthörern beim Abrufen von Tönen oder Tonarten aus dem LZG ist ein Hinweis darauf, dass die Tonarterinnerung von Nichtabsoluthörern weniger auf einer abgeschwächten Form absoluten Gehörs beruht als vielmehr auf anderen Gedächtnismechanismen. Ohnehin stellt sich die Frage, inwieweit eine Gedächtnisform „etwas“ oder gar „relativ“ absolut sein kann. Die Terminologie des „latenten“ absoluten Gehörs wurde vermutlich durch die vorrangige Betrachtung des Verhaltensergebnisses, nämlich der relativ genauen Produktion von Tonarten, motiviert. Der Gedächtnisprozess, der den genauen Tonartabruf ermöglicht, scheint aber gerade nicht *absolut* zu sein, sondern auf der Nutzung verschiedener Informationsquellen zu beruhen, zu denen die Nutzung relationaler Gedächtnisinhalte, die klangliche Erinnerung, die motorische Erinnerung wie auch die Erinnerung an andere Kontextreize zählen können.

Demgegenüber könnte ein wahrhaft „latentes“ absolutes Gehör im Sinne einer unbewussten und von der Benennung unabhängigen Erinnerung an einzelne absolute Tonhöhen ohne Kontext dann nachgewiesen werden, wenn mithilfe von Gedächtnisaufgaben (vgl. Kap. 2.2.3) der Nachweis einer überzufälligen Erinnerung bei Menschen ohne Notenkenntnis gelänge. Eine solche Fähigkeit nennen Ross, Olson, Marks & Gore (2004, S. 1798) „ability to perceptually encode” (APE), also sinngemäß „absolute Enkodierung”. Gelänge der Nachweis von APE-Besitzern in der musikalisch nicht ausgebildeten Allgemeinbevölkerung durch eine größere Repräsentativbefragung, so wäre dies ein Hinweis auf ein von der Tonbenennung unabhängiges Tongedächtnis. Von absoluter Tonerinnerung könnte man dann sprechen, wenn es tatsächlich gelänge, die Nutzung relationaler Hinweisreize zu verhindern. Ein solcher Nachweis steht bislang noch aus.

Obwohl hier angenommen wird, dass die Tonerinnerung von Absoluthörern und die Tonarterinnerung von Nichtabsoluthörern unterschiedliche Gedächtnisprozesse beanspruchen, stellt sich trotzdem die Frage, ob beiden Prozessen wiederum ähnliche Lernmechanismen zu Grunde liegen. Für Absoluthörer wurde als möglicher Lernmechanismus postuliert, dass sie in früher Kindheit die Assoziation bestimmter Klänge mit Namen dadurch lernen, dass sie einen Ton gleichzeitig spielen und hören. Diese Assoziationen sind so gut gelernt, dass durch Darbietung entweder eines Tonnamens oder eines Tons die jeweils andere Komponente des Assoziationspaares automatisch abgerufen wird. Bei Nichtabsoluthörern könnte das Lernen von Tonarten ebenfalls darauf beruhen, dass sie eine Melodie häufig in der gleichen Tonart hören, spielen oder singen und dadurch eine Assoziation Tonart-Melodie ent-

steht. Da eine Assoziation von Klängen mit Namen (auch ein Melodietitel ist ja nichts weiter als ein Name) im Erwachsenenalter weniger gut erlernbar ist als in der Kindheit, sind die Tonart-Melodie Assoziationen weniger automatisch abrufbar und störbarer als die Ton-Namen Assoziationen von Absoluthörern. Das häufige Hören von Transpositionen der gleichen Melodie scheint sich auf den Aufbau absoluter Tonartrepräsentationen im Erwachsenenalter ähnlich störend auszuwirken wie das häufige Hören von gleichen Tönen auf verstimmten Instrumenten den Erwerb eines absoluten Gehörs stört. Kurz zusammengefasst lässt sich also postulieren: Die Lernprozesse, die zum Aufbau von Ton-Namen sowie von Tonart-Melodie Assoziationen führen, ähneln sich. Einflussreich könnte sowohl die Hörhäufigkeit als auch das seltene Hören von Versionen in anderen Tonarten (Stimmungen) sein. Da das Lernen aber in unterschiedlichen Altersstufen stattfindet, unterscheidet sich das Lernergebnis, also die Stabilität der Assoziationen. Daher müssen Nichtabsoluthörer beim Tonabruf und Tonartabruf auf andere Strategien zurückgreifen als Absoluthörer.

Für Absoluthörer wurde der Effekt der „Assoziationsgüte durch Hörhäufigkeit“, anhand der in Kap. 2 berichteten tonspezifischen Effekte bereits häufiger nachgewiesen. In dieser Arbeit wird ein weiterer Nachweis dieses Effekts unter Nutzung einer neuen Methode, der Pupillometrie, vorgenommen, die in Kap. 8 näher beschrieben wird. Für Nichtabsoluthörer dagegen wurde ein Hörhäufigkeits-Effekt für Tonartrepräsentationen bisher nicht untersucht. Die Untersuchung eines solchen Effekts bildet daher einen wichtigen Teil dieser Arbeit.

Während sich die Untersuchung eines Hörhäufigkeits-Effektes auf die Genauigkeit von Tonartrepräsentationen theoretisch aus Befunden der Musikpsychologie zum absoluten Gehör ableiten lässt, stellt sich für die weitere Untersuchung von Tonartrepräsentationen die Frage, welche Hypothesen sich aus Befunden der Gedächtnispsychologie ableiten lassen. Dieser Frage wird sich das folgende Kapitel widmen.

4 Musik in der Gedächtnisforschung

In diesem Kapitel sollen Befunde aus der gedächtnispsychologischen Literatur zusammengefasst werden, die für Speicherung, Abruf und Vergessen musikalischer Information, insbesondere absoluter Toninformation, im LZG relevant sein könnten. Die Darstellung der grundlegenden Gedächtnisprozesse stützt sich im Wesentlichen auf die Übersichtsarbeiten von Squire & Kandel (1999), Buchner & Brandt (2002) und van der Meer (im Druck). Auf eine Diskussion der sensorischen Register wird hier verzichtet, da es in dieser Arbeit mit der Tonbenennung von Absoluthörern und dem Tonartabruf von Absolut- und Nichtabsoluthörern vorrangig um LZG-basierte Prozesse geht. Kurzfristiges Erinnern wird im Zusammenhang mit Befunden zum Übergang von Information aus dem KZG ins LZG diskutiert. In Abschnitt 4.4.5 wird außerdem im Zusammenhang mit der Diskussion motorischer Gedächtnisinhalte die Konzeption des Arbeitsgedächtnisses von Baddeley (1986), insbesondere seine Unterteilung der phonologischen Schleife, aufgegriffen, da es dort um ein Zusammenwirken von innerem Hören und Subvokalisation beim Melodieabruf geht.

4.1 Kategorisierung von Gedächtnissystemen

Obgleich es im Detail einige Kritikpunkte an modalen Gedächtnismodellen, z.B. dem Informationsverarbeitungsmodell von Broadbent (1958) oder dem Mehrspeichermodell von Atkinson & Shiffrin (1968) gibt, werden modale Modelle zur Schaffung eines Überblicks über Funktionsweisen des Gedächtnisses nach wie vor als hilfreich erachtet (Buchner & Brandt, 2002). Unterschieden wird dabei zwischen sensorischen Registern, dem Kurzzeit- oder Arbeitsgedächtnis und dem Langzeitgedächtnis. Die sensorischen Register, die eintreffende Information reiznah verarbeiten und für eine in verschiedenen Modalitäten unterschiedliche, insgesamt aber sehr kurze Zeit im Millisekunden- (auditiv: Sekunden-) bereich zur Verfügung halten, werden als Schnittstelle zwischen Wahrnehmung und Gedächtnis betrachtet. Im Arbeitsgedächtnis wird eine begrenzte Informationsmenge für die aktive Manipulation, z.B. den Vergleich von Tönen, für eine begrenzte Zeit zur Verfügung gehalten. Im LZG dagegen wird Information dauerhaft und vermutlich ohne Kapazitätsbeschränkung gespeichert. Auf Musik bezogen ist das Arbeitsgedächtnis notwendig, um aufeinander folgende Töne innerhalb eines gewissen Zeitfensters zu einer Melodie zu integrieren. Das LZG ermöglicht beim Hören ganzer Stücke eine (zumindest ungefähre) Erinnerung an bereits gehörte Motive und dadurch z.B. die Wahrnehmung musikalischer Form.

Die Trennung von LZG- und arbeitsgedächtnisbasierten Prozessen wird durch neuropsychologische Ergebnisse gestützt. Einige Berühmtheit erlangte dabei der von Milner über lange Zeit untersuchte Patient H.M. (zuerst berichtet in Scoville & Milner, 1957), dem zwecks Linderung seiner epileptischen Anfälle beidseitig Teile des Temporallappens inklusive Hippocampus entfernt worden waren. Dies führte dazu, dass er sich an Ereignisse seiner Jugend erinnern konnte, nicht aber fähig war, neue Erlebnisse ins LZG zu überführen. Neue Informationen blieben also nur für die kurze Zeit, in der seine Aufmerksamkeit darauf fokussiert war,

in H.M.s Gedächtnis. Beim Erwerb neuer Fertigkeiten, wie z.B. dem Nachzeichnen einer Figur, die nur im Spiegel gesehen wird, machte H.M. normale Fortschritte, ohne sich jedoch an einen früheren Lösungsversuch der gleichen Aufgabe erinnern zu können.

Eine grundsätzliche Unterscheidung im LZG betrifft die zwischen deklarativem und nichtdeklarativem Wissen (Squire, 1986). Dabei wird mit deklarativem Wissen Tatsachenwissen bezeichnet, über das Vpn explizit Auskunft geben können („wissen, dass"), nichtdeklaratives Wissen dagegen ist nicht verbalisierbar, sondern äußert sich im Verhalten eines Menschen („wissen, wie"). Angesichts des beschriebenen Patienten H.M. lässt sich die Trennung der beiden Wissensformen bestätigen, denn bei ihm kam es zur selektiven Störung deklarativen Wissens ohne Beeinträchtigung des nichtdeklarativen Wissens.

Innerhalb des deklarativen Gedächtnissystems unterscheidet Tulving (1972) zwischen episodischem und semantischem Gedächtnis. Dabei bezeichnet episodisches Gedächtnis die Erinnerung an konkrete Episoden wie z.B. den ersten Schultag oder den Urlaub des vorigen Jahres. Da räumliche und zeitliche Aspekte der Episoden erinnert werden, wird das episodische Gedächtnis auch als Quellengedächtnis bezeichnet. Semantisches Gedächtnis dagegen bezeichnet Sachwissen ohne eine Einbettung in konkrete Episoden. Dazu zählt beispielsweise das Wissen um historische Ereignisse oder die Kenntnis der Vornamen von Komponisten.

Nichtdeklarative Gedächtnisformen sind vielfältig. Sie umfassen assoziatives und nicht-assoziatives Lernen, Wahrnehmungslernen, Priming, emotionales Lernen sowie den Erwerb von Fertigkeiten und Gewohnheiten. Die verschiedenen Formen nichtdeklarativen Lernens treten auch bei vielen Tieren auf, und sie sind zuverlässig und andauernd. Durch sie entstehen Gewohnheiten, Dispositionen und Präferenzen, die dem Bewusstsein nicht zugänglich sind, aber einen wichtigen Teil der Persönlichkeit ausmachen.

4.2 Deklaratives Gedächtnis: episodisches und semantisches Wissen

Die Unterscheidung von episodischem und semantischem Wissen innerhalb des deklarativen Gedächtnisses wurde zwar z.B. von Schacter & Tulving (1994) auch auf neuronaler Ebene bestätigt, andererseits aber von Fuster (1995) und Buchner & Brandt (2002) als verzichtbare Unterscheidung kritisiert. Bei der Untersuchung des Gedächtnisses für Musik spielte die Unterscheidung zwischen episodischen und semantischen Gedächtnisinhalten in der Forschungsliteratur zwar eine wichtige Rolle (z.B. Nauck-Börner, 1984, 1988; Vazan, 2000), gleichzeitig wird z.B. bei der Betrachtung von Melodien deutlich, dass sie eigentlich eine Mischung episodischen und semantischen Wissens darstellen. Zunächst ist die Melodie eine Episode mit definierter räumlicher (Intervalle, Kontur, Tonart) und zeitlicher (Metrum, Tempo) Struktur. Bei wiederholtem Hören einer Melodie mit veränderter räumlicher oder zeitlicher Struktur entsteht semantisches Wissen im Sinne einer von Tonart und Tempo abstrahierten Repräsentation der Melodie. Diese ermöglicht z.B. das Wiedererkennen von Melodien, die in veränderten Tonarten, Klangfarben oder Tempi gehört werden. Andererseits gibt es bei Melodien die Besonderheit, dass sie z.B. beim Chorsingen oder beim Hören von Tonträgern

sehr häufig in gleicher oder zumindest sehr ähnlicher Form (hinsichtlich räumlicher und zeitlicher Struktur) gehört bzw. gesungen werden. Dies könnte dazu führen, dass zwar semantisches Wissen im Sinne einer von einzelnen Lernepisoden abstrahierten Melodie entsteht, aber auch die raum-zeitliche Struktur der Melodien langfristig verfügbar ist.

Eine Vielzahl von Experimenten widmet sich der Frage, inwieweit Musik „hierarchisch“ oder „reduziert“ im Sinne semantischen Wissens repräsentiert wird bzw. inwiefern auch „Oberflächeninformation“ langfristig erinnert werden kann. Unterschiedliche Methoden führen dabei grundsätzlich zu etwas unterschiedlichen Ergebnissen. Zum einen gibt es die Suche nach Regelsystemen, aufgrund derer gehörte Melodien „zusammengefasst“ werden können. Diskutiert werden sowohl harmonische (z.B. Bigand, 1990) als auch zeitliche (z.B. Large, Palmer & Pollack, 1995; siehe auch Lerdahl & Jackendoff, 1983) Reduktionsprinzipien. Im Experiment sind solche Reduktionsprinzipien z.B. dadurch untersuchbar, dass Stücke mit gleichem Harmonieschema, aber im Detail etwas unterschiedlicher Melodieführung dargeboten werden, ähnlich wie es in realer Musik beim Thema mit Variationen geschieht. Die Tatsache, dass auch leicht variierte Formen eines Grundthemas als äquivalent beurteilt werden, deutet auf eine auf die Haupttöne reduzierte Repräsentation der vorher unbekannten Melodien hin. Andererseits lässt sich z.B. beim Hören atonaler Melodien, für die viele Hörer keine harmonischen Reduktionsprinzipien kennen, eine sehr detaillierte Erinnerung an die Oberflächenstruktur (nur einmal) gehörter Musik nachweisen (Krumhansl, 1991). Ohne Aufschluss darüber zu geben, welche Repräsentationsform „wichtiger“ oder typischer sein könnte, verweist dies darauf, dass sowohl semantisches als auch episodisches Wissen gespeichert wird.

Für die Hypothesenbildung hinsichtlich absoluter Tonrepräsentationen ist eine Betrachtung der am deklarativen Gedächtnis beteiligten Prozesse vermutlich aufschlussreicher als der Versuch, zwischen vermutlich nicht eindeutig trennbaren episodischen und semantischen Gedächtnisinhalten zu unterscheiden. Daher werden im folgenden Abschnitt Prozesse des deklarativen Gedächtnisses betrachtet.

4.3 Gedächtnisprozesse beim deklarativen Erinnern

Analysiert man Gedächtnisleistungen auf der Basis der beteiligten Prozesse, so wird meist zwischen Enkodierung, Speicherung, Abruf und Vergessen von Information differenziert. Squire & Kandel (1999, S. 71) betonen, dass viele Ereignisse zwar aufwandsarm und anscheinend automatisch ins LZG gelangen können, dass Lernen und Erinnern aber keinesfalls als passiv oder gar automatisch anzusehen seien. Vielmehr sei es von einer Reihe von Faktoren, die sich meist auf den Zeitpunkt des Lernens beziehen, abhängig, ob eine Information später erinnert werde oder nicht. Wichtige Faktoren sind die Häufigkeit, mit der eine Information wahrgenommen oder gelernt wird, ihre Wichtigkeit, das Ausmaß, in dem der Wahrnehmende die Information organisieren und in sein Vorwissen integrieren kann und das Ausmaß an Rehearsal bzw. späterer Übung des Lerninhalts. Außer Rehearsal wirken sich alle diese Faktoren bereits auf die sehr frühe Informationsverarbeitungsphase, die Enkodierung, aus.

4.3.1 Enkodierung

Bei der Enkodierung werden wahrgenommene Informationen für die weitere Speicherung vorbereitet. Ein bei der Enkodierung wichtiger Faktor ist die Ausrichtung der Aufmerksamkeit auf persönlich wichtige Reize oder Reizmerkmale aus der aktuellen Reizumwelt. Squire & Kandel (1999, S. 72) betonen, dass selbst bei inzidentellem Lernen die persönliche Lerngeschichte eine wichtige Rolle bei der Auswahl relevanter Reizaspekte spiele: „When no particular effort is being made to record experiences for later, our interests and preferences direct our attention and determine the quality and quantity of encoding. Our interests and preferences thereby influence the nature and the strength of the resulting memory".

Auf das langfristige Tonartgedächtnis bezogen lässt sich aus der Bedeutung der Aufmerksamkeitsausrichtung für die Enkodierung die Hypothese ableiten, dass musikalische Expertise einen Einfluss auf die Erinnerungsgüte für Tonarten habe. Vpn mit längerer musikalischer Erfahrung haben vermutlich mehr Interesse für und ein reicheres Vorwissen über melodische Merkmale, so dass ihnen eine elaboriertere Melodieverarbeitung möglich ist als musikalischen Laien. Außerdem kann man annehmen, dass musikalische Experten durch ihr Vorwissen einen breiteren Aufmerksamkeitsfokus haben, so dass es ihnen besser als Laien möglich ist, neben den definierenden relationalen Melodiemerkmalen auch absolute Melodiemerkmale zu enkodieren. In bisherigen Experimenten zur Tonarterinnerung konnten Expertiseeffekte allerdings entweder nicht nachgewiesen werden (Levitin, 1994; Schellenberg & Trehub, 2003) oder aber sie wurden gar nicht erst untersucht (Corso, 1957; Terhardt und Mitarbeiter, 1982, 1983). Vitouch & Gaugusch (2000) fanden eher einen Vertrautheits- als einen Expertiseeffekt, da die Pianisten bessere Leistungen zeigten als die Nichtpianisten. Zu diesen Ergebnissen ist aber anzumerken, dass jeweils hinsichtlich musikalischer Expertise recht homogene Stichproben (in den meisten Experimenten: nur musikalische Laien) untersucht wurden. Der vermutete Expertiseeffekt sollte sich in einer größeren und hinsichtlich musikalischer Expertise heterogenen Stichprobe eher zeigen.

4.3.2 Speicherung

Dem Enkodierungsprozess nachgeschaltet ist die Phase der *Speicherung* von Informationen. Squire & Kandel (1999) schließen aus zahlreichen Experimenten mit Tieren und Menschen, dass es im Gehirn kein „Gedächtnis-Zentrum" gibt, in dem Gedächtnisinhalte quasi abgelegt werden. Vielmehr weisen viele Befunde darauf hin, dass die Kortexareale, die bei der Wahrnehmung und Verarbeitung einer bestimmten Reizklasse eine Rolle spielen, auch bei der Erinnerung an diese Reizklasse aktiv sind. Das bedeutet, dass das Gedächtnis z.B. für eine bestimmte Melodie als „Nachwirkung" der Verarbeitung dieser Melodie angesehen wird und die resultierenden kurz- und längerfristigen Veränderungen im Gehirn als „Engramme".

Für Zeiträume im Sekundenbereich wurde dies z.B. von Fuster (1995) nachgewiesen. Mithilfe von Einzelzellableitungen bei Affen konnte Fuster während einer Farbwahrnehmungsaufgabe Aktivität in einem mit der visuellen Wahrnehmung assoziierten Areal im Temporallappen nachweisen, die nach Verschwinden des Reizes noch andauerte („sustained activation")

und daher als Surrogat des Rehearsal-Prozesses interpretiert wurde. Da sich bei arbeitsgedächtnisbasierten Aufgaben auch der Frontalkortex als wichtig erwiesen hat (Goldman-Rakic, 1992), vermuten Squire & Kandel, dass beim kurzfristigen Erinnern eine Zusammenarbeit des Frontalkortex und der sensorischen Areale vorliegt, wobei der Frontalkortex (im Sinne einer zentralen Exekutive) den sensorischen Arealen signalisiert, welche Reize aufrecht zu erhalten sind.

Für das längerfristige Erinnern von Informationen ist eine Veränderung der Synapsenstärken (als Folge der ursprünglichen Reizverarbeitung) maßgeblich. Im einfachen Fall des Lernens von Reizpaaren beispielsweise reagiert ein Neuron, das ursprünglich nur auf einen Paarling reagiert hatte, nach dem Lernen auf beide Paarlinge: Diese sind durch verstärkte Synapsen assoziiert. Der Mechanismus des assoziativen Lernens, der auch für nichtdeklarative Lernformen wirksam ist, ist u.a. auch eine denkbare Erklärung für die Entstehung von Ton-Name-Assoziationen bei Absoluthörern. Wie aber kommt es zu solchen neuen Assoziationen bzw. Synapsen? Eine wichtige Rolle spielen dabei die Phase der Konsolidierung von Gedächtnisinhalten und eine Hirnstruktur im Temporallappen, der Hippocampus. Squire & Kandel (1999) vermuten, dass der Hippocampus während der Konsolidierungsphase „pointers“ im Sinne von „Zeigern“ speichere, die verschiedene Kortexareale, die die Repräsentation eines Ereignisses ausmachen, verbinden. Während der Phase der Konsolidierung, die bei unterschiedlichen Reizen unterschiedlich lange dauern kann, sind die Gedächtnisinhalte störanfällig für Interferenz, wie Müller & Pilzecker bereits 1900 nachweisen konnten. Nach der Konsolidierungsphase ist die Assoziation der entsprechenden Kortexareale stabil genug, um einen Abruf auch ohne Beteiligung des Hippocampus zu erlauben. Diese Funktionsweise bietet eine Erklärung dafür, dass Patienten mit Hippocampus-Läsionen wie z.B. der erwähnte H.M. sich an (konsolidierte) Ereignisse aus ihrer Jugend erinnern können, jedoch keine neuen Repräsentationen im LZG aufbauen können, obgleich ihr Arbeitsgedächtnis nicht beeinträchtigt ist. Der evolutionäre Vorteil eines solchen allmählichen Aufbaus von LZG-Repräsentationen liegt in einer größeren Stabilität des Gedächtnissystems, das sich nicht durch jede beliebige „Störung“ im Sinne abweichender Erfahrungen verändert, sondern sich allmählich an veränderte Umweltgegebenheiten anpassen kann.

Für die Beantwortung der Frage, wodurch Informationen aus dem Arbeitsgedächtnis ins LZG gelangen, ziehen Squire & Kandel (1999) einen biologischen Mechanismus heran, der im „Anschalten“ bestimmter Lern-Gene besteht. Die Funktion dieser Lern-Gene, die normalerweise erst bei mehr als einem Lerndurchgang, bei emotional bedeutsamen Reizen jedoch auch nach einem einzigen Lerndurchgang („flashbulb memory“) aktiviert werden, liegt in der Synthese neuer Proteine, die wiederum das Wachstum neuer synaptischer Verbindungen ermöglichen (genauer: vgl. Squire & Kandel, 1999, Kap. 7).

Betrachtet man den Aufbau von Tonart-Repräsentationen aus der Perspektive der Erkenntnisse zur Speicherung von Information im LZG, so wird deutlich, dass die Häufigkeit, mit der eine Melodie in der gleichen Tonart gehört bzw. gelernt wird, einen Einfluss darauf haben müsste, ob es zur Ausbildung einer langfristigen Assoziation dieser Melodie und ihrer Tonart

kommen kann. Wenn das Gedächtnis für eine Melodie tatsächlich als Nachwirkung der Informationsverarbeitung zu verstehen ist, so kann man bei sehr häufigem Hören der Melodie in gleicher Tonart den Aufbau einer „Gedächtnisspur“ für diese Melodie in der Originaltonart annehmen. Bei nur einmaligem Hören einer Melodie ist dagegen nur für Melodien, die in einer emotional sehr bedeutsamen Situation gehört wurden, überhaupt eine langfristige Repräsentation zu erwarten.

4.3.3 Abruf

Beim Abruf eines Gedächtnisinhalts, z.B. einer Melodie, müssen die Informationen aus unterschiedlichen Kortexarealen, die die Repräsentation der Melodie ausmachen, integriert werden. Die Güte des Abrufs hängt dabei entscheidend von gegebenen Hinweisreizen ab, denn unvollständige Hinweisreize können zum Abruf einer unvollständigen Episode führen. Interferierende (Fehl-) Informationen können den Abruf empfindlich stören, wie von Loftus, Miller & Burns (1978) nachgewiesen wurde. Dies deutet darauf hin, dass Ereignisse aus dem Gedächtnis nicht einfach abgerufen, sondern rekonstruiert werden. Bei diesem Prozess spielen das Vorwissen und gespeicherte Schemata im Sinne von gelernten Wissenskomplexen eine wichtige Rolle. Beim Erinnern eines gelesenen Textes z.B. wirkt sich vorhandenes Wissen auf die spätere Erinnerung an den Text aus. Passagen, die nicht mehr genau erinnert werden, werden durch Inferenzen ergänzt. Bei der Melodieerinnerung können Schemata für gängige Melodie- und Harmoniefolgen den Abruf beeinflussen.

Bei genauerer Betrachtung der am Abruf von Gedächtnisinhalten beteiligten Prozesse zeigt sich, dass der Abruf aus zwei Komponenten besteht. Der Abruf im eigentlichen Sinne bezeichnet die Aktivierung des Gedächtnisinhalts, so dass dieser im Arbeitsgedächtnis verfügbar ist. Um im Experiment die Erinnerungsgüte überprüfen zu können, ist jedoch noch eine zweite Komponente notwendig, nämlich die Externalisierung der im Arbeitsgedächtnis verfügbaren Gedächtnisinhalte. Diese wird in gedächtnispsychologischen Experimenten meist entweder über Wiedererkennungs- (Recognition) oder Produktionsaufgaben (Recall) verwirklicht. Dabei zeigt sich bei verbalen Aufgaben typischerweise, dass Wiedererkennungsaufgaben zu besseren Erinnerungsleistungen führen als Produktionsaufgaben. Dies könnte man dadurch erklären, dass durch die Darbietung verschiedener zu beurteilender Wörter mehr Hinweisreize vorhanden sind als beim freien Recall. Beim Vergleich von Tonidentifikations- und Tonproduktionsleistungen von Absoluthörern ergibt sich ebenfalls ein Vorteil für die erstgenannten. Beim Abruf von Tonarten durch Nichtabsoluthörer dagegen ist der Leistungsvorteil bei Wiedererkennungsaufgaben bisher nicht nachgewiesen worden. Da es kaum Experimente gibt, die den Abruf derselben Melodie sowohl durch Wiedererkennungs- als auch durch Produktionsmethoden gemessen haben, ist die Datenlage zur Beurteilung der relativen Schwierigkeit von Wiedererkennung und Produktion von Tonarten begrenzt. Lediglich Halpern (1989, vgl. Kap. 3.1) verglich die Produktion und Wiedererkennung von Volksliedern und fand etwas bessere Leistungen in der Produktionsbedingung. Das könnte ein Hinweis darauf sein, dass bei der Erinnerung an Volkslieder motorische Hinweisreize, die in Produktionsaufgaben besser genutzt werden können als in Wiedererkennungsaufgaben, wichtig

sind. Vorgespielte Vergleichstöne des Klaviers dagegen könnten im Sinne von Interferenz wirken und die Vorstellung der Tonart eher stören als fördern. Schellenberg & Trehub (2003, S. 264) berichten, dass in ihrer Tonart-Wiedererkennungsaufgabe das erste Trial signifikant besser gelöst wurde als alle Folgetrials und folgern daraus: „increasing exposure to pitch-shifted excerpts interfered with memory for the original pitch level“. Dass Interferenz auch bei Produktionsaufgaben auftreten kann, deutet sich in den Daten von Levitin (1994) an.

Die bisherigen Ergebnisse weisen darauf hin, dass sowohl Produktions- als auch Wiedererkennungsmethoden die Vorstellung nicht genau abbilden. Bei Produktionsaufgaben zeigt sich häufig, dass die Mehrzahl der Vpn zu tief singt. Dies könnte ein Hinweis darauf sein, dass sie ihre Vorstellung (unbewusst und blitzschnell) nach unten transponieren, bevor sie singen, da kaum jemand gerne hoch singt. Bei Wiedererkennungsaufgaben, die in der Darbietung verschiedener Transpositionen um das Original herum (also gleich weit nach oben und nach unten) bestehen, besteht dagegen die Gefahr, dass die eigene Vorstellung durch die Ermittlung einer „mittleren“ Tonartversion verzerrt wird. Da es auch den modernen bildgebenden Verfahren bislang nicht möglich ist, Gedächtnisinhalte „abzubilden“, ist nicht zu entscheiden, welchen Beitrag an der Größe der Erinnerungsfehler (im Sinne der Abweichung von den Originaltonarten) die Externalisierungsmethode hat bzw. inwieweit wirklich ungenaue Tonartrepräsentationen vorliegen.

Das von Tulving (1983) postulierte Prinzip der Enkodierungsspezifität besagt, dass für einen erfolgreichen LZG-Abruf auch die Ähnlichkeit von Lern- und Abrufkontext eine Rolle spielt. Dabei ist mit dem Kontext nicht nur die räumliche Umgebung gemeint, sondern auch die emotionale Verfassung und sogar vorgestellte Kontexte. Z.B. konnte Smith (1979) zeigen, dass Vpn, die sich nach einem Kontextwechsel die ursprüngliche Lernsituation vorstellten, fast genauso gute Abrufleistungen erbrachten wie Vpn ohne Kontextwechsel, während Vpn mit Kontextwechsel, aber ohne Vorstellung der Lernsituation, in der Abrufleistung abfielen.

Die Erkenntnisse zum Abruf von Informationen haben zum Einen praktische Konsequenzen. Für die Planung von Experimenten mit Abrufaufgaben ist es demnach wichtig, dass allen Vpn die gleichen Hinweisreize (z.B. Titel oder Text einer Melodie) gegeben werden, da diese den Abruferfolg stark beeinflussen können. Um einen möglichst guten Abruf zu ermöglichen, ist es außerdem sinnvoll, dass sich die Vpn die ursprüngliche Lernsituation vorstellen, sofern Experimente in genau gleichem Kontext nicht möglich sind. Zum Anderen ist festzuhalten, dass die Externalisierungsmethode des aktivierten Gedächtnisinhalts die Erinnerungsqualität beeinflusst, wobei sich für den Abruf von Tonarten nach der bisherigen Befundlage ein leichter Leistungsvorteil bei der Produktion im Vergleich zur Wiedererkennung andeutet. Die Möglichkeit der leichteren Aktivierung von motorischen Kontextinformationen bei Produktionsaufgaben sowie das Nichtvorhandensein der als Interferenz empfundenen transponierten Versionen könnten Ursachen für den Leistungsvorteil bei Produktionsaufgaben sein.

4.3.4 Vergessen

Die Beobachtung des Gedächtniskünstlers D.C. Shereshevskii durch den russischen Neuropsychologien Luria über 30 Jahre hinweg (beschrieben in Squire & Kandel, 1999, Kap. 4) zeigte, dass das Vergessen von Informationen und die Reduktion von Erlebnissen auf das Wesentliche wichtige Leistungen des menschlichen Gedächtnisses sind. Während Shereshevskiis reichhaltige visuelle Vorstellungskraft ihm das Behalten von langen Wortlisten ermöglichte, konnte er sich im Alltag kaum gegen die Flut an visuellen Assoziationen zu sämtlichen dargebotenen Informationen wehren. Dadurch hatte er zwar eine sehr detailreiche Erinnerung, jedoch Schwierigkeiten bei dem Versuch, die Bedeutung von Sätzen zu erfassen. Das „normale“ Gedächtnis dagegen wird als genau entgegengesetzt beschrieben, es ist wesentlich besser bei der Generalisierung, Abstraktion und Zusammenfassung als bei der exakten Reproduktion von Fakten oder Ereignissen. Das bedeutet, dass einzelne Erlebnisse nicht für immer im Gedächtnis fixiert sind, sondern mit der Zeit verblassen bzw. mit anderen Erlebnissen zusammen zu semantischem Wissen umgeformt werden. Für das Verblassen von Informationen oder Erlebnissen werden verschiedene Mechanismen verantwortlich gemacht: der spontane Zerfall von Information, die vorübergehende Nichtzugänglichkeit zumindest mit bestimmten Aufgaben sowie proaktive oder retroaktive Interferenz. Aktuell wird durch Tierstudien eher die Annahme gestützt, dass Vergessen zumindest teilweise den endgültigen Verlust von Informationen bedeutet, d.h., dass die lerninduzierten synaptischen Veränderungen wieder verschwinden bzw. durch neue Erfahrungen überformt werden. Squire & Kandel (1999) beschreiben Vergessen als einen kontinuierlichen Prozess, in dem Gelerntes abgeschwächt und modifiziert wird, jedoch nicht unbedingt vollständig. Zum Einen können beim Vergessen von Episoden nichtdeklarative Erinnerungen bleiben, z.B. Dispositionen, die infolge längst vergessener Ereignisse entstanden sind. Zum Anderen ist es von der Bedeutung und der Lernintensität abhängig, wie schnell Ereignisse vergessen werden. Sehr gut gelernte Informationen können ebenso wie emotional wichtige Erlebnisse über Jahre hinweg im Gedächtnis bleiben.

Die entlastende Funktion des Vergessens für das Gedächtnis kommt auch beim Hören von Musik zum Tragen: Ein Musikstück wie z.B. eine Mozart-Sinfonie mit einer halben Stunde Hördauer kann nach einmaligem Hören ebenso wenig exakt im Gedächtnis erhalten bleiben wie eine Konversation von der gleichen Dauer. Viele der Details werden daher vergessen (vgl. Kap. 4.2 zur Frage reduzierter Repräsentationen). Bei Melodien, die sehr häufig in gleicher Form gehört werden, ist jedoch von detaillierter Erinnerung auszugehen, wie sie bereits im Abschnitt *Speicherung* diskutiert wurde. Aufgrund der Befunde zum Vergessen lässt sich hier nochmals die Hypothese formulieren, dass die Tonart umso genauer repräsentiert ist, je besser die Melodie ursprünglich gelernt wurde.

Aufgrund der Betrachtung verschiedener Prozesse, die am deklarativen Erinnern beteiligt sind, lassen sich zusammengefasst die folgenden Schlussfolgerungen für die deklarative Tonarterinnerung ziehen:

Aufgrund des reicheren Vorwissens, in das aktuell gehörte Melodien während der Enkodierungsphase integriert werden können, ist ein Leistungsvorteil bei musikalisch erfahreneren Menschen bezüglich der Tonarterinnerung zu vermuten.

Aufgrund der Bedeutung, die der Lernintensität bei der Speicherung bzw. beim Vergessen von Informationen zukommt, ist für die Tonarterinnerung zu vermuten, dass die Tonarten gut gelernter Melodien besser abgerufen werden können als die Tonarten weniger gut gelernter Melodien. Ähnliches gilt für aufgrund emotionaler Faktoren persönlich wichtige Melodien.

Anhand der bisherigen Erkenntnisse zum Abruf von Tonarten aus dem Gedächtnis lässt sich vermuten, dass motorische Kontextinformationen für den Abruf wichtig sind und die Produktion im Vergleich zur Wiedererkennung von Tonarten erleichtern.

4.4 Formen nichtdeklarativen Erinnerns

Die Betrachtung verschiedener Formen nichtdeklarativen Erinnerns ist sowohl für die Hypothesenbildung bezüglich der Tonarterinnerung von Nichtabsoluthörern als auch für die allgemeinere Betrachtung des absoluten Gehörs wichtig, da einige der in Kap. 2 und 3 berichteten Befunde mit impliziten Methoden erhoben wurden, die nichtdeklaratives Erinnern im Sinne von Erfahrungsnachwirkungen messen sollen. Implizite Gedächtnistests haben anders als z.B. Wiedererkennungstests keinen expliziten Bezug zu einer Lernphase. Statt dessen wird inzidentell gelernt, wie man sich z.B. an der „Wortstammaufgabe“ verdeutlichen kann. Unter dem Vorwand, an einer Normierungsstudie teilzunehmen, lesen Vpn dort eine Reihe von Wörtern, die sie z.B. unter dem Aspekt der Auftrittshäufigkeit beurteilen sollen. Später werden bei der Aufgabe, Wortanfänge zu ergänzen, häufiger Wörter aus der „Lernphase“ ergänzt, obwohl keine explizite Erinnerung an diese Wörter vorliegt. Interessant ist der Vergleich zwischen impliziten und expliziten Tests bei Patientengruppen wie z.B. Amnestikern: Dort zeigt sich ein starker Leistungsabfall im Vergleich zu gesunden Vpn nur bei expliziten, nicht aber bei impliziten Gedächtnistests (Hamann & Squire, 1996). Eine solche Dissoziation kann man im Sinne funktional verschiedener Gedächtnissysteme interpretieren, also einer Trennung expliziter Erinnerung von unbewusster Erfahrungsnachwirkung.

Auf die Erinnerung bzw. Benennung einzelner Töne bezogen wäre es eine interessante Forschungsfrage, ob sich ähnliche Dissoziationen zwischen expliziter und impliziter Erinnerung zeigen lassen. Nichtabsoluthörer könnte man aufgrund ihrer Leistungen im (expliziten) Tonidentifikationstest auch als „Einzelton-Amnestiker“ bezeichnen, da sie Töne deutlich schlechter identifizieren können als Absoluthörer. Gelänge nun der Nachweis, dass sich Absoluthörer und Nichtabsoluthörer bei einem impliziten Tongedächtnistest nicht unterscheiden, so wäre dies ein Hinweis darauf, dass absolute Tonhöhen auch von Nichtabsoluthörern unbewusst erinnert werden. Ein solcher Nachweis liegt noch nicht vor, da sich bisherige implizite Tests für die Nachwirkung absoluter Tonhöhen in vergleichsweise kurzen Zeiträumen bewegen (vgl. Kap. 2). Außerdem warnen Buchner & Brand (2002) vor einer voreiligen Trennung von deklarativen und nichtdeklarativen Gedächtnissystemen aufgrund der Ergebnisse impliziter Testmethoden: Gefundene Dissoziationen könnten auch ein Hinweis darauf sein, dass

implizite Testmethoden einfach nur weniger reliabel und dadurch auch weniger trennscharf messen.

Squire & Kandel (1999) stellen die in vielen Fällen durch neuropsychologische Daten gesicherte Trennung von deklarativem und nichtdeklarativem Gedächtnis nicht in Frage, betonen aber, dass es auf der Ebene der molekularen Grundlagen beider Gedächtnisformen weniger große Unterschiede gebe als auf der Verhaltensebene. Bei beiden Gedächtnisformen werde zunächst eine instabile, kurzfristige Phase durchlaufen, die von einem stabilen, langfristigen Gedächtnis abgelöst werde. Dabei wirke sich Wiederholung positiv auf die Ausbildung stabiler Gedächtnisinhalte aus. Obgleich unterschiedliche Gehirnregionen an den beiden Gedächtnisformen beteiligt sind, bewirken ähnliche biochemische Prozesse sowohl in der deklarativen als auch in der nichtdeklarativen Form eine Veränderung von Synapsen. Die zur Veränderung von Synapsen notwendige Zeitspanne spiegele in etwa die Lebensdauer der Gedächtnisinhalte wieder. Im Folgenden werden einige Formen nichtdeklarativen Lernens und Erinnerns vorgestellt, um zu überprüfen, welche Prozesse bei der Tonarterinnerung wichtig sein könnten.

4.4.1 Priming

Priming bezeichnet eine Erleichterung der Wahrnehmung für Reize, die kurz zuvor in sehr ähnlicher Form wahrgenommen wurden. Beispielsweise können Wörter, die bereits einmal gelesen wurden, beim zweiten Lesen schneller erkannt werden. Die Funktion dieses Gedächtnismechanismus' ist die schnellere Reaktionsmöglichkeit in Umgebungen, die einem Individuum sehr vertraut sind. Dass Primingeffekte auch sehr lange andauern können, zeigte Cave (1997) mit einer Bildbenennungsaufgabe, bei der die schnelleren Reaktionszeiten auf bereits gesehene Bilder auch ein Jahr nach Bilddarbietung noch zu beobachten waren. Ein Beispiel für Priming im auditiven Bereich ist das in Kap. 3.1 beschriebene Experiment zur Worterkennung, das darauf hinweist, dass konstant gehaltene absolute Tonhöhen der Sprechstimmen die Erkennung der Wörter erleichterten. Unklar ist bislang die Lebensdauer solcher Repräsentationen, außerdem stehen Untersuchungen, die sich auf Melodien beziehen, noch aus. Könnte man eine erleichterte Wahrnehmung (z.B. schnellere Erkennung) von Melodien in vertrauter Tonart im Vergleich zu veränderten Tonarten nachweisen, so wäre dies ein Hinweis auf eine unbewusste Tonarterinnerung bei Nichtabsoluthörern.

4.4.2 Wahrnehmungslernen

Mit Wahrnehmungslernen wird die durch umfangreiche Übung erworbene verbesserte Fähigkeit bezeichnet, zwischen bestimmten Reizmerkmalen zu unterscheiden. Beispielsweise wurden Vpn von Karni & Sagi (1991) gebeten, auf einem Bildschirm, der mit kurzen horizontalen Linien bedeckt war, ein Muster von drei Schrägstrichen zu finden und anzugeben, ob diese horizontal oder vertikal angeordnet waren. Nachdem die Vpn diese Aufgabe zwei Wochen lang geübt hatten (täglich 1000 Trials), stiegen Erkennungsrate und -geschwindigkeit stark an. Ein Beispiel aus dem wirklichen Leben wäre die Fähigkeit von Weinkennern, zwi-

schen Weinen verschiedener Jahrgänge zu differenzieren, die für den „Wein-Laien“ nicht unterscheidbar sind, weil sie z.B. allesamt „trocken“ schmecken. Die erworbene Diskriminierungsfähigkeit ist stark aufgabenspezifisch. Beispielsweise war bei den oben beschriebenen Vpn kein Transfer von einer zu der anderen Hälfte des Gesichtsfeldes möglich. Der schlechte Transfer deutet darauf hin, dass sich Wahrnehmungslernen auf die frühen (sensorischen) Phasen der Reizverarbeitung auswirkt. Beispielsweise könnten durch Wahrnehmungslernen die in Abschnitt 4.3.1 beschriebenen Expertiseeffekte bei der Enkodierung von Information entstehen.

Goldstone (1998) unterscheidet zwischen verschiedenen Prozessen des Wahrnehmungslernens. Mit *attentional weighting* ist eine Ausrichtung der Aufmerksamkeit auf relevante und gleichzeitiges Ignorieren irrelevanter Reizdimensionen oder -merkmale gemeint, wodurch z.B. Kategorisierungsaufgaben erleichtert werden. Auf musikalische Information bezogen ließ sich kategorische Wahrnehmung im Sinne einer besseren Unterscheidung physikalisch unterschiedlicher Reize, wenn sie nicht der gleichen Kategorie angehörten, bei der Intervallwahrnehmung von Musikern, nicht aber musikalischen Laien zeigen (Burns & Ward, 1978; Zatorre & Halpern, 1979). Offensichtlich hat die im Musikstudium häufige Übung der Intervall-Kategorisierung bewirkt, dass die Aufmerksamkeit auf Unterschiede mit Halbton-Auflösung gerichtet wird und kleinere Unterschiede ignoriert werden. Dass eine der Aufgabe nicht angemessene Kategorisierung zu Leistungseinbußen bei den Musikern führte, erinnert an die in Kap. 2.4 beschriebenen Leistungseinbußen bei der Intervallbenennung durch Absoluthörer. Möglicherweise ist attentional weighting ein Mechanismus, der zur Entwicklung von Tonkategorien bei Absoluthörern beiträgt.

Ein weiterer Mechanismus des Wahrnehmungslernens ist das *imprinting*, die Entwicklung von Detektoren für wiederholt auftretende Reize, die eine schnellere Verarbeitung solcher Reize ermöglichen. Auf neuronaler Ebene konnte z.B. Weinberger (1993) zeigen, dass einzelne Zellen im auditorischen Kortex auf die Frequenz von sehr oft wiederholten Tönen „eingestimmt“ werden können. Für andere Zellen im inferioren Temporallappen konnte nach 300.000 Trials ein ähnliches „Einstimmen“ auf dreidimensionale Objekte gezeigt werden (Logothetis, Pauls & Poggio, 1995). Der imprinting-Mechanismus könnte sowohl für Lernprozesse von Absoluthörern als auch von Nichtabsoluthörern eine Rolle spielen, wobei festzuhalten ist, dass imprinting-Prozesse sehr viele Lerndurchgänge erfordern und somit nur bei sehr oft gehörten Melodien wirksam werden können.

Die Differenzierung (*differentiation*) erlaubt als dritter Mechanismus des Wahrnehmungslernens die nach längerem Training mögliche Unterscheidung von ursprünglich als gleich empfundenen Reizen. Beispielsweise ist es durch Training möglich, die Unterschiedsschwelle (just noticeable difference) von Tönen deutlich zu senken. Für die absolute Tonerinnerung spielt dieser Mechanismus vermutlich keine wichtige Rolle, denn wie in Kap. 2 beschrieben, ist Absoluthörern ja gerade keine bessere Tondiskriminierung möglich als Nichtabsoluthörern. Für beide Gruppen ist beim Hören realer Musik eine Differenzierung von Halbtönen ausreichend.

Der vierte von Goldstone beschriebene Mechanismus, die *unitization*, integriert ähnlich wie ein Chunking-Mechanismus Reizmerkmale zu einem Ganzen, das als Einheit verarbeitet wird. Beispielsweise deutet die schnellere Verarbeitung von Wörtern im Vergleich zu sinnlosen Silbenfolgen darauf hin, dass Wörter auf einem höheren Niveau als dem einzelner Buchstaben wahrgenommen werden (Salasoo, Shiffrin & Feustel, 1985). Eine solche Integration von Reizmerkmalen findet auf der Ebene einzelner Töne statt, die nicht als Grundfrequenz + Harmonische wahrgenommen werden, sondern als ein Komplexton. Die Rolle dieses Mechanismus bei der absoluten Tonerinnerung ist dagegen fraglich.

Zusammenfassend lässt sich feststellen, dass die für Wahrnehmungslernen als typisch beschriebene verbesserte Differenzierungsfähigkeit für die absolute Tonerinnerung gerade kein typischer Prozess ist. Trotzdem könnten die beschriebenen Prozesse der Aufmerksamkeitszuwendung und Reizeinprägung für die Entstehung musikalischer Expertise, die sich bereits auf die frühen Phasen der Reizverarbeitung auswirkt, wichtig sein.

4.4.3 Emotionales Lernen

Während durch Priming und Wahrnehmungslernen die Reizverarbeitung verbessert wird, machen Ergebnisse zum emotionalen Lernen deutlich, dass durch wiederholte Reizdarbietung auch die Bewertung dieses Reizes beeinflusst wird. Dies wird beispielsweise am „mere exposure“ Effekt, den Zajonc & Hazel (1982) zeigen konnten, deutlich: Ihre Vpn sahen für Zeiträume im Millisekundenbereich geometrische Formen, die sie in einem späteren Gedächtnistest nicht wiedererkannten. Bei der Aufgabe Präferenzurteile abzugeben, wurden jedoch bereits gesehene Reize neuen Formen deutlich vorgezogen. Umgekehrt wird in zahlreichen Experimenten deutlich, dass die emotionale Bewertung von Reizen sich auf die explizite Erinnerung auswirkt, indem die Erinnerung an emotional bewegende Reize besser ist als die an neutrale Reize. Als Ursache dieses Mechanismus sehen Cahill & McGaugh (1995, 1998) die Funktionsweise der Amygdala an, die sowohl mit verschiedenen Kortexregionen als auch mit dem Hippocampus vernetzt ist. Eine Aktivierung der Amygdala (über die Ausschüttung von Hormonen) könnte die Verarbeitung und damit auch das Gedächtnis für die Reize erleichtern, die gleichzeitig mit der emotionalen Aktivierung auftraten (bis hin zum „flashbulb memory“, der detaillierten Erinnerung an ein einzelnes emotional stark besetztes Erlebnis). Demnach kommt es beim emotionalen Lernen zu Wechselwirkungen zwischen deklarativem und nichtdeklarativem Lernen. Auf das Lernen von Tonarten bezogen führt dies zu der in Abschnitt 4.3 bereits vorgestellten Hypothese, dass die emotionale Bewertung von Melodien einen Einfluss auf die Tonarterinnerung haben könnte.

4.4.4 Fertigkeiten und Gewohnheiten

Unter dem Oberbegriff Fertigkeiten werden verschiedene erworbene Verhaltensmuster, z.B. motorische, kognitive und perzeptuelle Fertigkeiten zusammengefasst. Ähnlich wie Gewohnheiten werden Fertigkeiten durch Übung erworben, ohne dass sich der Übende dessen bewusst ist oder die Lerninhalte oder -episoden wie deklarative Gedächtnisinhalte benennen

könnte. Beispiele für *Gewohnheiten* sind das Händewaschen vor dem Essen oder das Sagen von „bitte“ oder „danke“, dies wird früh in der Ontogenese erlernt. Der Erwerb *motorischer Fertigkeiten*, zu denen z.B. das Fahrrad- oder Autofahren zählen, wird oft mit Hilfe von sequenziellen Lernaufgaben untersucht, bei denen Vpn per Tastendruck (verschiedene Tasten) auf verschiedene Reize reagieren sollen („tapping“). Übung erhöht die Reaktionsgeschwindigkeit erheblich, auch bei Amnestikern (Reber & Squire, 1998). Untersuchungen mit funktioneller MRT (Karni, 1997; Ungerleider, 1995) haben gezeigt, dass die Areale im motorischen Kortex, die beim Fingertappen aktiv sind, sich durch Übung vergrößern und dass diese Vergrößerung nach dem Training stabil bleibt, ebenso wie die höhere Reaktionsgeschwindigkeit. Offen ist bislang, ob motorische Gedächtnisinhalte langfristig im motorischen Kortex repräsentiert werden oder ob die Repräsentation eher auf einer verstärkten Verbindung zwischen motorischem Kortex und Neostriatum beruht.

Zu den *perzeptuellen Fertigkeiten* gehört z.B. das Lesen: Musen & Squire (1991) konnten zeigen, dass mehrmaliges Vorlesen eines Textes zu einem schnelleren Lesetempo (in gewissen Grenzen!) führte, das nicht mit der deklarativen Erinnerung an die gelesenen Inhalte zusammenhing. *Kognitive Fertigkeiten* wurden von Berry & Broadbent (1985) anhand des Umgangs mit computergenerierten Welten untersucht. Bei der Aufgabe, die Leitung einer Zuckerfabrik zu simulieren, machten ihre Vpn allmähliche Fortschritte, ohne eine bestimmte Strategie benennen zu können. Die Leistungssteigerung schien auf dem Aufbau intuitiven Wissens zu beruhen. Dass dies auch Amnestikern möglich ist, wurde von Squire & Frambach (1990) mit dem gleichen Aufgabentyp gezeigt.

Der Erwerb von Fertigkeiten ist für musikalisches Lernen zweifellos von großer Bedeutung: Es müssen Noten gelesen und mit bestimmten Griffen oder Tasten in Zusammenhang gebracht werden, und auf höherem Niveau üben z.B. Jazzmusiker das Spielen von Skalen und komplexeren Tonfolgen, die sie beim Improvisieren automatisiert abrufen können. Im Zusammenhang mit dem Gedächtnis für Tonarten ist der Beitrag motorischer Gedächtnisinhalte, zu dem es bereits einige Untersuchungen gibt, von größtem Interesse.

4.4.5 Motorische Gedächtnisinhalte

Der Beitrag motorischer Gedächtnisinhalte wurde für zeitliche Aspekte der Musikwahrnehmung in einer Vielzahl von Experimenten untersucht, die hier nicht referiert werden sollen, da die Frage, in welchem Ausmaß motorische Gedächtnisinhalte beim Abruf von Tonhöhen bzw. Tonarten wirksam werden, im Vordergrund steht. Dabei sind zwei Komponenten des motorischen Gedächtnisses zu unterscheiden: Zum einen gibt es die *singmotorische* Erinnerung, also das Wissen, wie sich das Singen z.B. eines zweigestrichenen *e* anfühlt. Dieses Wissen beruht auf impliziter Erinnerung z.B. an die Kehlkopfstellung beim Singen verschiedener Töne, die Sänger aufgrund langjähriger Erfahrung mit ihrem eigenen Stimmumfang und den Registerwechseln der Stimme aufbauen, es ist also abhängig von der Sing-Expertise. Zur Genauigkeit des singmotorischen Gedächtnisses gibt es nur sehr wenige empirische Befunde. Fündig wird man bei Studien zur Genauigkeit der Intonation beim Singen, die als

„Deprivationsstudien“ angelegt sind, also in irgendeiner Weise den Zugang zur auditiven Information erschweren bzw. verhindern. Diesen Experimenten ist ein Modell für die Tonhöhenkontrolle von Wyke (1974) vorauszuschicken, nach dem es beim Singen zwei Kontrollmechanismen für die Tonhöhe gibt: einen von der bewussten Wahrnehmung gesteuerten Weg, der auf auditive Rückkopplung gehörter Töne angewiesen ist, und einen unbewussten, an das Gehör gebundenen Reflexweg. Um der Frage, wie genau das singmotorische Gedächtnis für Tonhöhen sein kann, nachzugehen, wird hier also zunächst danach gefragt, inwieweit genaue Intonation ausschließlich über den Reflexweg, also ohne auditive Rückkopplung, möglich ist.

Um dieser Frage nachzugehen, ließ Murry (1990) Töne nachsingen und untersuchte den Ton-Onset, also die ersten 5 Schwingungen der produzierten Töne. Dies ist insofern aufschlussreich, als die auditive Rückkopplung erst nach einer kleinen Zeitverzögerung (im Bereich von 100-250 ms) einsetzt, so dass bei den ersten Schwingungen tatsächlich von „reflexhaft“ gesungenen Tönen auszugehen ist. Das Ergebnis war, dass die ersten Schwingungen vom endgültigen Ton beträchtlich abweichen: Bei professionellen Sängern betrug die Abweichung durchschnittlich nur einen Halbton, bei Laien dagegen durchschnittlich 3 (und bis zu 7) Halbtöne. Dies bedeutet natürlich nicht, dass die Sänger den ganzen Ton so ungenau gesungen haben, denn nach Einsetzen der auditiven Rückkopplung wird die Tonhöhe blitzschnell korrigiert. Bevor die auditive Kontrolle einsetzt, wird – für einen sehr kurzen Zeitraum – allerdings recht ungenau gesungen. Dieses Ergebnis deutet darauf hin, dass das reflexhafte Singen ohne auditive Rückkopplung zumindest bei nicht professionellen Sängern selbst beim bloßen Nachsingen von Tönen nicht sehr genau ist.

Eine andere Methode wandten Ward & Burns (1987) an: Sie ließen Sänger Tonleitern singen, während ihr Gesang durch lautes Rauschen (über Kopfhörer dargeboten) maskiert wurde. Hörten sich die Sänger beim Singen nicht, so wurde ihr Gesang deutlich ungenauer. Während beim unmaskierten Gesang eine Genauigkeit von 10-15 Cent berichtet wird, liegt diese beim maskierten Gesang nur noch bei 40-55 Cent (ein Halbton entspricht 100 Cent). Auch dieses Experiment deutet also darauf hin, dass Intonieren ohne auditive Rückkopplung deutlich ungenauer bzw. die singmotorische Erinnerung alleine nicht sehr genau ist.

In seiner Diplomarbeit wandte Kestler (1999) eine dritte Methode an: Er nutzte einen „pitch shifter“, um seinen Vpn falsches Feedback über die von ihnen gesungenen Töne zu geben. Während die Vpn einen Ton in bequemer Mittellage zu singen hatten, wurde ihr Gesang aufgenommen und ihnen über einen Kopfhörer und Lautsprecher vorgespielt. Nach einigen Sekunden wurde die gesungene Tonhöhe um einen Halbton nach oben oder unten verschoben, so dass die Vpn einen anderen Ton hörten als sie sangen. Aus der Tatsache, dass seine Vpn diese Verschiebungen schnell, genau (Sänger besser als Nichtsänger) und unbewusst (den Vpn war nicht bewusst, dass sie nach der Kompensation der Störung eine andere Tonhöhe als vor dem Sprung sangen) kompensieren konnten, schloss Kestler auf eine grundsätzliche Anpassungsfähigkeit des singmotorischen Tonhöhengedächtnisses und auf eine primär vorherrschende Tonhöhenkontrolle über das Gehör.

Insgesamt weisen die berichteten Ergebnisse darauf hin, dass Sänger ein genaueres singmotorisches Tonhöhengedächtnis haben als Nichtsänger, dass aber für die genaue Intonation, also gewissermaßen die „Feinregelung“ der Stimme, das Gehör für beide Gruppen unerlässlich ist. Dabei ist aber zu bedenken, dass in diesen Experimenten erstens eine relationale Aufgabe vorlag, also das Nachsingen vorgespielter Töne oder das Singen einer Tonleiter, und dass diese Ergebnisse nicht zwangsläufig auf den Abruf von Tönen aus dem LZG übertragbar sind. Zweitens sind die von Kestler (1999) und Ward & Burns (1987) berichteten „Ungenauigkeiten“ zwar groß, gemessen an der unter ungestörten Bedingungen möglichen Intonationsgenauigkeit, aber nicht so groß, dass sie singmotorische Gedächtnisinhalte als eine mögliche Grundlage für den Tonartabruf disqualifizieren. Die von Murry (1990) berichteten Abweichungen dagegen sind groß genug, um einen Beitrag singmotorischer Gedächtnisinhalte in Frage zu stellen. Die Frage, wie wichtig singmotorische Gedächtnisinhalte für den Tonartabruf sind, erfordert demnach noch weitere Untersuchungen.

Aufschlussreich sind dabei auch Untersuchungen aus der Arbeitsgedächtnis- und „Imagery“-Forschung, die auditive Vorstellungen unter dem Aspekt untersuchen, wie inneres Sprechen / Singen (Subvokalisation) und inneres Hören z.B. bei der Vorstellung von Sprache zusammenarbeiten. Experimente zu diesem Thema (Überblick bei Smith, Reisberg & Wilson, 1992) bedienen sich der Interferenzmethodik (vgl. Kap. 1), bei der selektiv entweder die phonologische Speicherkomponente oder die artikulatorische Komponente der phonologischen Schleife (im Sinne des Modells von Baddeley & Hitch, 1974) gestört werden. Ergebnisse von Interferenz-Experimenten weisen darauf hin, dass für Tonvergleiche bei bekannten Melodien sowohl die phonologische als auch die artikulatorische Komponente notwendig ist (Minderleistung bei beiden Arten von Interferenz). Für die Vorstellung von Klangfarben dagegen ist die phonologische mehr als die artikulatorische Komponente notwendig (keine Interferenz durch Subvokalisation) (Hespos, 1989; Smith et al., 1992). Aus der Tatsache, dass es auch Aufgaben wie z.B. die Beurteilung der Homophonie von Wörtern gibt, die durch beide Arten der Interferenz nicht beeinträchtigt werden, folgern Smith et al. (1992), dass es eine zweite phonologische Speicherkomponente geben könnte, die sie „lexikalisches Ohr“ nennen. Nach Besner (1987) kann das lexikalische Ohr z.B. ganze Wörter aus dem LZG abrufen, ohne die abgerufene Information jedoch manipulieren zu können. Smith et al. (1992) vermuten, dass auch gelernte Lieder durch das lexikalische Ohr aus dem LZG abgerufen werden können und dass demnach ein Abruf, der keine weitere Manipulation (z.B. Tonvergleiche) erfordert, interferenzresistent sein müsste. Eine experimentelle Prüfung dieser Hypothese auf der Verhaltensebene steht noch aus.

Experimente mit bildgebenden Verfahren stützen die Bedeutung singmotorischer Gedächtnisinhalte bei der Manipulation musikalischer Information im Arbeitsgedächtnis, nicht aber beim bloßen Abruf bekannter Melodien. Zatorre, Halpern, Perry, Meyer & Evans (1996) ließen ihre Vpn entscheiden, ob das Intervall zwischen den Tönen zweier Wörter einer vorgestellten bekannten Melodie auf- oder absteigend ist und konnten mittels PET u.a. eine Aktivierung des supplementären motorischen Kortex nachweisen. Da diese Region eine wichtige Rolle bei motorischen Aufgaben wie z.B. der Produktion von Sprache spielt, interpretieren

die Autoren die Aktivierung als Hinweis für die Notwendigkeit von Subvokalisation zur Lösung solcher Tonvergleiche. In einem weiteren Experiment spielten Halpern & Zatorre (1999) ihren Vpn die Anfänge bekannter Melodien vor, deren Fortsetzung sie sich vorstellen sollten, eine Aufgabe, die den Abruf aus dem LZG erfordert. In einer Kontrollaufgabe wurden den Vpn neue Tonsequenzen vorgespielt, und in einer zweiten Kontrollaufgabe sollten die Vpn sich die vorgespielten Tonsequenzen nochmals vorstellen. Durch Subtraktion der durch PET gemessenen Aktivierung beim Hören der neuen Tonsequenzen von der Vorstellung derselben wollten die Autoren Prozesse des tonalen Arbeitsgedächtnisses sichtbar machen. Durch Subtraktion der Aktivierung bei der Vorstellung neuer Tonsequenzen von der fortgesetzten Vorstellung bekannter Melodien sollte dagegen der Prozess des Abrufs aus dem LZG sichtbar gemacht werden. Die Ergebnisse zeigen, dass der supplementäre motorische Kortex bei der arbeitsgedächtnisbasierten Aufgabe aktiv war, nicht jedoch beim Abruf aus dem LZG. In einer weiteren Studie mit funktioneller MRT konnten Halpern, Zatorre, Bouffard & Johnson (2002) zeigen, dass es auch arbeitsgedächtnisbasierte Prozesse gibt, die ohne Subvokalisation auszukommen scheinen, denn bei einer Vergleichsaufgabe von Klangfarben war der supplementäre motorische Kortex nicht aktiviert. Dadurch wird das oben berichtete Ergebnis von Hespos (1989) bestätigt. Insgesamt verweisen die Ergebnisse darauf, dass Besners (1987) Annahme eines lexikalischen Ohrs korrekt sein könnte, zumindest was die Nichtbeteiligung der singmotorischen Komponente beim Melodieabruf aus dem LZG betrifft. Die PET-Ergebnisse von Halpern & Zatorre (1999) zeigten eine Aktivierung des sekundären, nicht aber des primären auditorischen Kortex beim LZG-Abruf, ein Ergebnis, mit dem man das Vorhandensein inneren Hörens nach dem derzeitigen Forschungsstand weder widerlegen noch bestätigen kann. Es sind also weitere Untersuchungen erforderlich, bevor man sich nach einem allgemeingültigeren Namen für das „lexikalische“ Ohr umsieht, der die Generalisierung auf Melodien sichtbar machen würde.

Eine weitere Komponente des motorischen Gedächtnisses ist die Erinnerung an *motorische Kontextinformationen*, die sich nicht auf ein inneres Singen von Tönen beziehen. Ein Beispiel für motorische Kontextinformationen ist die spielmotorische Erinnerung von Instrumentalmusikern im Sinne eines mit bestimmten Tönen assoziierten „Spielgefühls“, z.B. die Position der Hände, Zunge oder Lippen bei Spielern von Blasinstrumenten. Ein weiteres Beispiel wäre die Erinnerung an typische Hörumgebungen und damit assoziierte Körperhaltungen (u.a. Tanzbewegungen) bei den Sängern von Lieblings-Popsongs, die vermutlich eng mit der Erinnerung an typische emotionale Zustände beim Singen oder Hören bestimmter Musikstücke verknüpft ist. Eine Grundlage für die Annahme, dass motorische Kontextinformationen für den Abruf von musikalischer Information aus dem LZG eine Rolle spielen könnten, bilden Experimente mit bildgebenden Verfahren, die Wahrnehmungs- und Vorstellungsprozesse vergleichen. Dabei konnte gezeigt werden, dass der supplementäre motorische Kortex nicht nur bei der realen, sondern auch bei der vorgestellten Ausführung von Bewegungen aktiviert ist (Roland, Larsen, Lassen & Skinhoj, 1980; Rao, Binder, Bandettini, Hammeke, Yetkin, Jesmanowicz, Lisk, Morris, Mueller, Estkowski, Wong, Haughton & Hyde, 1993). Dies spricht dafür, dass man sich sogenannte „motor plans“ auch vorstellen kann und dass die Vorstel-

lung der Wahrnehmung ähnelt, ähnlich wie es auch für inneres Sprechen gezeigt werden konnte (Paulesu, Frith, Frackowiak, 1993; Wise, Chollet, Hadar, Friston, Hoffner & Frackowiak, 1991).

Als vorläufiges Fazit der Betrachtung motorischer Gedächtnisinhalte lässt sich festhalten, dass diese wichtige Kontextinformationen für die Vorstellung von Melodien liefern können, ohne dass es auf der Grundlage bisheriger Ergebnisse plausibel erscheint, dass ein LZG-Abruf von Melodien *alleine* auf dem Abruf eines „motor plan“ beruhen kann.

Aufgrund der Betrachtung verschiedener Prozesse, die am nichtdeklarativen Erinnern beteiligt sind, lassen sich zusammengefasst folgende Schlussfolgerungen für die nichtdeklarative Tonarterinnerung ziehen:

Ein langfristiger Primingeffekt wurde bezogen auf Tonarten bisher nicht untersucht. Gelänge der Nachweis eines solchen Effekts bei musikalischen Laien, so wäre dies ein Hinweis auf eine langfristige implizite Tonarterinnerung und somit eine Dissoziation impliziter und expliziter Gedächtnisprozesse für das Tongedächtnis. Allerdings wären außer einer Dissoziation auf der Verhaltensebene auch neuropsychologische Hinweise erforderlich, um eine Trennung von impliziten und expliziten Tonerinnerungsprozessen zu untermauern.

Der *imprinting*-Mechanismus des Wahrnehmungslernens könnte eine Erklärung für die Tonarterinnerung bei extrem oft gehörten Tönen, die zudem eine Art Signalcharakter haben, bieten. Dieser Mechanismus führt zu einer ähnlichen Hypothese wie sie aus der Betrachtung deklarativer Lernprozesse resultierte, nämlich, dass die Häufigkeit, mit der bestimmte Töne oder Melodien gehört werden, eine Auswirkung auf die Genauigkeit der Tonartrepräsentation haben sollte. Im Unterschied zu expliziten Lernprozessen, die eine Kategorienbildung ermöglichen, wird durch Wahrnehmungslernen aber eine punktuelle Genauigkeit der Tonrepräsentation vorhergesagt.

Motorische Fertigkeiten, die für die Erinnerung an einzelne Töne eine Rolle spielen könnten, sind das singmotorische Tongedächtnis bei hochtrainierten Sängern bzw. analog ein spielmotorisches Gedächtnis bei geübten Spielern von Musikinstrumenten.

Inwieweit motorische Gedächtnisinhalte im weiteren Sinne, also über spezielle Fertigkeiten hinaus, für den Tonartabruf von musikalisch untrainierten Personen eine Rolle spielen, erfordert weitere Untersuchungen. Die Ergebnisse der Arbeitsgedächtnis- und Imagery-Forschung ermutigen zu der Hypothese, dass für den bloßen Abruf von Melodien das singmotorische Gedächtnis nicht zwingend erforderlich ist.

Die Betrachtung deklarativer und nichtdeklarativer Gedächtnisprozesse verdeutlicht, dass sich aus vorhandenen gedächtnispsychologischen Befunden eine Reihe von Hypothesen für die Tonarterinnerung ableiten lassen. Welche der genannten Hypothesen in der vorliegenden Arbeit experimentell untersucht werden, wird im folgenden Kapitel dargestellt.

5 Fragestellung & Hypothesen der Untersuchungen

Da nicht alle in den bisherigen Kapiteln aufgeworfenen Fragestellungen und Hypothesen experimentell überprüft werden konnten, wurde in der vorliegenden Arbeit ein Schwerpunkt auf die Untersuchung der Tonartrepräsentation von Nichtabsoluthörern mithilfe von expliziten Methoden gelegt. Diese Schwerpunktsetzung erfolgte aus zwei Gründen:

Zum Einen erlauben die bislang vorliegenden Ergebnisse zur Tonartproduktion und –wiedererkennung noch keine Aussage darüber, wie verbreitet Tonartrepräsentationen in der Allgemeinbevölkerung tatsächlich sind. Ein erstes Ziel der Arbeit war es daher, die Generalisierbarkeit von Levitins (1994) Ergebnissen auf andere Versuchspersonen und Melodien zu überprüfen. Bei der Auswahl des experimentellen Paradigmas fiel die Wahl auf Levitins Methode (Messung der *Genauigkeit* von Produktionen) und nicht auf die Auswertung der Tonart*konstanz*, da in letzterer die Einflüsse einer konstant bevorzugten mittleren Stimmlage viel schwerer kontrollierbar sind. Außerdem wurde die Produktionsmethodik der Wiedererkennungsmethodik vorgezogen, da sie eine flexiblere Auswahl und somit eine größere Vielfalt von abzurufenden Melodien erlaubt. In zwei Experimenten wurden darüber hinaus Wiedererkennungsmethoden eingesetzt, um einen Methodenvergleich zu ermöglichen. Während in vielen vorliegenden Experimenten außerdem musikalisch nicht oder nur wenig vorgebildete Vpn untersucht worden sind, war es ein Anliegen dieser Arbeit, auch musikalisch ausgebildete Vpn auf ihr Tonartgedächtnis zu untersuchen, und zwar sowohl für inzidentell gelernte Melodien wie z.B. Werbejingles als auch für intentional gelernte Melodien wie z.B. Chormusik, und ihre Leistungen mit der von Nichtmusikern zu vergleichen. Da Nullergebnisse im Allgemeinen nicht publiziert werden, sollte in dieser Arbeit auch der Frage nachgegangen werden, unter welchen Umständen sich Vpn nicht bzw. weniger gut an Tonarten erinnern können. Kurz zusammengefasst ging es bei der Analyse der Verbreitung von Tonartrepräsentationen also darum festzustellen, *wer* sich unter *welchen* Bedingungen *wie* gut an *welche* Art von Melodien erinnern kann.

Da die Analyse der Verbreitung von Tonartrepräsentationen eine Untersuchung verschiedener Stichproben, die unterschiedliche Arten von Melodien auf unterschiedliche Weise gelernt haben, beinhaltet, wird dadurch auch die Analyse von Einflussfaktoren auf das Tonartgedächtnis möglich. Diese Möglichkeit der Identifikation von Einflussfaktoren auf die Tonarterinnerung von Nichtabsoluthörern war der zweite Grund für die genannte Schwerpunktsetzung dieser Arbeit. Ein wichtiges Ziel war es, Einflüsse von aus der allgemeinen Gedächtnisforschung abgeleiteten Faktoren auf die Tonarterinnerung von Nichtabsoluthörern zu überprüfen. Können solche Einflüsse nachgewiesen werden, so wäre dies ein Hinweis darauf, dass die Tonarterinnerung von Nichtabsoluthörern gewissermaßen ein „normaler“ Gedächtnisprozess ist. Andererseits sollte aber auch überprüft werden, inwieweit Zusammenhänge zwischen der hochspezialisierten Fähigkeit, Töne zu benennen, und der Tonarterinnerung von Nichtabsoluthörern vorliegen, die die bislang gängige Bezeichnung „latentes absolutes Gehör“ rechtfertigen würden. Im Einzelnen sollten folgende Hypothesen geprüft werden:

5.1 Hypothesen zur Tonarterinnerung von Nichtabsoluthörern

Aus den in Kap. 4.3 beschriebenen Befunden zur Speicherung und zum Vergessen gelernter Information lässt sich die Hypothese ableiten, dass die Häufigkeit, mit der Melodien in gleicher Tonart gehört werden, einen positiven Zusammenhang mit der Genauigkeit des Tonartabrufs aufweist. Da die bisher vorliegenden Produktions-Experimente von Levitin (1994) und Bergeson & Trehub (2002) mit seit sehr langer Zeit bekannten Melodien arbeiten, erlauben die Ergebnisse keine Rückschlüsse über einen solchen Zusammenhang. Die Überprüfung dieser Hypothese erfolgte in der vorliegenden Arbeit über den Vergleich der Leistungen unterschiedlicher Stichproben beim Tonartabruf gelernter Melodien, die über unterschiedlich lange Zeitdauern hinweg bzw. mit unterschiedlicher Übungshäufigkeit gelernt worden waren.

Hypothese 1: Häufigkeitseffekt

Die Tonarten von Melodien, die entweder über längere Lernphasen oder mit größerer Häufigkeit gehört, gesungen oder gespielt worden sind, werden besser abgerufen als die Tonarten von Melodien, die über kürzere Lernphasen bzw. seltener gehört, gesungen oder gespielt worden sind.

Aus den in Kap. 4.3 beschriebenen Befunden zur Enkodierung von Information lässt sich weiterhin die Hypothese eines Expertiseeffekts ableiten, da sich Enkodierungsprozesse als stark vorwissensabhängig erwiesen haben. Bisher vorliegende Befunde zu Expertiseeinflüssen auf die Tonarterinnerung (Levitin, 1994; Schellenberg & Trehub, 2003) konnten keine solchen Einflüsse finden. Diese Ergebnisse sind jedoch aufgrund der bezüglich musikalischer Expertise homogenen und insgesamt nicht sehr großen Stichproben in ihrer Aussagekraft beschränkt. In der vorliegenden Arbeit erfolgte die Überprüfung des angenommenen Expertiseeffekts sowohl über den Vergleich der Abrufleistung verschiedener Vpn innerhalb einer Stichprobe als auch über den Vergleich der mittleren Abrufleistungen verschiedener Stichproben. Der Intra-Stichproben-Vergleich hat den Vorteil, dass die Abrufleistungen bei gleichen oder zumindest vergleichbaren Melodien verglichen werden, aber den Nachteil, dass die Expertise in den meisten Stichproben recht homogen ist. Den Vergleich der Auswirkungen größerer Expertiseunterschiede erlaubt der Inter-Stichproben-Vergleich, der dafür den Vergleich unterschiedlicher Melodien in Kauf nehmen muss.

Hypothese 2: Expertiseeffekt

Die Tonarten gelernter Melodien können von Vpn mit einem reicheren musikalischen Erfahrungshintergrund besser abgerufen werden als von Vpn mit weniger musikalischer Vorerfahrung.

Im Zusammenhang mit dem erwarteten Expertiseeffekt ist auch die Tonbenennungsleistung zu erwähnen. Aus bisherigen Untersuchungen zum Tonartgedächtnis lässt sich bezüglich eines Zusammenhangs zwischen Tonartabruf und Tonbenennung keine Aussage treffen, da in keinem der in Kap. 3.1 beschriebenen Experimente zum „latenten“ absoluten Gehör die Fähigkeit der Vpn zur Benennung von Tönen überprüft wurde. In Anbetracht der theoretischen Überlegungen (Kap. 4.3.1) kann man annehmen, dass sich auch die in verschiedenen

Ausprägungen vorhandene Fähigkeit, Töne zu kategorisieren, als ein Aspekt musikalischer Expertise auf Enkodierungsprozesse auswirkt. Somit lässt sich auch für die Tonbenennungsleistung ein Zusammenhang mit der Tonarterinnerung annehmen.

Hypothese 3: Tonbenennungseffekt

Die Tonarten gelernter Melodien können von Vpn, die Töne besser benennen können, besser abgerufen werden als von Vpn, die Töne weniger gut oder gar nicht benennen können.

Aufgrund der in Kap. 4.4.5 beschriebenen Befunde zu motorischen Gedächtnisinhalten (Murry, 1990; Ward & Burns, 1987; Kestler, 1999; Halpern & Zatorre, 1999) wird angenommen, dass der Tonartabruf nicht alleine auf der Nutzung singmotorischer Gedächtnisinhalte beruht. Da bislang keine Untersuchungen vorliegen, die gezielt den Beitrag singmotorischer Gedächtnisinhalte auf den Abruf von Tonarten aus dem LZG untersuchen, erfolgt die Überprüfung dieser Hypothese in der vorliegenden Arbeit durch zwei Methoden. Zum Einen wird ein Experiment mit Interferenzmethodik (vgl. Kap. 1) durchgeführt. Zum Anderen wurde die Abrufleistung von durch Singen und durch Spielen eines Musikinstruments gelernten Melodien verglichen. Es wird angenommen, dass Instrumentalmusiker bei der Produktion von Instrumentalmelodien weniger gut auf singmotorische Gedächtnisinhalte zurückgreifen können als Chorsänger bei der Produktion von Vokalmelodien. Schlechtere Produktionsleistungen der Instrumentalmusiker (im Vergleich zu Chorsängern) wären demnach ein Hinweis darauf, dass der Tonartabruf die Nutzung singmotorischer Gedächtnisinhalte erfordert.

Hypothese 4: Motorisches Gedächtnis

4 a) Der Tonartabruf wird durch singmotorische Interferenzaufgaben nicht verschlechtert, da er nicht vorrangig auf der Nutzung singmotorischer Gedächtnisinhalte beruht.

4 b) Die Tonarten von Melodien, die durch Singen gelernt worden sind, können nicht besser produziert werden als die Tonarten von Melodien, die nicht durch Singen, sondern durch das Spielen von Instrumenten (das gleichzeitiges Singen ausschließt) gelernt worden sind.

5.2 Hypothesen zur Tonerinnerung von Absolut- und Nichtabsoluthörern

Um zu zeigen, dass der postulierte Häufigkeitseffekt einen allgemeingültigen Gedächtnismechanismus darstellt, der nicht auf die Tonarterinnerung beschränkt ist, wurde überprüft, ob er sich auch bei der Tonbenennungsleistung von Absolut- und Nichtabsoluthörern nachweisen lässt. Es wurde also untersucht, ob Töne, die in der persönlichen Lerngeschichte einer Vp häufiger aufgetreten waren, richtiger und schneller benannt werden können als Töne, die seltener gehört bzw. gelernt worden sind. Vorliegende Untersuchungen von Miyazaki (1989, 1990), Heyde (1987) und Takeuchi & Hulse (1991) bestätigen einen solchen Effekt auf der Verhaltensebene bei Absoluthörern, während Marvin & Brinkman (2000) ihn auch bei Nichtabsoluthörern nachweisen. In der vorliegenden Arbeit wird der Nachweis eines Häufigkeitseffekts bei Absolut- und Nichtabsoluthörern mit einer neuen Methode, der Pupillometrie, vorgenommen. Diese Methode hat sich für die Untersuchung des kognitiven Aufwands in verschiedensten Aufgaben bewährt (vgl. Kap. 8) und wurde in der vorliegenden Arbeit im Rah-

men eines Pilot-Experiments angewandt, um Unterschiede in der Ressourcenkonsumption abzubilden, die aus Verhaltensdaten nicht ablesbar sind.

In dieser Arbeit wird angenommen, dass ähnliche Lernprozesse bei Absoluthörern und Nichtabsoluthörern eine Rolle spielen, die aber in unterschiedlichen Altersstufen wirksam werden und daher zu unterschiedlichen Lernergebnissen bezüglich der Tonerinnerungsleistung führen (vgl. Kap. 3.2). Dies führt zu der Annahme, dass ähnliche Effekte bezüglich der Tonklasse und Klangfarbe gehörter Töne bei Absolut- und Nichtabsoluthörern auftreten, wenngleich bei Absoluthörern auf insgesamt höherem Leistungsniveau als bei Nichtabsoluthörern.

Hypothese 1a: Häufigkeitseffekt

Töne, die aufgrund ihrer Zugehörigkeit zu bestimmten Tonklassen oder aufgrund ihrer Klangfarbe in der persönlichen Lerngeschichte häufiger gehört worden sind, können richtiger, schneller und aufwandsärmer benannt werden als Töne, die aufgrund derselben Faktoren seltener gehört worden sind.

Die genannten Hypothesen wurden in einer Reihe von Experimenten überprüft. Der folgende Abschnitt gibt einen Überblick über die Schwerpunkte der einzelnen Experimente und den Aufbau der folgenden Kapitel.

5.3 Überblick über die Experimente[1]

Um die Verbreitung von Tonartrepräsentationen zu untersuchen und die Hypothesen 1 bis 4 zu überprüfen, wurden sechs Stichproben mit Produktionsanforderungen untersucht. Dabei wurden in einigen Stichproben mehrere Melodien produziert. Um Hypothese 1 zu überprüfen, wurden als musikalisches Material sowohl Melodien, die den Vpn seit längerer Zeit bekannt waren, als auch erst seit kürzerer Zeit gelernte Melodien genutzt. Um Hypothese 2 zu überprüfen, wurden musikalische Laien, musikalisch interessierte, aber wenig ausgebildete Schüler, semiprofessionelle Chorsänger und professionelle Instrumentalmusiker untersucht. Die Variation der musikalischen Expertise bewirkte gleichzeitig eine Variation der Fähigkeiten der Tonbenennung, die für die Überprüfung der Hypothese 3 wichtig war. Um Hypothese 4 zu überprüfen, wurden sowohl intentional gelernte Vokalmelodien als auch intentional gelernte Instrumentalmelodien genutzt.

Die Ergebnisse sämtlicher Produktions-Experimente werden in Kapitel 6 beschrieben. Im einzelnen wurden folgende Experimente durchgeführt: Eine Stichprobe musikalischer Laien produzierte ihre persönlichen Lieblingslieder (Kap. 6.2), während eine weitere Stichprobe musikalischer Laien ihre Handymelodien produzierte (6.7). Musikalisch interessierte Schüler produzierten über drei Monate intentional gelernte Melodien (6.3) sowie einen Werbejingle (6.6). Eine Stichprobe semiprofessioneller Chorsänger produzierte über drei Monate intenti-

[1] Alle in der vorliegenden Arbeit verwendeten Versuchsmaterialien werden auf Anfrage von der Autorin gerne zur Verfügung gestellt.

onal gelernte Melodien (6.4), während eine weitere Stichprobe semiprofessioneller Chorsänger über mehrere Jahre intentional gelernte Melodien (6.5) produzierte. Professionelle Instrumentalmusiker produzierten über mehrere Jahre intentional gelernten Instrumentalmelodien (6.8).

Nach der Darstellung sämtlicher Produktions-Ergebnisse wird in Kapitel 6 der Versuch einer zusammenfassenden Analyse aller Produktionsdaten beschrieben, und es werden die Hypothesen 1 bis 3 geprüft.

Da der Häufigkeitseffekt, der den Hypothesen 1 und 1a zugrunde liegt, für die vorliegende Arbeit von zentraler Bedeutung ist, wurde er in zwei Experimenten gesondert untersucht, die in den Kapiteln 7 und 8 beschrieben werden. In Kapitel 7 geht es zunächst um die Stichprobe, die bereits in Kapitel 6.3 und 6.6 beschrieben wurde. Mit den dort untersuchten Schülern wurde eine gezielte Variation der Lernintensität neu gelernter Melodien vorgenommen, deren Effekt auf die Produktions- und Wiedererkennungsleistungen überprüft wurde. In Kapitel 8 wird der Fokus der Arbeit vom Tonartabruf durch Nichtabsoluthörer auf die Tonbenennung erweitert. Dies ist das einzige Experiment dieser Arbeit, in dem Absoluthörer und Nichtabsoluthörer explizit verglichen werden. Dabei wird vermutet, dass Hörhäufigkeitseffekte sich bei beiden Gruppen auf die Tonbenennungsleistung auswirken. Die für Absoluthörer bereits mehrfach nachgewiesenen tonspezifischen Effekte, die sich vor allem auf Tonklassen und Klangfarben beziehen, sollen hier erstens auf ihre Gültigkeit auch für Nichtabsoluthörer überprüft und zweitens durch eine neue Methode, die Pupillometrie, abgesichert werden.

Ein weiterer Aspekt, der in zwei Experimenten gesondert untersucht wurde, betrifft den Beitrag motorischer Gedächtnisprozesse für den Tonartabruf. Dieser wird in Kapitel 9 dargestellt. Dabei werden zum Einen weitere Ergebnisse aus dem unter 6.8 beschriebenen Experiment mit Instrumentalmusikern dargestellt, das auch eine Wiedererkennungsanforderung beinhaltete. Um Hypothese 4 b zu überprüfen, werden die Produktionsleistungen der Instrumentalmusiker mit denen der Chorsänger verglichen. Um Hypothese 4 a zu überprüfen, werden zusätzliche Ergebnisse aus den unter 6.2 und 6.4 beschriebenen Experimenten berichtet, die Effekte von spezifischen Interferenzaufgaben auf die Geschwindigkeit und Genauigkeit des Tonartabrufs untersuchen.

In Kapitel 10 werden die Ergebnisse aller Experimente zusammengefasst und Einflussfaktoren auf das Ton- und Tonartgedächtnis diskutiert.

6 Verbreitung von Tonartrepräsentationen

In diesem Kapitel sollen überblicksartig Ergebnisse zur Verbreitung absoluter Tonhöhenrepräsentationen dargestellt werden, damit ein vollständigeres Bild davon entsteht, *wer* sich unter *welchen* Bedingungen *wie* gut an *welche* Art von Melodien erinnern kann. Daher werden in diesem Kapitel Methodik, Stichproben und abzurufende Melodien so beschrieben, dass ein Vergleich der Abrufleistungen und somit eine Prüfung der folgenden Hypothesen ermöglicht wird:

1. *Häufigkeitseffekt*: Vergleich der mittleren Abrufleistungen verschiedener Stichproben, die unterschiedliche Melodien unterschiedlich lange gelernt haben.

2. *Expertiseeffekt:* Vergleich der Abrufleistungen von Vpn mit unterschiedlicher Dauer der musikalischen Ausbildung.

3. *Tonbenennungseffekt*: Vergleich der Abrufleistungen von Vpn mit unterschiedlichen Leistungen in einer Tonidentifikationsaufgabe.

An die Beschreibung der Tonarterinnerungsleistungen in den einzelnen Experimenten schließt sich der Versuch an, die Daten aller Experimente zusammenzufassen, um die Hypothesen an der zusammengefassten Stichprobe zu untersuchen.

6.1 Methodik der Produktions-Experimente

Nachdem die Vpn in einer schriftlichen Instruktion über den Ablauf des jeweiligen Experiments informiert worden waren, wurden sie zunächst aufgefordert, sich die zu erinnernde Melodie so genau wie möglich vor ihr „inneres Ohr“ zu rufen. Handelte es sich bei der zu erinnernden Melodie um ein mehrstimmiges Chorstück, so wurden die Vpn gebeten, sich die Melodie ihrer jeweiligen Stimme vorzustellen. Sobald sie eine genaue Vorstellung der zu erinnernden Melodie generiert hatten, signalisierten sie dies der Versuchsleiterin mit einem nonverbalen Zeichen, z.B. durch Kopfnicken. Die Vpn wurden gebeten, während und nach der Vorstellungs-Generierung nicht zu sprechen, um eine mögliche Interferenz durch Sprache auszuschließen. Auch die Versuchsleiterin war während der Versuchsdurchführung vollkommen ruhig. Auf das Zeichen der Vpn hin startete sie die Aufnahme, woraufhin die Vpn mit der Produktion der jeweiligen Melodie begann. Um die meist vorhandene Scheu zu singen zu reduzieren, aber auch um eine möglichst genaue Produktion der vorgestellten Melodie zu bewirken, wurden die Vpn in der Instruktion darauf hingewiesen, dass sie möglichst genau die Töne singen sollen, die sie im Kopf haben, und keinesfalls etwas höher oder tiefer zu singen, damit es schöner klänge. Es wurde ausdrücklich betont, dass Schönklang in den Experimenten keine Rolle spiele und es nur auf eine möglichst genaue Produktion ihrer Vorstellung ankäme. Diese Bemerkung richtete sich vor allem an die Chorsänger, da sich in einer Voruntersuchung (Hahn, 2002) gezeigt hatte, dass einige Chorsänger absichtlich etwas tiefer als vorgestellt gesungen hatten, einerseits um sich zu schonen, andererseits, weil nicht klar genug instruiert worden war, dass es bei der Produktion gerade auf die Tonart ankäme.

Die Vpn produzierten meist eine Melodiephrase, also einen Abschnitt, den sie auf einen Atemzug singen konnten. In den Experimenten, in denen mehr als eine Melodie produziert werden sollte, wurde die Reihenfolge der abzurufenden Melodien über die Vpn randomisiert, um Interferenzeffekte zu minimieren.

Die Produktionen wurden auf Mini Disc aufgenommen, um spätere Tonhöhenschwankungen, wie sie bei Analogaufnahmen manchmal auftreten, zu vermeiden. Die Produktionsdaten wurden per Digitalkabel auf den PC überspielt und dort mit dem Programm Cool Edit 2000 (Syntrillium) analysiert. Vor der Analyse der Anfangstöne wurden die produzierten Melodien zunächst daraufhin überprüft, ob sie erstens wiedererkennbar und zweitens in der Tonart stabil waren. Melodien, die nicht als die gewünschte Melodie erkennbar waren (z.B. Singen einer fremden Chorstimme oder sonstige Erinnerungslücken), wurden von weiteren Analysen ausgeschlossen. Bei Melodien, die in der Tonart nicht stabil blieben, wurde sowohl der Anfangston als auch die späteren Tonartwechsel notiert. Zum Vergleich mit der Originaltonart wurde die erste erkennbare Tonart benutzt. War der Anfangston sehr kurz und dadurch schwer zu hören, so wurde einer der nächsten Töne für die Analyse ausgewählt. Die Anfangstöne wurden mit zwei Methoden analysiert. Zum einen wurde mit Cool Edit eine Fourier-Analyse des Tons durchgeführt, wobei Onset und Offset der Töne abgeschnitten wurden, da die Töne dort am meisten schwanken, so dass nur der stabilste Abschnitt der Töne analysiert wurde. Dabei trat bei einem kleinen Teil der Töne das Problem auf, dass die Obertöne relativ laut bzw. der Grundton relativ leise war, so dass das Programm Cool Edit, das jeweils die Frequenz der lautesten Komponente ausgibt, sich in diesen Fällen um eine Oktave (beim 1. Oberton) oder Oktave + Quinte (beim 2. Oberton) „irrte“. Um diese Fehler zu korrigieren, wurde während der Fourier-Analyse der Ton manuell per Stimmgabel (Kammerton A = 440 Hz) analysiert. Das Ergebnis dieser doppelten Analyse war eine Liste mit gesungenen Anfangstönen (bzw. Folgetönen), die nun mit den originalen Anfangstönen (bzw. Folgetönen) verglichen werden konnten. Als originale Anfangstöne wurden im Fall von Chor- und Instrumentalmelodien die in der Partitur notierten Töne definiert. Bei Melodien, für die keine Notation vorliegt, z.B. Popsongs oder Handymelodien, wurden die Anfangstöne der Originalmelodien auf die gleiche Art analysiert wie die gesungenen Anfangstöne.

Um die gesungenen Anfangstöne mit den originalen vergleichen zu können, wurde zwischen diesen beiden die Differenz in Halbtönen bestimmt. Eine kleinere Maßeinheit (z.B. Cents) zu nutzen wurde für nicht sinnvoll befunden, da viele der gesungenen Töne nicht auf den Cent stabil bleiben, sondern durch Vibrato in der Tonhöhe um bis zu einen Viertelton schwanken. Außerdem hat sich der Halbton als kleinste musikalische Einheit für die Tonhöhe in der westlichen Musik durchgesetzt und stellt damit für diese Aufgabe eine angemessene Einheit dar.

Aus den Differenzen wurden zwei Verteilungen berechnet: Für Verteilung 1 wurde bei der Differenzbildung die Richtung der Abweichung mit beachtet, so dass man in der Verteilung danach differenzieren kann, wie viele Vpn zu hoch bzw. zu tief gesungen haben. Für Verteilung 2 wurde die Differenz ohne Vorzeichen notiert, um die mittlere Abweichungsgröße unabhängig davon, ob zu tief oder zu hoch gesungen wurde, berechnen zu können. In einigen

Experimenten wurde die Oktavlage, in der die Vpn sangen, mit bewertet. Dies war immer dann der Fall, wenn Chorlieder produziert werden sollten: Diese Lieder sollten genau in der Lage produziert werden, in der sie auch geübt worden waren, und es gab keinen plausiblen Grund dafür, sie im Experiment nach oben oder unten oktaviert wiederzugeben. Etwas anders sah es bei den Handymelodien, dem Werbejingle, den persönlichen Lieblingsliedern aus der „Unterhaltungsmusik“ und den instrumentalen Melodien aus. Die Originalmelodien lagen dort in vielen Fällen außerhalb (meist oberhalb) des Stimmumfangs der Vpn und hätten demnach im Experiment nicht in der Originallage gesungen werden können. Daher wurde in diesen Fällen das von Levitin (1994) genutzte Vorgehen gewählt und die Oktavlage, in der die Vpn sangen, nicht mit bewertet.

Um die Ergebnisse der Produktionsaufgabe auf Signifikanz zu überprüfen, wurde zunächst (in Anlehnung an Levitin, 1994) Verteilung 1 mit einem chi²-Test auf Gleichverteilung überprüft. Eine Gleichverteilung der Differenzen auf alle betrachteten Abweichungs- oder Fehlerkategorien würde bedeuten, dass eine zufällige Tonartproduktion vorliegt. Um die Ergebnisse aller Experimente vergleichen zu können, wurde für diese Analyse eine einheitliche Anzahl von Fehlerkategorien festgelegt. Da man sich beim Versuch, eine bestimmte Tonart zu treffen, innerhalb einer Oktave um maximal 6 Halbtöne nach oben oder unten irren kann, so dass es inklusive der Kategorie „Fehler = 0“ 13 Fehlerkategorien gibt, wurde immer gegen eine Verteilung von -6 bis +6 getestet. Eine solche Verteilung lag bei allen Experimenten vor, in denen Oktavlagenfehler nicht mit bewertet worden waren (Handymelodien, Werbejingle, persönliche Lieblingslieder, Instrumentalmelodien). In den Experimenten mit Chormelodien, bei denen die Verteilung 1 auch Oktavlagenfehler berücksichtigt hatte, wurde diese für die Berechnung des chi²-Tests umgewandelt in eine Verteilung ohne Oktavlagenfehler. Zu diesem Zweck wurden alle negativen Abweichungen, die größer als -6 waren, eine Oktave nach oben transponiert (aus -7 wird +5, aus -8 wird +4, etc.). Signifikante Tonarterinnerung liegt dann vor, wenn die Fehler nicht gleichverteilt sind, sondern sich systematisch um die Kategorie „Fehler = 0 Halbtöne“ verteilen. Nach Levitin (1994) sind diese Art von Abweichungs-Verteilungen als zirkuläre Daten zu behandeln und entsprechende Testverfahren anzuwenden. Der chi²-Test wird von Fisher (1993) als angemessenes Verfahren zur Überprüfung von gruppierten zirkulären Daten auf Gleichverteilung gegen eine nicht spezifizierte Alternativhypothese beschrieben, allerdings nur, wenn die erwarteten Häufigkeiten pro Kategorie größer als 2 sind (Fisher, 1993, S. 67f). Waren die erwarteten Häufigkeiten kleiner als 2 (bei Stichproben mit $n < 26$), so wurde approximativ ein Kolmogorov-Smirnov-Test auf Gleichverteilung durchgeführt, wie von Bortz & Lienert (1998, S. 218) empfohlen. Um bei signifikantem chi²-Test zu prüfen, ob die Verteilung einer Normalverteilung genügt, wurde als nächstes Verteilung 1 mit dem Kolmogorov-Smirnov-Test auf Normalverteilung getestet.

Neben der Signifikanzprüfung anhand der Verteilungsanalysen, die sich an Levitin (1994) anlehnt, wird hier noch ein weiteres Signifikanzkriterium vorgeschlagen. Diese weitere Analyse baut auf folgender Argumentation auf: Sind die Daten nicht gleichverteilt, sondern normalverteilt, so können dennoch die Fehler so breit streuen, dass zwar nicht von einer zufälligen, aber doch von einer sehr ungenauen Tonartproduktion ausgegangen werden muss.

Diese zweite Analyse richtet sich also bei nachgewiesener Nicht-Gleichverteilung der Fehlerverteilungen auf die Genauigkeit der Tonartproduktion. Zu diesem Zweck wurde der mittlere Fehler der Verteilungen berechnet. Da es bei dieser Analyse auf die Fehler*größe* (= Anzahl an Halbtönen zwischen Original und Produktion) und nicht auf die Fehler*richtung* (= Abweichung nach oben oder nach unten) ankam, erfolgte die Fehlermittelung anhand von Verteilung 2. Mittelt man die Fehler aus Verteilung 1, so erhält man auch bei sehr großen Abweichungen nach oben und unten immer dann einen Mittelwert nahe „0“, wenn die Abweichungen sich symmetrisch nach oben und unten verteilen (bzw. wenn einige große positive Abweichungen viele kleinere negative Abweichungen „ausgleichen“); es könnte dadurch also künstlich der Eindruck eines sehr kleinen Durchschnittsfehlers hergestellt werden. Diese Gefahr besteht bei der Mittelung vorzeichenloser Fehlergrößen nicht, so dass die Nutzung der vorzeichenlosen Verteilung 2 für eine Betrachtung der Fehlergrößen vorzuziehen ist.

Betrachtet man aufgrund der Vergleichbarkeit der verschiedenen Experimente auch für Verteilung 2 nur die Fehler innerhalb einer Oktave (und transponiert analog dem oben beschriebenen Vorgehen größere Fehler in diese Oktave), so können diese von 0 bis 6 variieren. Diese Tatsache wurde für die Festlegung eines „Zufallswertes“ für die Signifikanzprüfung genutzt: Bei zufälliger Tonartproduktion wäre zu erwarten, dass sich die vorzeichenlosen Differenzen gleichmäßig auf die 7 Kategorien von Verteilung 2 verteilen, so dass sich eine mittlere Abweichung von 3 (= (0+1+2+3+4+5+6) / 7) ergäbe. Um zu überprüfen, ob sich die mittleren vorzeichenlosen Differenzen signifikant vom Mittelwert 3 unterscheiden, wurden die Daten der Verteilung 2 mit einem Einstichproben t-Test geprüft. Der Test wurde zweiseitig durchgeführt, da meines Wissens in der berichteten Literatur keine Differenz berichtet wird, die sicher auf absolutes Tongedächtnis verweist. Die Voraussetzung der Normalverteilung der Daten wurde durch den Kolmogorov-Smirnov-Test in allen Experimenten überprüft.

6.2 Ausgangspunkt Levitin (1994): Erinnerung an persönliche Lieblingslieder

6.2.1 Stichprobe

Untersucht wurden 30 Vpn (15 weiblich, 15 männlich, Durchschnittsalter 35,4 Jahre) mit keiner oder wenig musikalischer Vorbildung (durchschnittliche Instrumentalerfahrung 2,1 (SD = 3,3) Jahre). Sie wurden aus dem Bekanntenkreis der Versuchsleiterin rekrutiert und nahmen am Experiment freiwillig und ohne Bezahlung teil.

6.2.2 Produzierte Melodien

Produziert werden sollten die Lieblingslieder der Vpn, die sie entweder seit langer Zeit durch das Hören von CDs oder Schallplatten kannten oder in letzter Zeit häufig gehört hatten. Bei diesen Liedern handelte es sich teilweise um Evergreens z.B. von den Beatles oder ABBA, teilweise aber auch um neuere Lieder, z.B. von der ein halbes Jahr vor dem Experiment erschienenen CD „Mensch“ von Herbert Grönemeyer, und teilweise um Kinderlieder, die von Vpn mit Kindern fast täglich gehört wurden. Zu den Kinderliedern sei angemerkt, dass es

sich nicht um mündlich tradierte Lieder wie z.B. „Hänschen klein“ handelte, sondern um Lieder, die auf Tonträgern verbreitet und somit zumindest vorrangig in einer bestimmten Tonart gehört werden. Musikalisch ist das hier gesungene Liedmaterial mit den bei Levitin (1994) produzierten Melodien vergleichbar, es handelt sich in beiden Fällen um leicht singbare Unterhaltungsmusik, deren Tonart durch die Referenzeinspielung eindeutig festgelegt ist. Von der Vertrautheit her ist das Material jedoch nicht in allen Fällen mit Levitins (1994) vergleichbar, denn die Dauer der Bekanntheit variierte von einigen Monaten bis 35 Jahre (m = 10 Jahre, SD = 10 Jahre). 30 % der Vpn kannten ihr Lied erst höchstens ein Jahr, weitere 20 % kannten es 2 - 6 Jahre, während nur 50 % der Vpn ihr Lied seit über 10 Jahren kannten. Fraglich ist demnach vor allem, ob die 30 % der Vpn, die erst seit kürzerer Zeit intensiv gehörte Lieder sangen, eine vergleichbar gute Tonart-Repräsentation der Lieder aufweisen wie Levitins (1994) Probanden. Im Unterschied zu Levitin (1994), der die eingesetzten Popsongs vor dem eigentlichen Experiment einer Standardisierung unterzogen hatte (vgl. Kap. 3.1.2), waren die hier produzierten Melodien also nicht hinsichtlich ihrer Bekanntheit standardisiert und als Evergreens definiert, sondern es wurde den Vpn freigestellt, welches ihrer Lieblingslieder sie singen durften. Dieses Vorgehen wurde gewählt, um den Vpn, die die Aufgabe als sehr schwierig empfanden, zu ermöglichen, Lieder zu singen, die sie am besten kennen, auch wenn dies erst seit kürzerer Zeit der Fall war.

6.2.3 Besonderheiten bei der Versuchsdurchführung

Da in diesem Experiment die Bedingungen, unter denen sich die Vpn an ihre Lieblingslieder erinnern sollten, variiert und dabei auch Reaktionszeiten erfasst wurden (vgl. Darstellung dieses Experiments in Kap. 9), fand die Versuchssteuerung am PC statt und unterschied sich damit vom Versuchsablauf der übrigen Experimente. Die Vpn wurden auf dem Bildschirm angewiesen, sich eine Phrase aus ihrem Lieblingslied vorzustellen und zunächst eine Taste der Tastatur zu drücken, sobald sie eine genaue Vorstellung von dieser Phrase hatten. Erst nach dem Tastendruck wurden die Vpn durch eine Instruktion auf dem Bildschirm angewiesen, die Phrase zu singen. Diese doppelte Reaktionsanforderung (1. Tastendruck, 2. Produktion) hat einige Vpn nach eigenen Angaben etwas verunsichert. Von daher ist dieses Experiment vom Ablauf her nicht vollständig vergleichbar mit den übrigen Experimenten oder mit Levitins (1994) Experiment.

6.2.4 Besonderheiten bei der Datenauswertung

Da die hier produzierten Lieder nicht durch Singen, sondern durch häufiges Hören von Tonträgern gelernt wurden und viele außerdem außerhalb des Stimmumfangs der Vpn lagen, z.B. wenn ein Mann versuchte, das Lied einer weiblichen Solistin nachzusingen, wurden Oktavlagenfehler in diesem Experiment nicht als Fehler bewertet. Bei Levitin (1994) hatte sich gezeigt, dass seine Vpn die Melodien, die nicht in ihrem Stimmumfang lagen, oktaviert hatten. Da es sich hier um eine vergleichbare Stichprobe und vergleichbare Melodien handelt,

erscheint ein analoges Vorgehen gerechtfertigt. Das Ergebnis waren maximale Abweichungen von der Original-Tonart um 6 Halbtöne nach oben oder unten.

6.2.5 Ergebnis und Diskussion

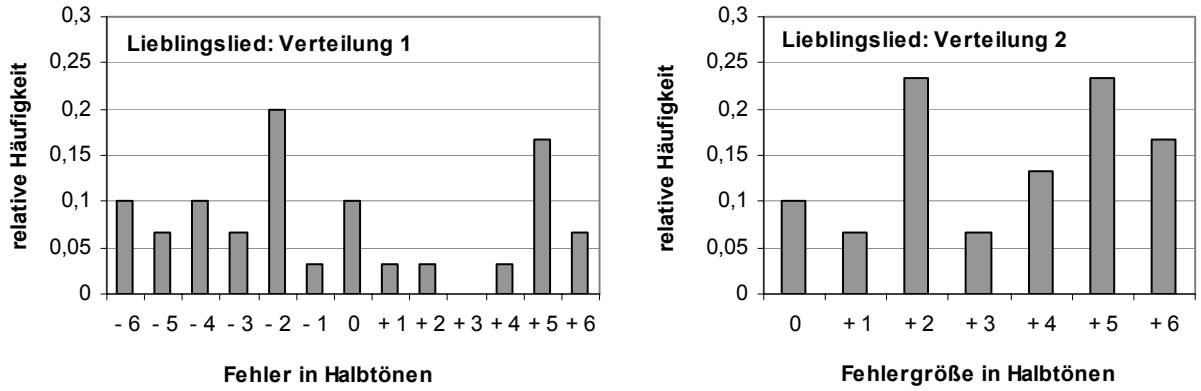

Abb. 6.1: Fehlerverteilungen bei der Produktion eines persönlichen Lieblingsliedes. Eine Abweichung von 0 bedeutet die Produktion der Originaltonart. Negative Abweichungen in Verteilung 1 bedeuten, dass zu tief gesungen wurde, positive Abweichungen, dass zu hoch gesungen wurde. In Verteilung 2 ist die absolute Fehlergröße ohne Beachtung der Richtung abgetragen. Dargestellt sind die relativen Fehlerhäufigkeiten von 30 Vpn.

Die Analyse der Fehlerverteilung ergab kein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 15,12, p = .238), so dass die Nullhypothese der Gleichverteilung nicht abgelehnt werden kann. Der Kolmogorov-Smirnov-Test auf Normalverteilung ergab ebenfalls kein signifikantes Ergebnis (Z = 0,987, p = .284). Der t-Test gegen den Wert 3 wurde für Verteilung 2 nicht signifikant (t (29) = 1,200, p = .240). Insgesamt haben nur 16,6 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung getroffen.

Das Ergebnis der Signifikanztests weist darauf hin, dass die hier produzierten Fehler nicht von einer zufälligen Verteilung abweichen, und dass die mittlere Fehlergröße viel zu groß ist, um den Zufallswert 3 signifikant zu unterschreiten. Das bedeutet, dass bei der hier getesteten Stichprobe keine Tonarterinnerung an ihre Lieblingslieder vorliegt. Das Ergebnis von Levitin (1994) konnte somit nicht repliziert werden, obwohl es sich um eine vergleichbare Stichprobe und vergleichbares musikalisches Material handelte. Die Unterschiede zu Levitins Experiment, die für dieses unerwartete Ergebnis verantwortlich sein könnten, liegen vermutlich einerseits in der Versuchsdurchführung, andererseits in der Vertrautheit des Materials bei ca. einem Drittel der Vpn. Um zu überprüfen, ob bei den 19 Vpn, die ihre Lieblingslieder seit mindestens 5 Jahren kennen, eine genauere Tonarterinnerung nachweisbar ist, wurde für diese Vpn eine weitere Verteilung berechnet. Auch für diese reduzierte Verteilung ergab weder der (aufgrund der kleineren Stichprobe durchgeführte) Kolmogorov-Smirnov-Test (Z = 1,095, p = .181) noch der t-Test (t (18) = 0,364, p = .720) ein signifikantes Ergebnis. Ein t-Test für unabhängige Stichproben ergab keine signifikant bessere Erinnerung der 19 Vpn im Vergleich zu den 11 Vpn, die ihre Lieblingslieder erst seit kürzerer Zeit kannten (t (28) = 1,003, p = .324). Tendenziell ist der mittlere Fehler bei den Vpn mit höherer Lied-Vertrautheit

mit m = 3,16 Halbtönen aber etwas kleiner als der der Vpn mit geringerer Lied-Vertrautheit (m = 3,91 Halbtöne). Dies ist ein Hinweis darauf, dass sowohl die kürzere Bekanntheitsdauer mancher hier getesteter Lieblingslieder als auch der kompliziertere Versuchsablauf die im Vergleich mit Levitins (1994) Ergebnissen wesentlich schlechtere Tonarterinnerung der hier getesteten Stichprobe erklären können. Eine weitere Erklärung könnte darin bestehen, dass die Vpn ihre Lieblingslieder häufig spontan singen und dabei nicht auf die Tonart achten, so dass die Tonartrepräsentation beeinträchtigt wird. Diese Gefahr ist bei Lieblingsliedern mit „Ohrwurm-Potenzial“, aber auch bei Kinderliedern gegeben und kann nur durch die Erhebung entsprechender Befragungsdaten überprüft werden, die hier nicht vorliegen.

6.3 Kurzzeitig gelernte Melodien: Schüler

6.3.1 Stichprobe

Untersucht wurden 77 Schüler des Droste-Hülshoff-Gymnasiums in Berlin (53 weiblich, 24 männlich, Altersdurchschnitt 15,5 Jahre). 32 Schüler gehörten zu einer musikbetonten 9. Klasse, die im Rahmen ihres Musikunterrichts wöchentlich über drei Monate die hier getesteten Melodien lernte. Die übrigen 45 Schüler gehörten dem Schulchor an, der im Rahmen der normalen Chorproben alle 2-3 Wochen über drei Monate die hier getesteten Melodien lernte. Die Lernsitzungen dauerten jeweils 10 Minuten. Die beiden Gruppen unterscheiden sich nicht signifikant hinsichtlich ihrer musikalischen Ausbildung (mittlere Dauer des Instrumentalunterrichts 9. Klasse = 5,7 (SD = 3,4) Jahre, Chor 6,6 (SD = 3,7) Jahre; t (75) = - 1,068, p = .289). Da der Unterschied in der Tonarterinnerung zwischen den beiden Gruppen nur sehr klein ist (vgl. weitere Darstellung dieses Experiments in Kap. 7), wird hier die über beide Gruppen zusammengefasste Fehlerverteilung dargestellt.

6.3.2 Produzierte Melodien

Produziert werden sollten zwei Melodien, die sich musikalisch folgendermaßen charakterisieren lassen: Melodie 1, „By the waters of Babylon“, ist ein bekannter Kanon, der leicht erlernbar und singbar ist. Er besteht überwiegend aus Sekundschritten und enthält nur an den Zeilenwechseln größere Intervalle. Der Ambitus des Kanons beträgt eine Dezime. Produziert werden sollte die erste Zeile, deren Ambitus eine Oktave beträgt. Bei Melodie 2, “Andray soulet”, handelt es sich um einen Kanon aus dem 14. Jahrhundert von Perusio. Da es eine modale Melodie ist, enthält dieser Kanon einige schwierige Tonfolgen, beispielsweise die Folge von vier Ganztönen in Takt 2 oder den großen Septimen-Sprung zwischen Takt 7 und 8. Produziert werden sollte auch hier nur die erste Zeile, deren Ambitus eine Oktave beträgt. Beide Melodien wurden hinsichtlich der Tonarten so gelegt, dass der höchste vorkommende Ton ein c’’ ist, da höhere Töne von ungeübten Sängern nicht ohne weiteres singbar sind.

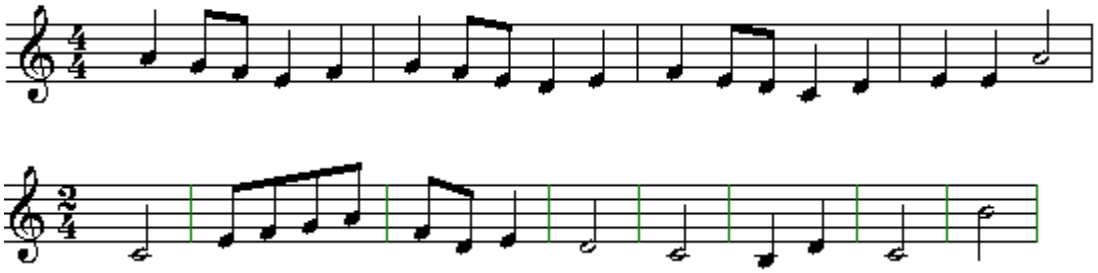

Abb. 6.2: Melodie 1: By the waters of Babylon (oben) und Melodie 2: Andray soulet (unten), jeweils die zu produzierende erste Melodiezeile.

Die beiden Kanons unterscheiden sich vor allem hinsichtlich der Art, in der sie gelernt wurden, von den bei Levitin (1994) getesteten Melodien. Während Levitins Popsongs quasi nebenbei durch das häufige Hören der Lieblingslieder gelernt wurden, wurden die Kanons intentional gelernt, also aktiv durch häufiges Singen. Auch die Lerndauer ist unterschiedlich, denn während Levitin (1994) betont, dass seine Vpn die produzierten Popsongs schon jahrelang kannten, wurden die Kanons nur über 3 Monate regelmäßig gesungen. Musikalisch ist Melodie 1 mit den Popsongs vergleichbar, da sie eine ähnlich einfache Struktur aufweist. Melodie 2 dagegen ist nicht mit einem Popsong vergleichbar, da sie wesentlich komplizierter ist. Dies bedeutete auch Unterschiede in der Lernzeit für die beiden Melodien: Melodie 1 war einem kleineren Teil der Schüler schon bekannt und wurde von den übrigen Schülern schnell gelernt, so dass sie ab der 2. Lernsitzung pro Sitzung nur noch einige Male durchgesungen wurde, und zwar sowohl einstimmig als auch dreistimmig im Kanon. Melodie 2 dagegen war gänzlich unbekannt und außerdem schwerer zu lernen. Dadurch dauerte es 4 Sitzungen, bis die Melodie überhaupt einstimmig gesungen werden konnte, und erst in der 8. bis 10. Sitzung konnte sie sicher im zweistimmigen Kanon gesungen werden. Pro Sitzung wurde auf diese Melodie also mehr Lernzeit aufgewandt. Die erwarteten Erinnerungsunterschiede zwischen den beiden Melodien sind also nicht eindeutig auf Lernzeit *oder* musikalische Struktur zurückführbar, da die beiden Faktoren konfundiert sind.

6.3.3 Besonderheiten bei der Datenauswertung

Beim Vergleich von gesungenen und originalen Anfangstönen wurden Oktavfehler als Fehler bewertet, da die Vpn die Melodien genau in der Lage zu singen geübt hatten, die sie auch im Experiment wiedergeben sollten. Da es bei Chorsängern und offensichtlich auch bei singenden Schülern die Tendenz gibt, eher zu tief als zu hoch zu singen (vgl. Hahn, 2002), resultiert dieses Vorgehen in einer Verteilung, die nicht symmetrisch um den Nullpunkt angeordnet ist, sondern linksschief, d.h. die Kategorien mit negativem Vorzeichen sind erstens stärker besetzt, und zweitens gibt es viel größere negative (bis „-11“) als positive (bis „+2“) Abweichungen. Um die Verteilungen auf Signifikanz zu prüfen, wurden sie umgewandelt in eine Verteilung ohne Oktavlagenfehler.

Da sich 7 Vpn im Experiment nicht mehr an Melodie 2 erinnern konnten, beruhen die Verteilungen für Melodie 2 nur auf 70 Produktionen, während für Melodie 1 Produktionen aller 77 Vpn vorlagen.

6.3.4 Ergebnis und Diskussion

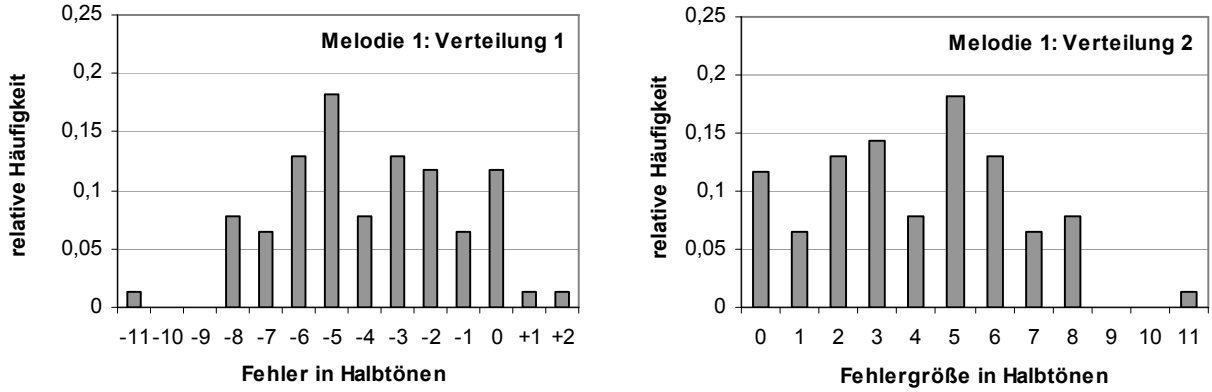

Abb. 6.3: Fehlerverteilungen bei der Produktion von Melodie 1. Beide Verteilungen berücksichtigen Oktavlagenfehler. Dargestellt sind die relativen Fehlerhäufigkeiten von 77 Vpn.

Die Analyse der Fehlerverteilung für Melodie 1 ergab ein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 38,46, p < .001), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant (Z = 1,026, p = .244), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = -4,10) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 nicht signifikant (t (76) = 1,839, p = .070). Insgesamt trafen 18,2 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

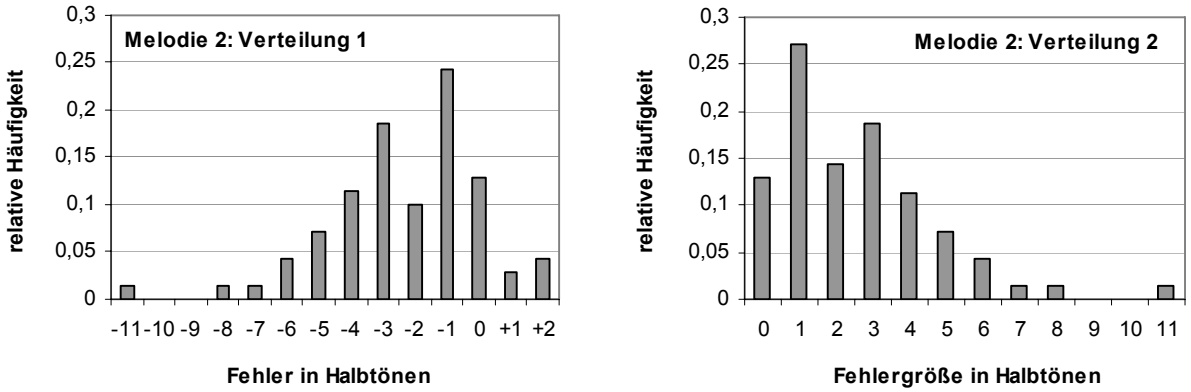

Abb. 6.4: Fehlerverteilungen bei der Produktion von Melodie 2. Beide Verteilungen berücksichtigen Oktavlagenfehler. Dargestellt sind die relativen Fehlerhäufigkeiten von 70 Vpn.

Die Analyse der Fehlerverteilung für Melodie 2 ergab ein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 60,94, p < .001), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1

nicht signifikant (Z = 1,251, p = .087), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = -2,42) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 signifikant (t (69) = -3,304, p = .002). Insgesamt trafen 40 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

Das Ergebnis der Signifikanztests bezüglich Melodie 1 zeigt, dass die beobachtete Fehlerverteilung zwar signifikant von einer zufälligen Verteilung abweicht, dass sich die Fehler aber nicht systematisch um die Kategorie „Fehler = 0 Halbtöne“ verteilen, sondern um die Kategorie „-4“. Die Abweichungen sind im Mittel viel zu groß, um die durch Raten zu erzielende Abweichung zu unterschreiten, wie man an Verteilung 2 deutlich sehen kann. Auffällig ist die hohe Anzahl von nahezu 20 % der Vpn, die 5 Halbtöne zu tief und damit in der Dominant-Tonart sangen. Dies verweist darauf, dass der Ambitus der Melodie mit einer Oktave eventuell zu groß war, um die ganze Melodie zu oktavieren. Vpn, denen die Melodie im Experiment „zu hoch“ erschien, haben sie nicht oktaviert, sondern die am nächsten verwandte Tonart genutzt, die nur 5 statt 12 Halbtöne tiefer lag. Das Ergebnis der Produktion von Melodie 1 zeigt, dass bei einer Melodie, die so leicht ist, dass sie nicht sehr intensiv geübt werden muss, durch das intentionale Erlernen über 3 Monate keine Repräsentation der Originaltonart aufgebaut wurde, die mit der von Levitins (1994) Popsongs vergleichbar wäre.

Der Vergleich mit Melodie 2 zeigt aber, dass auch innerhalb von 3 Monaten eine Repräsentation der Originaltonart aufgebaut werden kann: Auch bei Melodie 2 liegt der Mittelpunkt der Verteilung zwar bei „-2“ statt bei „0“, aber die Fehler streuen weniger breit als die bei Melodie 1 und sind mit einem Mittelwert von 2,3 überzufällig klein. Zusammenfassend kann man also feststellen, dass der Aufbau einer Tonart-Repräsentation auch innerhalb von kurzer Zeit möglich ist, aber nicht mit jeder beliebigen Melodie. In diesem Experiment scheint die Schwierigkeit, aber auch die ungewohnte Struktur von Melodie 2 beim Lernen eine hohe Konzentration bewirkt zu haben, die den Aufbau einer Tonart-Repräsentation erleichtert hat.

6.4 Kurzzeitig gelernte Melodien: Chorsänger

6.4.1 Stichprobe

Untersucht wurden 25 Chorsänger (15 weiblich, 10 männlich, Durchschnittsalter 32,2 Jahre) aus den Chören Ensemberlino Vocale und Ensemble Vocal de Bardou (Leitung: Ralf Sochaczewsky). Beide Ensembles zeichnen sich durch qualitativ hochwertige musikalische Arbeit aus und nehmen so eine Zwischenstellung zwischen professionellen und Amateur-Chören ein. Hinsichtlich ihrer musikalischen Ausbildung unterscheiden sich die beiden Ensembles nicht signifikant (mittlere Instrumentalerfahrung Ensemberlino = 11,8 (SD = 5,5) Jahre, Bardou 10,7 (SD = 5,8) Jahre; t (23) = 0,466, p = .646). Die Durchführung des Experiments wurde zeitlich so gelegt, dass die Konzerte für die Vpn 1-3 Wochen zurücklagen. Das Ensemberlino Vocale hat sein Konzertprogramm über 3 Monate hinweg wöchentlich geprobt, während das Ensemble Vocal de Bardou sein Programm im Rahmen einer zweiwöchigen Proben- und Konzertreise in täglichen Proben einstudiert hatte. Da sich die Produk-

tions-Ergebnisse der beiden Chöre im Mittel nicht signifikant unterschieden (vgl. weitere Darstellung dieses Experiments in Kap. 9), werden hier die Ergebnisse beider Gruppen zusammengefasst dargestellt.

6.4.2 Produzierte Melodien

Die Vpn wurden aufgefordert, eines ihrer Lieblingslieder aus dem jeweiligen Konzertprogramm zu produzieren. Das Konzertprogramm des Ensemberlino Vocale beinhaltete Werke von Helmut Barbe (Volkslieder), Francis Poulenc (8 Chansons Francaises) und Hugo Distler (Mörike-Chorliederbuch, op. 19). Im Vergleich mit Levitins (1994) Popsongs ist diese Literatur musikalisch komplexer und dadurch schwieriger zu singen. Das gleiche gilt für die Johannes-Passion von J.S. Bach, die vom Ensemble Vocal de Bardou gesungen wurde. Produziert wurden sowohl Teile aus den Chorsätzen als auch aus den leichter singbaren Chorälen. Da die Stichprobe jedoch nicht groß genug war, konnten Unterschiede in der Erinnerung an Choräle vs. andere Chorsätze nicht untersucht werden. Hinsichtlich der Vertrautheit gilt für die hier produzierten Melodien Ähnliches wie für die im vorigen Abschnitt beschriebenen: Sie waren den Vpn zwar nicht jahrelang bekannt wie Popsongs, wurden dafür aber in der dem Experiment vorausgehenden Zeit sehr intensiv geprobt, so dass sie manche Vpn laut Selbstaussage während der Probenphase als Ohrwürmer regelrecht verfolgt haben.

6.4.3 Besonderheiten bei der Versuchsdurchführung

Da in diesem Experiment, ähnlich wie in dem in Abschnitt 6.2 beschriebenen Experiment mit Lieblingsliedern, die Bedingungen, unter denen sich die Vpn an ihre Lieblings-Chorlieder erinnern sollten, variiert und dabei auch Reaktionszeiten erfasst wurden (vgl. Darstellung dieses Experiments in Kap. 9), fand die Versuchssteuerung am PC statt und unterschied sich damit vom Versuchsablauf der übrigen Experimente. Die Vpn wurden auf dem Bildschirm angewiesen, sich eine Phrase aus ihrem Lieblingslied vorzustellen und zunächst eine Taste der Tastatur zu drücken, sobald sie eine genaue Vorstellung von dieser Phrase hatten. Erst nach diesem Tastendruck wurden die Vpn durch eine Instruktion auf dem Bildschirm angewiesen, die Phrase zu singen. Auch hier ist also durch die doppelte Reaktionsanforderung mit einer leichten Verunsicherung der Vpn zu rechnen.

6.4.4 Besonderheiten bei der Datenauswertung

Wie im vorigen Experiment wurden auch hier Oktavlagenfehler als Fehler gewertet, da die Chorsänger in der Lage sein sollten, die Melodien im Experiment in der Oktavlage zu singen, die sie lange geübt hatten. Das Ergebnis sind wieder Verteilungen, die nach links mehr und stärker besetzte Kategorien aufweisen als nach rechts, wobei die negativen Abweichungen, die größer als „-6“ waren, für die Signifikanztests nach oben transponiert wurden. Da eine Vp aus dem Alt sich nicht mehr an ihre eigene Stimme erinnern konnte und die Sopran-Melodie sang, wurden ihre Daten in die weitere Analyse nicht mit einbezogen. Folglich sind in den Fehlerverteilungen die Daten von 24 Vpn abgetragen.

6.4.5 Ergebnis und Diskussion

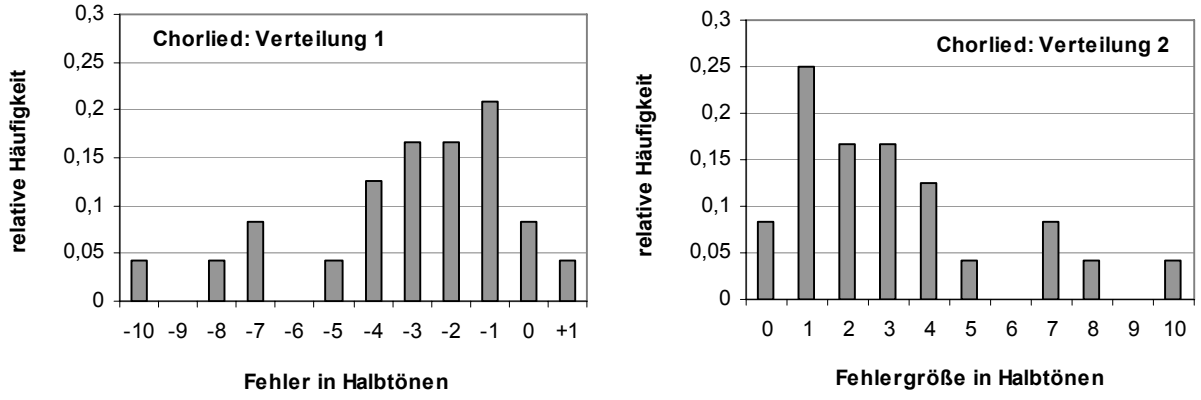

Abb. 6.5: Fehlerverteilungen bei der Produktion des persönlichen Lieblings-Chorliedes. Beide Verteilungen berücksichtigen Oktavlagenfehler. Dargestellt sind die relativen Fehlerhäufigkeiten von 24 Vpn.

Die Analyse der Verteilung 1 erfolgte aufgrund des kleinen Stichprobenumfangs mit dem Kolmogorov-Smirnov-Test. Dieser ergab ein signifikantes Ergebnis (Z = 1,511, p = .021), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant (Z = 0,991, p = .280), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = -3,04) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 nicht signifikant (t (23) = -1,701, p = .102). Insgesamt wurde die Original-Tonart von 33,3 % der Vpn mit höchstens einem Halbton Abweichung getroffen.

Das Ergebnis der Signifikanztests verweist darauf, dass beim Singen von Lieblings-Chorliedern keine völlig zufälligen Tonarten produziert werden, denn die Fehler streuen systematisch um die Kategorien „Fehler = -1“ bis „Fehler = -3“. Jedoch ist der Betrag der Fehler im Mittel zu groß, um den zufällig zu erwartenden Fehler signifikant zu unterschreiten. Das bedeutet, dass für Chorlieder, die zwar intensiv, aber erst seit kürzerer Zeit geprobt wurden, keine Tonarterinnerung gezeigt werden konnte, die mit der von Levitins (1994) Popsongs vergleichbar wäre. Der Grund dafür liegt vermutlich zum einen in der geringeren Vertrautheit der hier getesteten Melodien, zum anderen aber auch in der für die Vpn etwas schwieriger zu handhabenden Versuchsdurchführung. Vergleicht man die hier nachgewiesene Erinnerung an Chormusik mit der im vorigen Abschnitt beschriebenen Erinnerung an im Schulunterricht geübte Melodien, so ist die Erinnerung an Chormusik nicht ganz so gut wie die an Melodie 2, jedoch besser als die an Melodie 1. Vermutlich wurden die Erinnerungsunterschiede zumindest teilweise durch die Prägnanz bzw. Einzigartigkeit der gelernten Melodien hervorgerufen: Während Melodie 1 sehr eingängig, aber wenig prägnant ist, ist Melodie 2 einzigartig in der Weise, dass solche Musik von Schülern normalerweise nicht gehört wird. Zudem ist sie einstimmig und damit vermutlich wiederum prägnanter als die Chorlieder dieses Experiments. Nimmt man dieses und das vorige Experiment zusammen, so scheint die Tonarterinnerung nicht nur von der Vertrautheit mit dem musikalischen Material abhängig zu sein, sondern ebenso von der Struktur der gelernten Melodien.

6.5 Erinnerung an über längere Zeit gelernte Melodien

6.5.1 Stichprobe

Untersucht wurden 17 Mitglieder der Berliner Domkantorei (Leitung: Herbert Hildebrandt) und 20 Mitglieder der Kantorei Potsdam (Leitung: Ud Joffe). Diese beiden Chöre sind qualitativ vergleichbar und wurden für das Experiment ausgewählt, weil sie jedes Jahr das Weihnachtsoratorium von J.S. Bach zur Aufführung bringen. Die Versuchsteilnehmer (29 weiblich, 8 männlich, Altersdurchschnitt 37,2 Jahre, mittlere Instrumentalerfahrung 8,3 (SD = 7,2) Jahre) meldeten sich freiwillig und wurden für die Teilnahme nicht bezahlt. Da die Unterschiede in der Tonart-Produktion zwischen den Sängern beider Chöre nicht signifikant waren (mittlere Abweichung bei „Jauchzet“: Berlin 1,8 (SD = 0,5) Halbtöne, Potsdam 2,3 (SD = 0,5) Halbtöne; t (35) = - 0,643, p = .524; mittlere Abweichung „Choral“: Berlin 3,3 (SD = 0,6) Halbtöne, Potsdam 4,1 (SD = 0,7) Halbtöne; t (32) = - 0,950, p = .349), ist hier die zusammengefasste Fehlerverteilung dargestellt.

6.5.2 Produzierte Melodien

Produziert werden sollte der Beginn des ersten Chorsatzes aus dem Weihnachtsoratorium, „Jauchzet, frohlocket“. Mit seinem einstimmigen Einsatz, der eine Wiederholung des von der Pauke gespielten Anfangsmotivs des Satzes ist, ist dieser Beginn sehr prägnant und dadurch auch sehr einprägsam. Hinsichtlich der Vertrautheit ist dieser Choreinsatz mit Levitins (1994) Popsongs vergleichbar, denn viele der Vpn kennen das Weihnachtsoratorium vom Hören her von Kindheit an, auch wenn sie es erst wenige Male gesungen haben. Hinsichtlich der musikalischen Struktur ist der ganze Chorsatz mit seiner polyphonen Struktur nicht mit den Popsongs vergleichbar, sondern ist wesentlich komplexer. Die Vpn wurden aber ausdrücklich instruiert, nur den Beginn des Satzes zu singen, um die Anforderung von der Schwierigkeit her in Grenzen zu halten. Weiterhin sollte noch einer von zwei Chorälen produziert werden, und zwar entweder „Wie soll ich dich empfangen“ oder „Brich an, o schönes Morgenlicht“. Die beiden Choräle sind von der Schwierigkeit und vom Stimmumfang miteinander vergleichbar. Verglichen mit „Jauchzet, frohlocket“ sind sie zwar musikalisch weniger komplex, aber auch weniger einprägsam, da in den Chorälen nur der Sopran die Melodie singt, während die übrigen Stimmen begleiten. Es gibt einen wichtigen Unterschied zwischen Levitins (1994) Popsongs und den Chorälen: Da Bach diese Choräle in anderen Werken mit neuen Texten und in anderen Tonarten wieder verwendet hat, ist es möglich, dass einige Vpn die hier getesteten Choräle in mehreren Versionen kennen. Dadurch, wie auch durch die Tatsache, dass die vergleichsweise leichten Choräle weniger oft geprobt werden als der Eingangschor, ist bei jenen eine weniger genaue Tonarterinnerung zu erwarten als beim Eingangschor (und Popsongs), von dem es nur eine Version gibt.

6.5.3 Besonderheiten bei der Datenauswertung

Ähnlich wie bei dem vorigen Experiment wurden auch hier Oktavlagenfehler als Fehler gewertet, da die Chorsänger in der Lage sein sollten, die Melodien im Experiment in der Oktavlage zu singen, die sie lange geübt hatten. Es resultieren auch hier Verteilungen, die nach links mehr und stärker besetzte Kategorien aufweisen als nach rechts, wobei die negativen Abweichungen, die größer als „-6“ waren, für die Signifikanztests wieder nach oben transponiert wurden. Bei den Chorälen trat bei 3 Vpn das Problem auf, dass sie sich an ihre eigene Stimme nicht mehr erinnern konnten und daher die Sopranstimme, also die Melodie, sangen. Diese Fälle wurden von den weiteren Analysen ausgeschlossen, so dass in die Fehlerverteilung der Choräle nur 34 Fälle eingingen.

6.5.4 Ergebnis und Diskussion

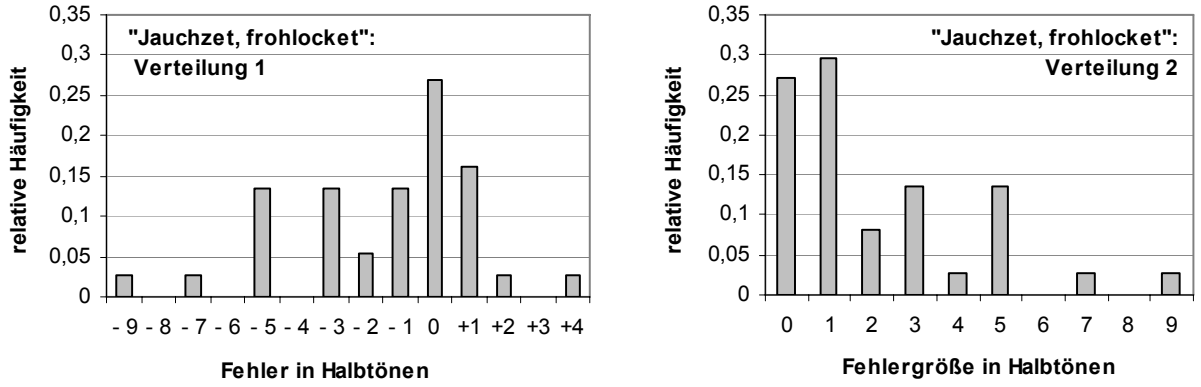

Abb. 6.6: Fehlerverteilungen bei der Produktion von „Jauchzet, frohlocket“. Beide Verteilungen berücksichtigen Oktavlagenfehler. Dargestellt sind die relativen Fehlerhäufigkeiten von 37 Vpn.

Die Analyse der Fehlerverteilung für „Jauchzet, frohlocket“ ergab ein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 40,61, p < .001), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant (Z = 1,146, p = .145), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = -1,43) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 signifikant (t (36) = -3,875, p < .001). Insgesamt trafen 56,7 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

Bei den Chorälen wurde zunächst geprüft, ob sich die Produktionsgenauigkeit zwischen den beiden Chorälen unterschied. Der t-Test für unabhängige Stichproben ergab für diesen Vergleich kein signifikantes Ergebnis (mittlere Abweichung „Brich an“ = 4,11; mittlere Abweichung „Wie soll“ = 3,67; t (32) = 0,532, p = .589). Dadurch wird eine gemeinsame Analyse beider Choräle gerechtfertigt. Die Analyse der Fehlerverteilung für die Choräle ergab kein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 11,95, p = .455), so dass die Nullhypothese der Gleichverteilung nicht abgelehnt werden kann. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde ebenfalls nicht signifikant (Z = 0,888, p = .410). Der t-Test gegen

den Wert 3 wurde für Verteilung 2 nicht signifikant (t (33) = 0, p = 1,0). Insgesamt trafen 23,5 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

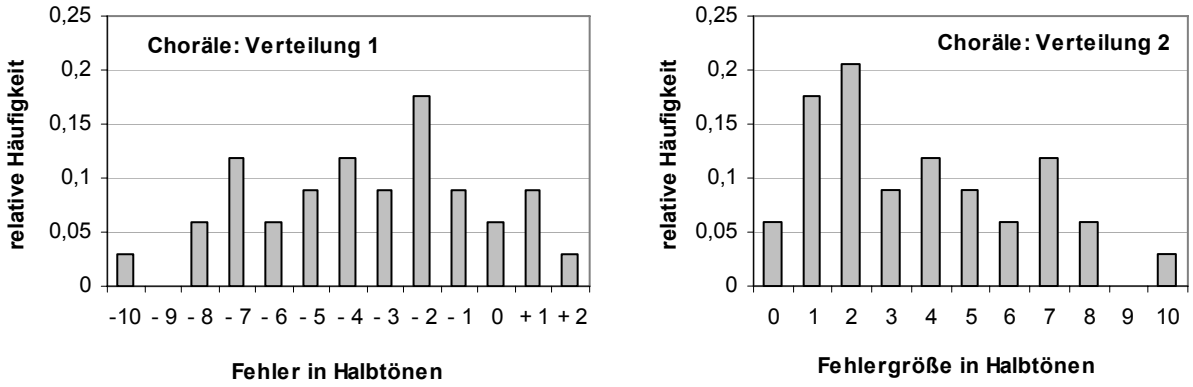

Abb. 6.7: Fehlerverteilungen bei der Produktion der Choräle. Beide Verteilungen berücksichtigen Oktavlagenfehler. Dargestellt sind die relativen Fehlerhäufigkeiten von 34 Vpn.

Die Ergebnisse der Signifikanztests bezüglich „Jauchzet, frohlocket“ verweisen darauf, dass für diesen Chorsatz eine signifikante Tonarterinnerung vorliegt. Zwar singen die Chorsänger im Mittel etwas über einen Halbton zu tief, aber die durchschnittliche Fehlergröße liegt deutlich unter der Ratewahrscheinlichkeit, und die Streuung der Fehler ist mit der bei Levitins (1994) Popsongs vergleichbar. Demnach kann durch das intentionale Lernen eines Chorsatzes über mehrere Jahre hinweg eine mit der von Lieblingsliedern vergleichbare Repräsentation der Originaltonart aufgebaut werden. Das Ergebnis bezüglich der Choräle zeigt aber, dass dies nicht auf alle Teile des Weihnachtsoratoriums gleichermaßen zutrifft, denn bei den Chorälen wurde kein überzufälliges Ergebnis erzielt. Die gesungenen Tonarten streuen sehr breit, und fast alle Chorsänger sangen deutlich zu tief. Dieses Ergebnis ist nach Angaben der Vpn darauf zurückzuführen, dass sie sich nicht mehr sehr gut an ihre Stimme erinnern konnten, falls es eine Mittelstimme war, sondern nur noch an die Melodie. Dies wiederum könnte daran liegen, dass die Choräle sehr selten geprobt wurden und manche Sänger sie aus anderen Werken Bachs kannten. Insgesamt lässt sich aus den Ergebnissen dieses Experiments schlussfolgern, dass eine absolute Tonartrepräsentation bei über lange Zeit gelernter Chormusik nur dann aufgebaut wird, wenn die zu erinnernde Melodie erstens prägnant und zweitens charakteristisch nur für ein einziges Werk ist.

6.6 Erinnerung an einen Werbejingle

6.6.1 Stichprobe

Die Stichprobe bestand aus den 77 Schülern des in Abschnitt 6.3 beschriebenen Experiments plus 17 Schülern aus einem Musik-Basiskurs der Anna-Freud-Oberschule in Berlin (13 weiblich, 4 männlich, Durchschnittsalter 18,1 Jahre, durchschnittliche Instrumentalerfahrung 4,1 (SD = 4,0) Jahre), die zusätzlich getestet wurden.

6.6.2 Produzierte Melodie

Produziert werden sollte der Werbejingle der Telekom, der aus einem fünftönigen Motiv besteht. Charakteristisch für dieses Motiv ist die häufige Wiederholung des Tons *c* sowie der Terz-Sprung aufwärts.

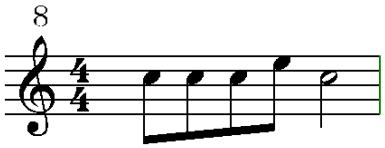

Abb. 6.8: Telekom-Werbejingle.

Der Telekom-Werbejingle ist sehr häufig in Radio und TV zu hören, so dass er den Vpn, die diese Medien regelmäßig nutzen, sehr vertraut ist. Da es ihn außerdem schon seit längerer Zeit gibt, ist er von der Vertrautheit her mit Levitins (1994) Popsongs vergleichbar. Allerdings hat er nicht die gleiche positive Konnotation wie persönliche Lieblingslieder. Da er aber einerseits prägnant und andererseits in der Werbung sehr präsent ist, ist eine ähnlich genaue Tonart-Repräsentation zu erwarten wie für die Popsongs.

6.6.3 Besonderheiten bei der Datenauswertung

Wie bei den Lieblingsliedern wurden in diesem Experiment Oktavlagenfehler nicht als Fehler bewertet, da der Jingle so hoch liegt, dass er in der Originaltonart für kaum jemanden singbar ist.

Insgesamt 23 Vpn gaben an, so selten Radio zu hören oder fernzusehen, dass ihnen der Telekom-Jingle nicht bekannt sei. Daher sind in die Fehlerverteilungen nur die Daten von 71 Vpn eingegangen.

6.6.4 Ergebnis und Diskussion

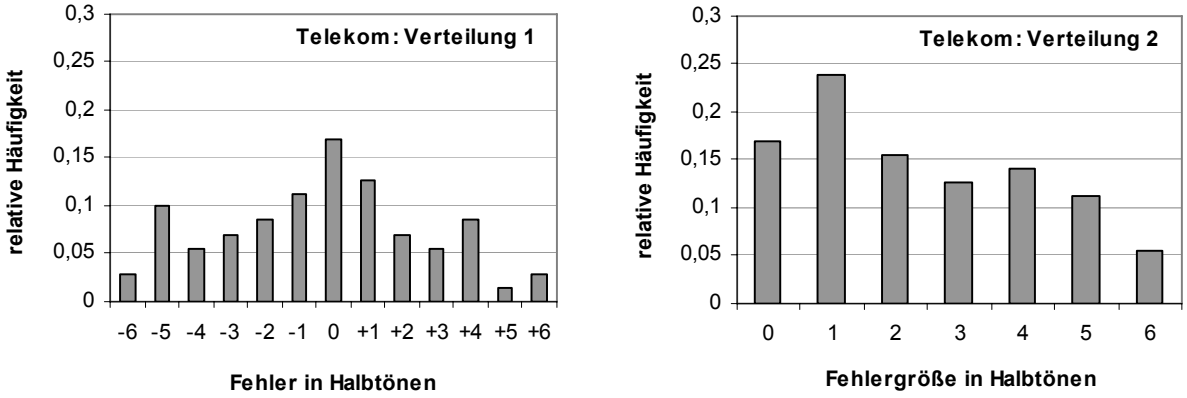

Abb. 6.9: Fehlerverteilungen bei der Produktion des Telekom-Werbejingles. Dargestellt sind die relativen Fehlerhäufigkeiten bei den 71 Vpn, denen der Werbejingle sehr vertraut war.

Der chi²-Test wurde für Verteilung 1 nur tendenziell signifikant (*chi²* = 20,73, *p* = .054), so dass die Nullhypothese der Gleichverteilung nicht abgelehnt werden kann. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant (*Z* = 0,823, *p* = .508), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = - 0,37) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 signifikant (*t* (70) = - 2,762, *p* = .007). Insgesamt trafen 40,9 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

Die Ergebnisse der Signifikanztests bedeuten, dass die beobachtete Fehlerverteilung zumindest tendenziell von einer zufälligen Verteilung abweicht und sich die Fehler systematisch um die Kategorie „Fehler = 0 Halbtöne“ verteilen. Da der durchschnittliche Fehler signifikant unter der Ratewahrscheinlichkeit liegt, ist insgesamt von einer Repräsentation der Tonart des Telekom-Werbejingles auszugehen. Jedoch ist diese etwas weniger genau als die von Levitins (1994) Popsongs, denn die Fehler streuen insgesamt breiter. Außerdem zeigte sich auch hier die Tendenz, dass neben der Originaltonart häufig die Dominant-Tonart gesungen wurde. Insgesamt lässt sich aus diesem Experiment schließen, dass das bloße Hören eines Werbejingles, wenn es oft genug geschieht, zu einer Repräsentation der Tonart führen kann, die fast so genau ist wie die von Levitins (1994) Popsongs.

6.7 Erinnerung an die eigene Handymelodie

6.7.1 Stichprobe

Untersucht wurden 55 musikalisch nicht oder wenig vorgebildete Vpn (37 weiblich, 18 männlich, Altersdurchschnitt 23,5 Jahre, mittlere Instrumentalerfahrung 2,6 (SD = 4,1) Jahre), die überwiegend aus der Studentenschaft der Humboldt-Universität zu Berlin rekrutiert wurden. Sie nahmen am Versuch freiwillig und ohne Bezahlung teil.

6.7.2 Produzierte Melodien

Produziert werden sollte der Klingelton des eigenen Handys durch Singen, Summen oder Pfeifen. Handymelodien sind mit den Popsongs aus Levitins (1994) Experiment insofern vergleichbar, als sie durch sehr häufiges Hören gelernt werden. Wie Popsongs sind Handymelodien außerdem meist musikalisch einfach strukturiert und einprägsam. Zum Teil werden auch persönliche Lieblingslieder als Klingeltöne auf das Handy geladen oder einprogrammiert. Dadurch ergibt sich ein Unterschied der Handymelodien zu Levitins (1994) Popsongs: Während Levitin nur Popsongs nutzte, von denen nicht mehr als eine Version auf Tonträgern existierte (also keine Cover-Versionen in anderen Tonarten), werden bei zu Handymelodien umgewandelten Lieblingsliedern zum Teil die Tonarten verändert. Dies geschieht vor allem dann, wenn der Nutzer die Melodie selbst einprogrammiert und dabei zur Vereinfachung vorzeichenlose Tonarten benutzt. In den Fällen, in denen der Handybesitzer seinen Klingelton in mehreren Tonarten regelmäßig hört, ist keine genaue Tonarterinnerung zu erwarten. Da die Information, ob es außer ihrem Klingelton noch andere Versionen der jeweiligen Me-

lodie gibt, die regelmäßig gehört werden, bei diesen Vpn nicht erhoben wurde, ist hier nur ein Vergleich zwischen den Produktionsleistungen von Nutzern „reiner“ Handymelodien und Nutzern potenziell in mehreren Versionen vorhandener Melodien möglich. Dies geschah über einen t-Test für unabhängige Stichproben. Die Handymelodien wurden durch einen musikalischen Experten nach den beiden Kategorien „reine Handymelodie“ und „gecoverte Melodie“ klassifiziert.

6.7.3 Besonderheiten bei der Datenauswertung

Wie bei den Lieblingsliedern und dem Werbejingle wurden hier bei der Auswertung Oktavlagenfehler nicht als Fehler gewertet, da viele Klingeltöne so hoch liegen, dass man sie kaum in der Original-Lage wiedergeben kann, zumal hier ungeübte Sänger untersucht wurden. In sechs Fällen waren die Produktionen so ungenau, dass der Klingelton nicht erkennbar war. Diese Fälle wurden von weiteren Analysen ausgeschlossen, so dass für die Abweichungs-Verteilung noch 49 Fälle bleiben.

6.7.4 Ergebnis und Diskussion

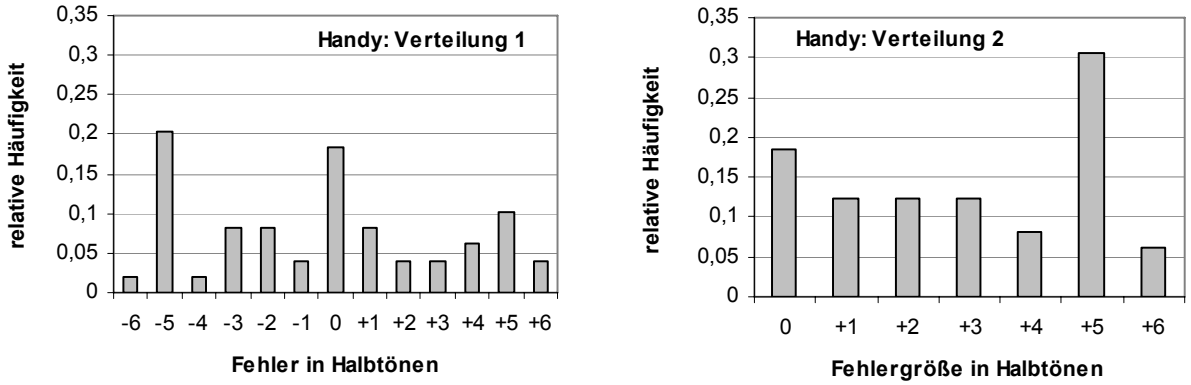

Abb. 6.10: Fehlerverteilungen bei der Produktion der eigenen Handymelodie. Dargestellt sind die relativen Fehlerhäufigkeiten bei den 49 Vpn, deren Produktionen eindeutig als ihre Handymelodien erkennbar waren.

Der chi²-Test wurde für Verteilung 1 signifikant ($chi^2 = 25{,}55$, $p = .012$), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant ($Z = 0{,}869$, $p = .437$), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = -0,39) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 nicht signifikant ($t\ (48) = 0{,}139$, $p = .890$). Insgesamt trafen 30,7 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung. Der Vergleich zwischen den Nutzern von originalen Handymelodien (n = 18) und Nutzern von gecoverten Melodien (n = 31) ergab einen signifikanten Unterschied in der Produktionsgenauigkeit [$t\ (47) = -4{,}739$, $p < .001$], wobei die originalen Handymelodien mit einer mittleren Abweichung von 1,4 Halbtönen besser getroffen wurden als die gecoverten Melodien mit einer mittleren Abweichung von 3,8 Halbtönen.

Das Ergebnis der Signifikanztests bedeutet, dass die beobachtete Fehlerverteilung zwar signifikant von einer zufälligen Verteilung abweicht und sich die Fehler systematisch um die Kategorie „Fehler = 0 Halbtöne“ verteilen, dass aber die Streuung relativ groß ist und dadurch die Abweichungen im Mittel zu groß, um die durch Raten zu erzielende Abweichung signifikant zu unterschreiten. Wie man an Verteilung 2 deutlich sieht, schlägt besonders die hohe Anzahl an Vpn, die 5 Halbtöne zu tief sangen und damit ihren Klingelton in der Dominante, also in der mit der Originaltonart am engsten verwandten Tonart, wiedergaben, zu Buche. Das deutet darauf hin, dass nicht alle Vpn, deren Handymelodie zu hoch zum Singen war, die Melodie oktaviert haben, sondern teilweise „nur“ eine Quarte tiefer sangen.

Das Ergebnis dieser Produktionsaufgabe zeigt, dass bei musikalisch ungeschulten Handynutzern keine mit der von Levitins (1994) Popsongs vergleichbare Tonarterinnerung vorliegt, denn die Fehler streuen wesentlich breiter als bei Levitin. Allerdings werden auch keine zufälligen Tonarten produziert, sondern es gibt eine Tendenz, entweder die Originaltonart oder deren Dominante zu singen. Ein Problem bei der Interpretation dieser Daten besteht sicherlich darin, dass nicht erhoben wurde, ob die Vpn ihren Klingelton häufig auch in einer anderen Tonart hören. Fans einer bestimmten Melodie, die sich diese als Klingelton einstellen, können dann eine unklare Tonartrepräsentation des Klingeltons haben, wenn die Originalmelodie erstens eine andere Tonart hat als der Klingelton und zweitens häufig gehört wird. Andererseits kann man argumentieren, dass der Handy-Klingelton aufgrund seines Signalcharakters herausragende Eigenschaften hat und durch das häufige Hören des Klingeltons, meist täglich mehrmals, dessen Tonart repräsentiert wird. Der t-Test zeigte aber, dass zu Handymelodien umfunktionierte Popsongs, TV-Jingles oder Hits der klassischen Musik insgesamt ungenauer repräsentiert waren als originale Handymelodien. Dadurch wird die Annahme einer herausragenden Bedeutung der Handymelodie und dadurch stabilen Tonartrepräsentation nicht gestützt. Um die Repräsentation der eigenen Handymelodie genauer zu untersuchen, sind weitere Untersuchungen mit der genaueren Erhebung von Hörgewohnheiten der Handynutzer notwendig.

6.8 Erinnerung an ein Instrumental-Konzert

6.8.1 Stichprobe

Untersucht wurde eine Stichprobe von 40 Instrumentalmusikern (25 weiblich, 15 männlich, Durchschnittsalter 24,0 Jahre). Es handelte sich um Studierende der beiden Berliner Musikhochschulen, die ein Holzblasinstrument (Querflöte, Klarinette, Oboe, Fagott), im Hauptfach „Orchestermusik“ studierten. Im Unterschied zu allen bisher berichteten Experimenten bestand diese Stichprobe also aus professionellen Musikern, die über eine durchschnittliche Instrumentalerfahrung von 16,6 (SD = 3,4) Jahren verfügten und ihr erstes Instrument im Mittel mit 7,4 (SD = 2,9) Jahren zu spielen begonnen hatten. Viele der Vpn wurden über ihre Hauptfach-Lehrer rekrutiert, die übrigen Vpn wurden per Schneeball-System, also durch Weiterempfehlung einer anderen Vp gefunden. Die Teilnahme am Experiment erfolgte frei-

willig und ohne Vergütung. Da sich die Produktionen der Musiker verschiedener Instrumente nicht signifikant unterschieden (vgl. Darstellung dieses Experiments in Kap. 9), werden hier die Daten aller Vpn zusammengefasst dargestellt.

6.8.2 Produzierte Melodien

Es gibt von W.A. Mozart jeweils ein Solokonzert für die Oboe, die Klarinette und das Fagott, sowie zwei Konzerte für Querflöte. Diese Solokonzerte sind im Repertoire von Orchestermusikern aus dem Grund herausragend, weil sie bei allen Probespielen für eine Position im Orchester vorgespielt werden müssen. Daher fangen die meisten Hauptfach-Studenten schon früh im Studium an, das Mozart-Konzert für ihr Instrument zu üben. Die hier getesteten Vpn haben „ihr“ Mozart-Konzert im Mittel seit 4,8 Jahren regelmäßig geübt. Hinsichtlich der Vertrautheit ist es insofern mit der von Levitins (1994) Popsongs vergleichbar, als es zwar mit knapp 5 Jahren noch nicht so lange geübt wie manches Lieblingslied gehört wird, dafür aber aufgrund der herausragenden Bedeutung für den Beruf sehr viel regelmäßiger. Musikalisch sind die Anfänge komplexer als Popsongs und mit Ausnahme des Klarinetten-Konzerts auch schwieriger zu singen: Das Oboen-Konzert fängt mit einem Triller an, der schwer zu singen ist, gefolgt von einer schnellen Aufwärtsbewegung über eine kleine None. Das Flöten-Konzert in D-Dur ist das gleiche wie das Oboen-Konzert (einen Ganzton höher), insofern hat es die gleichen Schwierigkeiten, während das Flöten-Konzert in G-Dur ebenso wie das Fagott-Konzert aufgrund des relativ großen Tonumfangs schwer zu singen ist. Um die Aufgabenschwierigkeit in Grenzen zu halten, wurden die Musiker angewiesen, nur den Beginn des Konzerts zu singen und Töne, die außerhalb ihres Stimmumfangs lagen, wegzulassen.

Als weitere Melodie sollten die Vpn eine seit längerer Zeit geübte Orchesterstelle produzieren, also eine für ihr Instrument besonders schwierige Passage aus einem Orchesterwerk. Solche Orchesterstellen werden ähnlich wie das Mozart-Konzert sehr intensiv geübt, weil sie ebenfalls bei Probespielen vorgespielt werden müssen. Die hier getesteten Orchesterstellen wurden im Mittel seit 3,2 Jahren regelmäßig geübt, also etwas kürzer als das Mozart-Konzert. Hinsichtlich der Vertrautheit sind sie aufgrund der Regelmäßigkeit des Übens trotzdem ebenfalls vergleichbar mit den Popsongs, und musikalisch sind sie ebenfalls komplexer. Außerdem waren viele der Orchesterstellen sehr schwer zu singen, weil es sich per definitionem um besonders schwierige (= oft besonders schnelle) Passagen handelt.

6.8.3 Besonderheiten bei der Datenauswertung

Da die Instrumentalmusiker die zu produzierenden Melodien nicht primär durch Singen gelernt haben, sondern durch Üben am Instrument, und die Melodien zum Teil sehr hoch liegen und vor allem für die männlichen Vpn nicht in der Originallage singbar sind, wurden bei der Auswertung Oktavlagenfehler nicht als Fehler bewertet.

6.8.4 Ergebnis und Diskussion

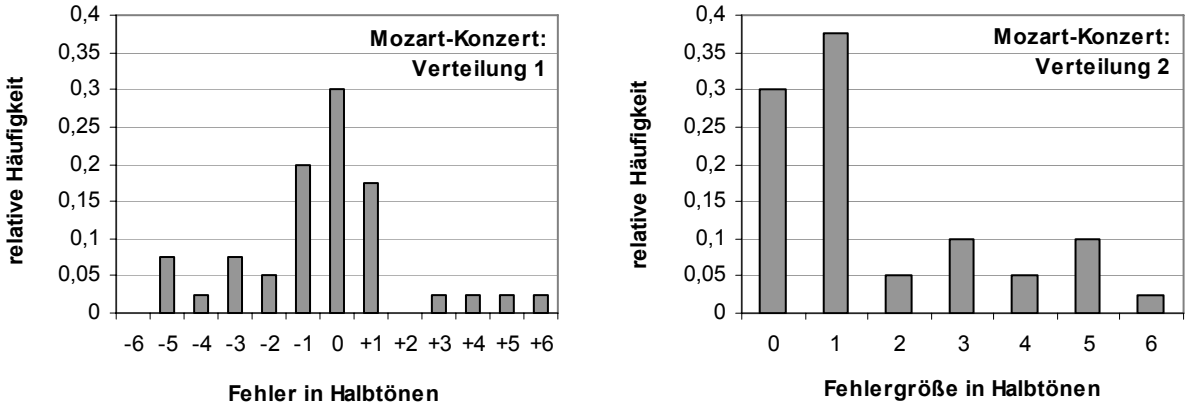

Abb. 6.11: Fehlerverteilungen bei der Produktion des Mozart-Konzerts des eigenen Hauptinstruments. Dargestellt sind die relativen Fehlerhäufigkeiten bei 40 Vpn.

Die Analyse der Fehlerverteilung für das Mozart-Konzert ergab ein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 51,91, p < .001), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant (Z = 1,151, p = .142), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = - 0,38) verteilen. Der t-Test gegen den Wert 3 wurde für Verteilung 2 signifikant (t (39) = - 4,929, p < .001). Insgesamt trafen 67,5 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

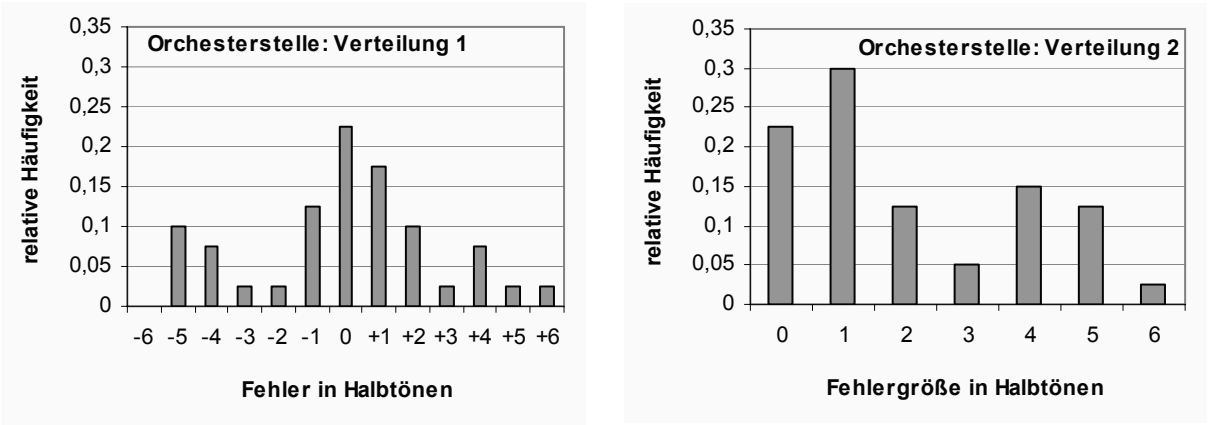

Abb. 6.12: Fehlerverteilungen bei der Produktion einer Orchesterstelle für das eigene Hauptinstrument. Dargestellt sind die relativen Fehlerhäufigkeiten bei 40 Vpn.

Die Analyse der Fehlerverteilung für die Orchesterstelle ergab ein signifikantes Ergebnis im chi²-Test für Verteilung 1 (chi^2 = 28,04, p = .005), so dass die Nullhypothese der Gleichverteilung abzulehnen ist. Der Kolmogorov-Smirnov-Test auf Normalverteilung wurde für Verteilung 1 nicht signifikant (Z = 0,971, p = .302), so dass davon auszugehen ist, dass sich die Fehler systematisch um den Mittelpunkt der Verteilung (m = - 0,03) verteilen. Der t-Test ge-

gen den Wert 3 wurde für Verteilung 2 signifikant (t (39) = - 3,147, p = .003). Insgesamt trafen 52,5 % der Vpn die Original-Tonart mit höchstens einem Halbton Abweichung.

Die Signifikanztests zeigen, dass sowohl für das Mozart-Konzert als auch für die Orchesterstellen die Fehlerverteilungen nicht zufällig sind, sondern es eine Häufung um die Kategorie „Fehler = 0 Halbtöne“ herum gibt. Da die mittleren Fehler überzufällig klein sind, ist davon auszugehen, dass professionelle Instrumentalmusiker eine Tonart-Repräsentation für intensiv geübte Stücke aufgebaut haben, die mit der von Levitins (1994) Popsongs vergleichbar ist. Dies liegt vermutlich einerseits an ihrer höheren musikalischen Expertise und Fähigkeit, Töne auf dem Tonhöhenkontinuum zu kategorisieren (vgl. Darstellung weiterer Ergebnisse in Kap. 9), andererseits aber auch an der mit der von Lieblingsliedern vergleichbaren Vertrautheit der hier produzierten Melodien, die intensiver geübt wurden als alle anderen Melodien der bisher beschriebenen Experimente.

6.9 Zusammenfassende Diskussion der Produktions-Ergebnisse

Um einen Überblick darüber zu erhalten, wer sich unter welchen Umständen wie gut an die Tonarten verschiedener Arten von Melodien erinnern kann, wurden in allen beschriebenen Experimenten insgesamt die Tonarten von 472 produzierten Melodien ausgewertet und zusammenfassend dargestellt. Unter den Vpn befanden sich sowohl musikalische Laien als auch semiprofessionelle und professionelle Musiker. Produziert wurden persönliche Lieblingslieder und Handymelodien, ein Werbejingle, verschiedene Varianten von durch Singen in Chor oder Schule gelernten Melodien sowie Melodien, die sonst auf dem Instrument geübt werden. Sie variieren sowohl in der persönlichen Vertrautheit als auch hinsichtlich der musikalischen Komplexität.

Unter den musikalischen Laien, die durch häufiges Hören gelernte Melodien produzierten, wurden folgende Ergebnisse gefunden: Zunächst konnte das Ergebnis von Levitin (1994) bei der hier untersuchten Stichprobe nicht repliziert werden, denn nur 16,6 % statt wie bei ihm über 50 % trafen die Originaltonart (mit Halbtontoleranz) eines Lieblingsliedes, damit ist die Tonarterinnerung nicht besser als der Zufall. Dieser große Unterschied in den Ergebnissen ist wohl nur teilweise durch die Unterschiede im Versuchsablauf erklärbar. Daneben spielt wahrscheinlich der Faktor Vertrautheit eine Rolle, denn aufgrund der geringeren Standardisierung der hier produzierten Melodien wurden nicht nur seit Jahren bekannte Evergreens produziert, sondern von ca. einem Drittel der Vpn auch seit kürzerer Zeit bekannte Lieblingslieder. Dieses Ergebnis verweist darauf, dass Levitins (1994) Ergebnis nicht auf alle möglichen Arten von Popsongs verallgemeinerbar ist, sondern nur für manche Melodien eine absolute Tonartrepräsentation aufgebaut wird, nämlich für seit sehr langer Zeit bekannte.

Auch für Handymelodien, die als Versuchsmaterial gewählt wurden, weil sie zwar nicht seit so langer Zeit wie persönliche Lieblingslieder gehört werden, dafür aber täglich, konnte keine mit Levitins (1994) Ergebnis vergleichbare Tonarterinnerung nachgewiesen werden, obwohl mit 30,7 % wesentlich mehr Vpn die Originaltonart trafen als bei Lieblingsliedern. Obgleich Handymelodien Signalcharakter haben und somit für ihre Nutzer persönlich sehr wichtig und

auch einzigartig sind, ist deren Vertrautheit vermutlich nicht hoch genug, um mit der von Evergreens vergleichbar zu sein. Dabei muss auch berücksichtigt werden, dass manche Vpn ihre Handymelodie häufiger als einmal pro Jahr wechseln und somit keine extreme Vertrautheit mit einer Melodie aufgebaut werden konnte. Außerdem wird die genaue Repräsentation der Tonarten von Handymelodien durch die Tatsache erschwert, dass über die Hälfte der produzierten Handymelodien gecoverte Versionen sonstiger Melodien sind, die auf dem Handy nicht in allen Fällen in der Originaltonart erklingen.

Die beste Tonart-Erinnerung unter musikalischen Laien beim Produzieren von durch häufiges Hören gelernter Musik wurde bei der Produktion des Telekom-Werbejingles erzielt, hier war die Trefferquote mit 40,9 % fast mit der von Levitin (1994) vergleichbar. Dies lässt sich vermutlich einerseits auf die starke Präsenz dieses Jingles in der Werbung und auf seine hohe Prägnanz zurückführen, andererseits aber auch auf seine Kürze und dadurch sehr geringe Schwierigkeit. Während viele Vpn bei Lieblingsliedern und auch bei der Handymelodie eine große Scheu zu singen an den Tag legten, die auch durch die Instruktion nur zum Teil verringert werden konnte, war der Jingle offensichtlich kurz und einfach genug, um jeder Vpn das Gefühl zu geben, ihn „einfach nachsingen“ zu können. Zusammen belegen diese Ergebnisse, dass auch bei vielen Vpn ohne musikalische Ausbildung eine signifikante Tonarterinnerung nachweisbar ist, deren Genauigkeit von der Art der Melodie und deren Vertrautheit abzuhängen scheint.

Wie bei durch häufiges Hören gelernten Melodien gab es auch bei den intentional gelernten Melodien ein breites Leistungsspektrum hinsichtlich der Tonarterinnerung. Die vergleichsweise schlechtesten Leistungen, gemessen am Prozentsatz der Vpn, die mit Halbtontoleranz die richtige Tonart sangen, traten mit 18,2 % bei der Erinnerung an Melodie 1 aus dem Schüler-Experiment sowie mit 23,5 % bei der Erinnerung an einen Choral aus dem Weihnachtsoratorium auf, beides Melodien, die aufgrund ihrer allgemeinen Bekanntheit und Einfachheit nicht sehr intensiv geübt wurden. Die Erinnerung an die erst seit kurzer Zeit gelernte Chormusik war mit einer Trefferquote von 33,3 % zwar etwas besser, aber auch hier streuen die Fehler wesentlich breiter als bei Levitins (1994) Popsongs. Die besten Ergebnisse bei durch Singen gelernten Melodien traten mit 40 % bei Melodie 2 aus dem Schüler-Experiment sowie mit 56,7 % bei der Produktion des Eingangschores aus dem Weihnachtsoratorium auf. Eine Gemeinsamkeit dieser beiden Melodien ist, dass sie beide einstimmig beginnen und sehr prägnant sind. Der Eingangschor ist prägnant aufgrund seiner herausragenden Position am Anfang des Weihnachtsoratoriums und seines paukenartigen Motivs, während Melodie 2 aufgrund ihrer ungewohnten musikalischen Struktur prägnant ist. Ein vermutlich entscheidender Unterschied zwischen den beiden Melodien ist, dass Melodie 2 nur über 3 Monate gelernt wurde, das Weihnachtsoratorium aber über mehrere Jahre. Dieser Unterschied könnte der Grund dafür sein, dass nur die Trefferquote des Eingangschores aus dem Weihnachtsoratorium mit der von Levitin (1994) vergleichbar ist. Offensichtlich wird bei intentional gelernten Melodien also nur dann eine Tonart-Repräsentation aufgebaut, wenn die Melodien sowohl hoch vertraut als auch prägnant (auch im Sinne von ungewöhnlich) sind.

Die Ergebnisse aus dem Experiment mit Instrumentalmusikern weisen in die gleiche Richtung. Auch hier wurden hoch vertraute und für die Musiker persönlich wichtige Melodien produziert. Die Trefferquote war mit 67,5 % für die Produktion des Mozart-Konzerts die höchste aller hier dargestellter Experimente, während die für die Orchesterstelle mit 52,5 % vergleichbar ist mit der Trefferquote des Weihnachtsoratoriums-Eingangschores.

Zusammenfassend lässt sich festhalten, dass sich aus dem Vergleich der sieben Produktions-Experimente mehrere Einflussfaktoren auf die Tonarterinnerung ergeben. Diese liegen zum Einen auf der Seite der Melodie: Hier scheint die Prägnanz oder auch die Einzigartigkeit einer Melodie wichtig zu sein. Bei sehr prägnanten bzw. einzigartigen Melodien scheint die Tonart eher repräsentiert zu sein als bei gewöhnlicheren Melodien. Allerdings ist fraglich, inwieweit dieser Faktor mit der Lernintensität von Melodien konfundiert ist, denn bei intentionalem Lernen werden schwierige Melodien intensiver geübt als eingängige Melodien, und die Vertrautheit (im Sinne von Lerndauer oder Lernintensität) scheint ein weiterer wichtiger Einflussfaktor auf die Tonart-Repräsentation zu sein. Hinsichtlich der Lernart – durch häufiges Singen oder häufiges Hören - lässt sich vorerst nur feststellen, dass durch beide Varianten der Aufbau absoluter Tonrepräsentationen möglich ist. Ein weiterer Einflussfaktor scheint der musikalische Hintergrund der Vpn zu sein, denn die besten Produktionsergebnisse der hier beschriebenen Experimente wurden von musikalisch aktiven Vpn (den Amateuren aus dem Repertoirechor und den professionellen Instrumentalmusikern) erzielt.

Angesichts der Vielzahl der untersuchten Stichproben und Melodien stellte sich die Frage, ob ein Gesamtvergleich aller Daten möglich ist, um die Einflüsse der erwähnten Faktoren nicht nur „per Augenmaß“ festzustellen, sondern statistisch abzusichern. Dieser Vergleich wird im Folgenden beschrieben. In die Analyse werden sowohl mehrere Merkmale der Melodien eingehen als auch Merkmale des Lernens und der musikalischen Expertise der Versuchspersonen. Dabei wird es auch um die Frage gehen, welche Rolle neben der musikalischen Ausbildung die mehr oder weniger ausgeprägte Fähigkeit, Töne zu benennen (absolutes Gehör) für die Tonarterinnerung spielt. Denn denkbar wäre es, dass die Instrumentalmusiker auch deshalb die höchste Trefferquote erzielten, weil sie die höchste Quote an Absoluthörern besitzen und nicht nur, weil ihre Melodien am vertrautesten sind.

6.10 Methodik der zusammenfassenden Analyse

Um die Daten aller Produktions-Experimente in einer zusammenfassenden Analyse untersuchen zu können und so Einflussfaktoren auf die Tonarterinnerung ausfindig zu machen, war eine numerische Charakterisierung sowohl der musikalischen Expertise der Vpn als auch der potenziell relevanten Eigenschaften der Melodien sowie der relevanten Lernfaktoren notwendig.

6.10.1 Charakterisierung der Versuchspersonen

In allen Experimenten wurden die Vpn nach der Dauer ihrer Instrumentalausbildung, gemessen in Jahren, befragt. Dieses Maß wird für den Gesamtvergleich als Expertise-Kriterium

genutzt, da es mehr Varianz erzeugt als eine bloße Einteilung der Vpn in „Musiker“ und „Nicht-Musiker“. Schließlich gibt es viele „Nicht-Musiker“, die jahrelang Instrumentalunterricht erhielten und zum Teil quasi professionelle Leistungen erbringen, obgleich sie einen nicht-musikalischen Beruf ergriffen haben. Bei den Vpn aus den Chor-Stichproben wurden weitere Merkmale hinsichtlich ihrer musikalischen Ausbildung erhoben, z.B. Chorerfahrung und Gesangsunterricht. Diese liegen aber erstens nicht für alle Vpn vor und spielen zweitens vermutlich nur für die intentional gelernten Melodien eine Rolle. Daher wurden sie in den Gesamtvergleich nicht einbezogen, sondern statt dessen in den Experimenten mit Chor-Stichproben einzeln analysiert.

Neben der Dauer der Instrumentalausbildung liegen für alle Vpn Screening-Test-Daten zum absoluten Gehör vor, die in den Gesamtvergleich einbezogen wurden. Der Screening-Test auf absolutes Gehör bestand aus 12 Dreieckstönen, die jeweils für 4 Sekunden (inkl. 200 ms Fade-in und Fade-out) dargeboten wurden. Es wurden künstliche Töne dargeboten, weil sich in Experimenten mit Absoluthörern gezeigt hat, dass die Benennungsleistung u.a. abhängig von der Klangfarbe ist: Töne in vertrauten Klangfarben werden häufiger richtig benannt als Töne in weniger vertrauten Klangfarben (vgl. Kap. 2.2.1). Da in einem kurzen Screening-Test die Nutzung vieler verschiedener Klangfarben nicht möglich war, der Test aber auch möglichst „gerecht“ für die Spieler der verschiedenen Instrumente sein sollte, wurden die allen Vpn gleichermaßen fremden Dreieckstöne als Klangfarbe ausgewählt. Die Vpn wurden gebeten, jeden dargebotenen Ton zu benennen. Dafür wurde ihnen maximal 10 Sekunden Zeit gegeben. Die zeitliche Begrenzung des Urteils wurde vorgenommen, um die (mehr Zeit erfordernde) Nutzung relationaler Strategien wie z.B. die Orientierung am eigenen Stimmumfang bei der Tonbenennung zu verhindern und nur absolute Tonbenennungen zu erfassen. Nach den 10 Sekunden erklangen Störklänge (Klaviertöne in zufälliger Reihenfolge), die den jeweils davor gehörten Ton aus dem Arbeitsgedächtnis löschen sollten. Die Vpn wurden angewiesen, nicht weiter über den soeben gehörten Ton nachzudenken, sobald die Störklänge ertönten. Außerdem wurden sie angewiesen, möglichst nur absolute Tonurteile zu fällen, also nicht zu versuchen, Intervalle zu vorher gehörten Tönen zu bilden. Ein Trial verlief folgendermaßen:

4 s Tondarbietung

10 s Reaktion (Tonurteil)

5 s Störklänge

5 s Pause (darauf folgt der nächste Ton)

Die Dreieckstöne wurden mit der Software Cool Edit (Syntrillium) hergestellt und mit den Störklängen und Urteilspausen nach dem beschriebenen Zeitschema zusammengeschnitten. Den Vpn wurde der Screening-Test von der CD über Kopfhörer vorgespielt, so dass ein gleicher zeitlicher Ablauf für alle Vpn gesichert werden konnte. Das Ergebnis des Screening-Tests war die Anzahl korrekter Antworten, die maximal 12 betragen konnte.

6.10.2 Charakterisierung der Lern-Faktoren

Zwei Faktoren, die sich auf das Lernen der Melodien beziehen, wurden in den Gesamtvergleich mit einbezogen. Zum Einen wurde die Lernart charakterisiert als intentional oder inzidentell. Als intentional gelernt wurden alle Melodien charakterisiert, die aktiv, also durch das Singen im Chor oder das Spielen von Instrumenten gelernt wurden. Als inzidentell gelernt wurden alle Melodien charakterisiert, die vorrangig durch das Hören von Tonträgern, Radio, etc. gelernt wurden. Einen Sonderfall stellen die Lieblingslieder dar, die zwar durch das Hören von Tonträgern gelernt, aber auch häufig mitgesungen werden. Für den Gesamtvergleich wurden Lieblingslieder als inzidentell gelernt charakterisiert, weil angenommen wurde, dass sie zwar häufig mitgesungen, aber nicht mit der gleichen Intensität aktiv geübt werden wie Chor- oder Instrumentalmelodien. Bezüglich eines Einflusses der Lernart auf die Tonarterinnerung wurde in Kap. 5 keine Hypothese formuliert, diese Variable wurde als Kontrollvariable in die Gesamtanalyse einbezogen.

Zum Anderen wurde die Vertrautheit der jeweiligen Melodien charakterisiert als mittel oder hoch. Ideal wäre eine individuelle Einschätzung der Vpn hinsichtlich der Vertrautheit jeder produzierten Melodie gewesen, die eine feinere Kategorisierung der Vertrautheit ermöglicht hätte. Da eine solche Einschätzung nicht bei allen Stichproben vorlag, konnte die Vertrautheit nur im Nachhinein als mittel (= seit einigen Monaten bis höchstens einem Jahr bekannt) oder hoch (= seit über einem Jahr bekannt) kategorisiert werden.

6.10.3 Charakterisierung der Melodien

Eine Charakterisierung der Melodien war notwendig, um die Ergebnisse der verschiedenen Experimente vergleichbar zu machen. Eine Möglichkeit wäre es gewesen, einen Faktor „Melodie“ in die Gesamtanalyse einzuführen und die Melodien nur zu benennen. Obgleich dieses Vorgehen einfach und wenig fehleranfällig gewesen wäre, wurde es hier nicht gewählt, weil man an einem signifikanten „Melodie“-Effekt nicht weiter ablesen kann, worauf er beruht – Melodien können durch viele verschiedene Parameter gekennzeichnet werden und sich in sehr vielen Merkmalen unterscheiden. Daher wurde die kompliziertere Variante einer Parametrisierung der Melodien durch musikalische Experten gewählt. Durch die Parametrisierung sollten die Melodien nicht etwa vollständig „abgebildet“, sondern theoriegeleitet anhand für die Tonarterinnerung potenziell relevanter Merkmale gekennzeichnet werden. Daher wurde die Anzahl an Merkmalen möglichst klein gehalten. Dies sollte erstens die Beurteilungs-Prozedur in einem überschaubaren Rahmen halten, und zweitens sollten vorzugsweise solche Merkmale bewertet werden, für die sich Hypothesen bezüglich ihres Einflusses auf die Tonarterinnerung generieren lassen.

Ein Merkmal, das sich angesichts der Datenlage als relevant erwiesen hat, ist die Prägnanz von Melodien. Daher wurde nach Kriterien gesucht, anhand derer man die Prägnanz von Melodien kennzeichnen kann. Das Prägnanzerleben wurde auf der visuellen Ebene bereits häufiger untersucht (u.a. Attneave, 1954; Rausch, 1966; Prinz, 1966; Stadler et al., 1979; Hüppe, 1984). Bezogen auf Melodien stammt die erste umfangreichere Untersuchung von

Winkelhaus (2004), die sich an das Prägnanzkonzept von Rausch (1966) anlehnt. Dieses unterscheidet zwischen formal-syntaktischen Prägnanzaspekten wie z.B. Gesetzmäßigkeit, Integrität oder Komplexität und inhaltlichen Aspekten wie der Ausdrucks- und Bedeutungsfülle. Unter den Erstgenannten lässt sich nochmals eine Unterteilung in Gesetzmäßigkeits-Aspekte und Fülle-Aspekte treffen. Aufgrund ihrer Hörversuche, die eine verbale Beurteilung verschiedener Melodien beinhalteten, kommt Winkelhaus (2004) zu dem Ergebnis, dass sich die Kriterien von Rausch, die für die visuelle Wahrnehmung entwickelt wurden, nicht ohne weiteres auf die Melodiewahrnehmung übertragen lassen. Statt dessen haben sich die Kriterien Eingängigkeit, Regelhaftigkeit, Ausdruck, Bedeutung und Gefallen als besonders relevant und reliabel für die Kennzeichnung melodischer Prägnanz herausgestellt.

Für die Beurteilung der Melodien wurden die Kriterien Eingängigkeit und Regelhaftigkeit als Gesetzmäßigkeits-Aspekte und die Kriterien Ausdrucksstärke und Bedeutungsfülle als inhaltliche Aspekte ausgewählt. Auf eine Beurteilung des Gefallens wurde verzichtet, erstens, da bei Winkelhaus das Gefallen sehr hoch mit der Ausdrucksstärke korreliert war, und zweitens, da es für den Gefallens-Aspekt problematisch erschien, ihn durch einen Beurteiler im nachhinein valide einzuschätzen, da letztlich nur die persönliche Einschätzung durch die entsprechende Versuchsperson relevant ist. Weiterhin wurden für die Melodiebeurteilung noch zwei Fülle-Aspekte ausgewählt, die auch als Maße für die Schwierigkeit von Melodien verstanden werden können. Da sich der von Winkelhaus untersuchte Fülle-Aspekt „Komplexität“ als kein reliables Kriterium zur Bestimmung melodischer Prägnanz herausgestellt hat, wurden hier die folgenden zwei Fülle-Maße ausgewählt: Zum Einen wurde der horizontale Umfang der Melodie als Umfang des charakteristischen Motivs bestimmt und zum Anderen der vertikale Umfang als Tonumfang. In beiden Aspekten gibt es große Unterschiede zwischen den genutzten Melodien, die einen Einfluss auf die Tonarterinnerung haben könnten.

Die 6 Melodiemerkmale Eingängigkeit, Regelhaftigkeit, Ausdrucksstärke, Bedeutungsfülle, Motiv-Umfang und Tonumfang wurden 3 Beurteilern vorgelegt. Diese bewerteten die insgesamt 137 Melodien aus allen Experimenten nach diesen Merkmalen. Die Beurteilung erfolgte im Einzelversuch. Den Beurteilern wurde eine standardisierte Instruktion vorgelegt, in der sie angewiesen wurden, jede Melodie anhand jedes Merkmals auf einer Skala von 1 bis 7 zu bewerten. Dabei entsprach die 7 einer vollen Ausprägung eines Merkmals (z.B. extremer Regelhaftigkeit), die 1 einer nicht vorhandenen Ausprägung eines Merkmals (z.B. keine Regelhaftigkeit erkennbar), während die 4 einer mittleren Merkmalsausprägung entsprach. Die Melodien wurden in der Form vorgelegt, in der sie am leichtesten zu beurteilen sind. Bei den Melodien, die in Notenform vorliegen (sämtliche Chorlieder und die Instrumentalwerke), wurden die Noten vorgelegt. Auf eine akustische Darbietung wurde hier verzichtet, da nur die jeweils relevante Stimme beurteilt werden sollte, die aus dem Gesamtklang nicht immer einzeln herauszuhören ist. Dies gilt insbesondere für viele Orchesterstellen, aber auch für die Mittelstimmen der meisten Chorstücke. Melodien, für die keine notierten Fassungen vorliegen (Lieblingslieder, Telekom-Jingle und Handymelodien), wurden akustisch dargeboten, da hier die relevante Stimme immer gut heraushörbar war. Die Reihenfolge der dargebotenen Melodien wurde zwischen den drei Beurteilern variiert, um Reihenfolgeeffekte zu verringern.

Als Beurteiler wurden drei musikalische Experten ausgewählt, da es wichtig war, dass die Beurteiler ein breites musikalisches Hintergrundwissen haben. Dies spielt z.B. bei der Beurteilung der Regelhaftigkeit eine entscheidende Rolle, da sie nur aufgrund der Kenntnis musikalischer Regeln beurteilbar ist. Außerdem waren Experten gefragt, weil die Durchführung der Beurteilung es erforderte, dass man sich notierte Melodien klanglich vorstellen kann. Bei den Beurteilern handelt es sich um einen Musiktheoretiker, eine Musikwissenschaftlerin und einen Orchestermusiker.

Nach erfolgter Melodiebeurteilung wurde zunächst die Beurteilerübereinstimmung über die Berechnung der Intraclass-Korrelationen bestimmt. Folgende Werte wurden erzielt:

Eingängigkeit	$r = .61$	Regelhaftigkeit	$r = .46$
Ausdrucksstärke	$r = .80$	Bedeutungsfülle	$r = .73$
Umfang des Motivs	$r = .50$	Tonumfang der Melodie	$r = .81$

Um die Melodiemerkmale auszuwählen, die in die Gesamtanalyse eingehen sollten, wurden die Werte der Intraclass-Korrelationen, die Korrelationen der 6 Merkmale untereinander (gemittelte Urteile), sowie Beobachtungen während der Melodiebeurteilung berücksichtigt. Letztere zeigten, dass einige Merkmale mehr Probleme bereiteten als andere. Während die Beurteilung der Eingängigkeit den Beurteilern wenig Probleme bereitete, gab es bei der Beurteilung der Regelhaftigkeit zum einen das Problem, dass sie stark abhängig ist vom musikalischen Vorwissen und somit von den Regeln, die ein Beurteiler kennt - z.B. fragte Beurteiler 1 nach, ob er den Perusio-Kanon aufgrund seines Wissens über die Musik des 14. Jahrhunderts oder aufgrund der Annahme, was der musikalische „Ottonormalverbraucher“ wissen könnte, beurteilen soll. Von Beurteiler 3 dagegen gab es den Kommentar, dass alle hier beurteilten Melodien „ziemlich regelhaft“ seien, mit der einzigen Ausnahme der teilweise durchaus zufällig wirkenden Handymelodien. Bei der Beurteilung der inhaltlichen Melodieaspekte Ausdruck und Bedeutung zeigte sich, dass es durchaus möglich ist, die Ausdrucksstärke eines kleinen Melodieausschnitts zu beurteilen, dass es aber Probleme bereitete, die Bedeutungsfülle bei einem Melodieausschnitt zu beurteilen, dessen Kontext man sehr gut kennt. Dies war z.B. der Fall beim Weihnachtsoratorium und bei vielen Orchesterstellen, hier hat die Bedeutung des ganzen Werks die des Melodieausschnitts quasi „überschattet“. Die Beurteilung des Tonumfangs wiederum fanden die Beurteiler insgesamt leichter als die Beurteilung des Motivumfangs. Bei letzterem waren die vorgespielten Ausschnitte aus den Lieblingsliedern besonders problematisch, weil hier nur die ersten 5-10 Sekunden der Lieder vorlagen und charakteristische Motive teilweise abgeschnitten waren. Dies liegt daran, dass die Liedanfänge ursprünglich nur zum Zweck der Tonhöhenanalyse aufgenommen wurden, für die ein sehr kleiner Ausschnitt ausreicht.

Korrelationsanalysen zwischen den 6 Melodiemerkmalen mit den über alle Beurteiler gemittelten Werten ergaben eine hohe Korrelation zwischen Eingängigkeit und Regelhaftigkeit ($r = .72$, $p < .001$) sowie zwischen Ausdrucksstärke und Bedeutungsfülle ($r = .87$, $p < .001$) und zwischen Tonumfang und Motivumfang ($r = .86$, $p < .001$). Aufgrund der niedrigen Beurtei-

lerübereinstimmung wurden die Regelhaftigkeit und der Motivumfang von den weiteren Analysen ausgeschlossen. Aufgrund der hohen Korrelation mit Ausdrucksstärke und der beobachteten Schwierigkeiten bei der Beurteilung wurde auch die Bedeutungsfülle von den weiteren Analysen ausgeschlossen. Die Eingängigkeit dagegen wurde trotz der unbefriedigenden Beurteilerübereinstimmung in die weiteren Analysen einbezogen, weil ansonsten kein Gesetzmäßigkeits-Aspekt mehr vertreten wäre, der inhaltlich durchaus von Interesse ist. Somit werden die Melodien im Gesamtvergleich durch die Merkmale Eingängigkeit, Ausdrucksstärke und Tonumfang charakterisiert, womit sowohl Gesetzmäßigkeitsaspekte als auch inhaltliche und Fülleaspekte des Prägnanzerlebens berücksichtigt werden.

Bei der Formulierung von Annahmen bezüglich des Einflusses der Melodiemerkmale auf die Tonart-Erinnerung kann nur auf wenige Literaturbefunde zurückgegriffen werden. Auf der visuellen Ebene konnten Stadler et al. (1979) auf der Grundlage eines „Transformationsscheinbewegungs-Experiments“ nachweisen, dass eine prägnante oder ausgezeichnete Figur eine geringere „Identitätstoleranz“ hat als eine weniger prägnante oder nicht-ausgezeichnete Figur. Auf Melodien übertragen könnte man darauf aufbauend annehmen, dass prägnante Melodien sich weniger leicht transponieren lassen, also eher in der Originaltonart repräsentiert sein sollten als weniger prägnante Melodien. Winkelhaus (2004) fand heraus, dass eine prägnante Melodiegestaltung weitgehend, aber nicht extrem gesetzmäßig und einfach ist und einen mittleren Komplexitätsgrad aufweist. Außerdem wurde eine prägnante Melodiegestaltung als stark ausdrucks- und bedeutungsvoll beurteilt.

Vor dem Hintergrund dieser Befunde wurde für die Gesamtanalyse angenommen, dass die Tonarterinnerung an Melodien, deren Ausdrucksstärke als hoch eingeschätzt wurde, genauer ist als die an Melodien mit niedriger Ausdrucksstärke. Nimmt man an, dass für die in Winkelhaus’ erstem Experiment nicht untersuchte Eingängigkeit ähnliche Prinzipien gelten wie für Gesetzmäßigkeit, Einfachheit und Komplexität, so lässt sich weiterhin annehmen, dass die Tonarterinnerung an Melodien mit mittlerer Eingängigkeit genauer ist als die an Melodien mit extrem hoher oder niedriger Eingängigkeit. Bezüglich eines Einflusses des Tonumfangs von Melodien konnten keine Vorhersagen getroffen werden, weil es diesbezüglich weder theoretische noch empirische Vorbefunde gibt.

6.11 Ergebnisse der zusammenfassenden Analyse

In der folgenden Tabelle sind die Ergebnisse aller vorgestellten Produktions-Experimente zusammengefasst. Dabei werden die Versuchspersonen anhand der mittleren Dauer ihrer Instrumentalausbildung und der mittleren Trefferquote einer Stichprobe im Test auf absolutes Gehör charakterisiert. Die Lernart wird als intentional oder inzidentell charakterisiert und die Vertrautheit als mittel oder hoch. Die produzierten Melodien werden hinsichtlich der (ggf. mittleren) Eingängigkeit, Ausdrucksstärke und des Tonumfangs charakterisiert. Als Ergebnisse werden sowohl die mittleren Fehlerbeträge als auch die Mittelpunkte der jeweiligen Verteilung 1 (in den Chor-Stichproben mit Berücksichtigung von Oktavfehlern) angegeben, sowie die jeweiligen Standardfehler, um zu verdeutlichen, wie breit die Fehler streuen. Die

Angaben zur Signifikanz beruhen auf den Ergebnissen des chi²-Tests (bzw. bei einer Stichprobe mit n < 26 des Kolmogorov-Smirnov-Tests) für Verteilung 1 sowie des t-Tests für Verteilung 2. Um die Ergebnisse anschaulicher zu machen und einen Vergleich mit Levitins (1994) Ergebnissen zu ermöglichen, ist außerdem der Prozentsatz an Vpn notiert, die bei der Produktion der jeweiligen Melodie um maximal einen Halbton falsch sangen. Bei Levitin (1994) waren es 57 bzw. 51 % der Vpn, die in den zwei Durchgängen dieses Kriterium erfüllten.

Tab. 6.1: Ergebnisse der Produktions-Experimente

Experiment	Versuchs-personen	Lernart	Melodien	Ergebnisse
Lieblings-lieder	n = 30 I = 2,10 AG = 0,33	inzidentell V = mittel / hoch	Popsongs: Lieblingslieder & Evergreens E = 5,97 A = 3,69 T = 3,52	m = 3,43 (SE = 0,36) mv = - 0,50 (SE = 0,73) chi² p = .238 t-Test p = .240 16,6 % korrekt
Telekom-Jingle	n = 71 I = 5,54 AG = 0,86	inzidentell V = hoch	Telekom-Werbejingle E = 6,67 A = 5,67 T = 2,33	m = 2,39 (SE = 0,22) mv = - 0,37 (SE = 0,36) chi² p = .054 t-Test p = .007 40,9 % korrekt
Handy-melodie	n = 49 I = 2,41 AG = 0,00	inzidentell V = mittel	eigene Handymelodie E = 5,27 A = 2,35 T = 3,75	m = 2,96 (SE = 0,29) mv = - 0,39 (SE = 0,52) chi² p = .012 t-Test p = .890 30,7 % korrekt
kurze Zeit in der Schule gelernte Melodien	n = 77 (Melodie 1) I = 6,18 AG = 0,97	intentional V = mittel	Melodie 1: bekannter Kanon E = 6,67 A = 5,0 T = 5,33	m = 4,20 (SE = 0,27) mv = - 4,10 (SE = 0,29) chi² p = .000 t-Test p = .070 18,2 % korrekt
	n = 70 (Melodie 2) I = 6,46 AG = 0,97	intentional V = mittel	Melodie 2: Kanon 14. Jh E = 3,0 A = 4,33 T = 5,33	m = 2,63 (SE = 0,23) mv = - 2,42 (SE = 0,25) chi² p = .000 t-Test p = .002 40 % korrekt
kurze Zeit im Chor gelernte Melodien	n = 24 I = 11,5 AG = 2,08	intentional V = mittel	Chorwerke von Bach, Poulenc, Distler, Barbe E = 5,62 A = 4,58 T = 4,43	m = 3,13 (SE = 0,54) mv = - 3,04 (SE = 0,56) KST p = .021 t-Test p = .102 33,3 % korrekt

Experiment	Versuchs-personen	Lernart	Melodien	Ergebnisse
lange Zeit im Chor gelernte Melodien	n = 37 (Jauchzet) I = 8,36 AG = 1,57	intentional V = hoch	Eingangschor „Jauchzet, frohlocket“ E = 7,0 A = 6,67 T = 3,67	m = 2,08 (SE = 0,37) mv = - 1,43 (SE = 0,44) chi² p = .000 t-Test p = .000 56,7 % korrekt
	n = 34 (Choral) I = 8,40 AG = 1,71	intentional V = hoch	Choral aus dem Weihnachtsoratorium E = 5,43 A = 5,34 T = 5,17	m = 3,71 (SE = 0,45) mv = - 3,41 (SE = 0,52) chi² p = .455 t-Test p = 1,00 23,5 % korrekt
Instrumen-talmusik	n = 40 I = 16,6 AG = 2,93	intentional V = hoch	Mozart-Konzert E = 5,82 A = 4,61 T = 5,73	m = 1,63 (SE = 0,28) mv = - 0,38 (SE = 0,38) chi² p = .000 t-Test p = .000 67,5 % korrekt
		intentional V = hoch	Orchesterstelle E = 5,78 A = 5,05 T = 4,91	m = 2,08 (SE = 0,29) mv = - 0,03 (SE = 0,44) chi² p = .005 t-Test p = .003 52,5 % korrekt

Die Vpn einer Stichprobe werden charakterisiert durch die mittlere Dauer (in Jahren) des Instrumentalunterrichts (I) sowie den mittleren Score (Anzahl korrekt benannter Töne von 12) im Test auf absolutes Gehör (AG). Die Lernart wird charakterisiert als inzidentell oder intentional und die Vertrautheit (V) als mittel oder hoch. Die in einem Experiment produzierten Melodien werden charakterisiert nach ihren (ggf. über alle Melodien des Experiments gemittelten) Werten für Eingängigkeit (E), Ausdrucksstärke (A) und Tonumfang (T), jeweils auf einer Skala von 1-7, bei der die 7 der vollen Ausprägung der Eigenschaft entspricht. Die Ergebnisse werden dargestellt anhand von:

m = mittlere Fehlergröße in Halbtönen, SE = Standardfehler des Mittelwerts,
mv = Mittelwert der Fehlerverteilung 1, SE = Standardfehler des Mittelwerts,
chi² = Ergebnis des chi²-Tests auf Gleichverteilung,
KST = Ergebnis des Kolmogorov-Smirnov-Tests auf Gleichverteilung,
t-Test = Ergebnis des Einstichproben t-Tests gegen den Wert 3, und
% korrekt = Prozentsatz der Vpn, die maximal einen Halbton neben der Original-Tonart sangen.

Um zu überprüfen, ob die musikalische Expertise, die Fähigkeit Töne zu benennen, die Lern-Faktoren sowie die drei Melodiemerkmale einen signifikanten Einfluss auf die Genauigkeit der Tonarterinnerung haben, wurde der Einfluss jedes Merkmals auf die Tonart-Produktion zunächst einzeln analysiert. Der Grund für dieses Vorgehen ist das Vorliegen eines Untersuchungsplans mit teilwiederholten Messungen, der die Analyse über die mehrfaktorielle Varianzanalyse erschwert, da diese unabhängige oder aber vollständig wiederholte Messungen erfordert. Auf der einen Seite produzierte ca. die Hälfte der Vpn zwei oder drei verschiedene Melodien (in den Experimenten mit den Schülern, den Weihnachtsoratorium-Sängern sowie den Instrumentalmusikern). Auf der anderen Seite gab es einige Experimente, in denen sehr viele bzw. alle Vpn die gleiche Melodie produzierten (dies betrifft die gleichen Experimente). Bortz (persönliche Mitteilung) empfiehlt für dieses Design eine gesonderte Analyse pro Merkmal, auch wenn dadurch nicht alle Interaktionen analysierbar sind. Trotzdem wird nach der „Einzelanalyse“ explorativ eine mehrfaktorielle Varianzanalyse durchgeführt, um Interaktionen der verschiedenen Einflussgrößen auszuschließen.

6.11.1 Expertise- und Tonbenennungseffekte

Um den Einfluss der musikalischen Expertise auf die Genauigkeit der Tonart-Produktion zu untersuchen, wurden zunächst für jene Vpn, die mehrere Melodien produzierten, die Differenzen aller Produktionen gemittelt, so dass keine Vp doppelt in die Analysen einging. Für die Instrumentalmusiker und Weihnachtsoratorium-Sänger wurden zwei Produktionen gemittelt und für die Schüler drei Produktionen. Dadurch entstand eine Stichprobe von 268 Vpn. Um herauszufinden, ob der prognostizierte kleine bis mittlere Zusammenhang zwischen musikalischer Expertise und Tonarterinnerung (Hypothese 2) sowie zwischen Tonbenennung und Tonarterinnerung (Hypothese 3) vorliegt, wurden zwei Korrelationsanalysen durchgeführt. Abb. 6.13 zeigt die Zusammenhänge zunächst als Scatterplot.

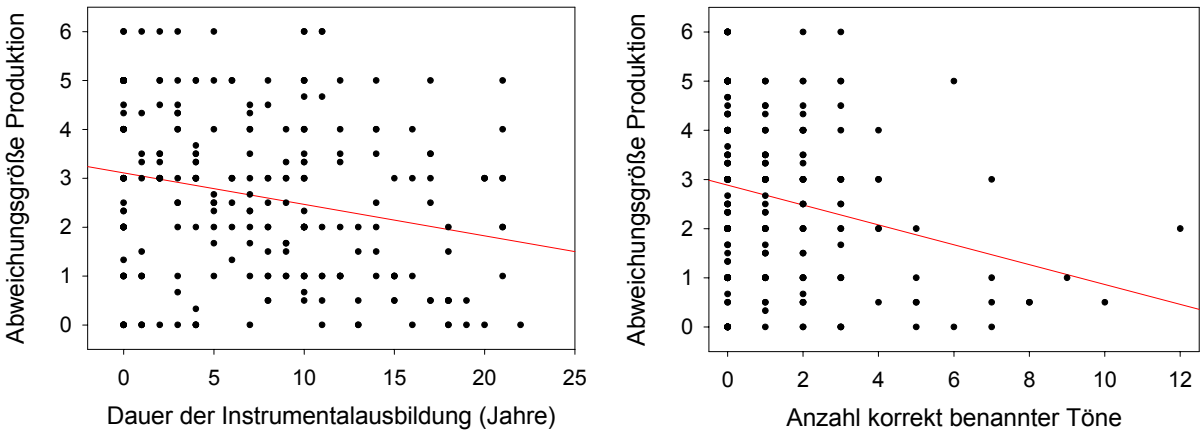

Abb. 6.13: Zusammenhänge zwischen Produktionsgenauigkeit (y-Achse: Abweichungen in Halbtönen) und der musikalischen Expertise sowie der Fähigkeit, Töne zu benennen.

Die Korrelationsanalyse ergab kleine, aber signifikante Korrelationen der Genauigkeit der Tonart-Produktion mit der Dauer der Instrumentalausbildung (r = - 0,25, p < .001) sowie mit der Leistung im Test auf absolutes Gehör (r = - 0,23, p < .001). Die Dauer der Instrumentalausbildung korrelierte dagegen höher mit der Leistung im Test auf absolutes Gehör (r = 0,46, p < .001). Das letzte Ergebnis bestätigt Literaturbefunde, die Zusammenhänge zwischen der Leistung von Absoluthörern und der musikalischen Ausbildung gefunden haben (vgl. Kap. 2.4). Außerdem wird die Hypothese eines Zusammenhangs zwischen musikalischer Expertise und Tonarterinnerung bestätigt, ebenso wie die Hypothese eines Zusammenhangs zwischen der Tonbenennungsleistung und Tonarterinnerung. Dass diese Zusammenhänge nur moderat sind, wird aus Abb. 6.13 ersichtlich: Es gibt unter den Vpn mit nur zufälliger Tonbenennungsleistung (0-1 Richtigurteile) das ganze Spektrum von sehr genauer bis zufälliger Tonarterinnerung, während insgesamt nur sehr wenige Absoluthörer untersucht wurden, bei denen jedoch insgesamt eine eher genaue Tonarterinnerung vorliegt. Dies weist darauf hin, dass die Tonbenennungsleistung zwar einen Einfluss auf die Tonart-Erinnerung hat, deren Genauigkeit alleine aber nicht erklären kann. Ähnliches gilt für die musikalische Expertise: Viele musikalisch besser ausgebildete Vpn weisen recht genaue Produktionsleistungen auf, aber auch unter den musikalisch nicht ausgebildeten Vpn findet sich das ganze Leistungsspektrum von genauer bis zufälliger Tonart-Produktion.

6.11.2 Einfluss der Lernfaktoren

Um den Einfluss der Lernart auf die Genauigkeit der Tonart-Produktion zu überprüfen, wurden die mittleren Abweichungsgrößen der intentional gelernten Melodien mit denen der inzidentell gelernten Melodien mit einem t-Test für unabhängige Stichproben verglichen. Um sicher zu stellen, dass unabhängige Stichproben vorliegen, wurden bei der einen Hälfte der Vpn des Schüler-Experiments die Produktionsdaten des Telekom-Jingles (inzidentell) und bei der anderen Hälfte die Produktionsdaten der beiden gelernten Lieder (intentional) einbezogen. Die Auswahl erfolgte zufällig. Aus den anderen beiden Experimenten mit Messwiederholungen (Weihnachtsoratorium, Instrumentalmusiker) wurden alle Daten einbezogen, da alle Daten nur in eine Gruppe (intentional) gehören und somit die Unabhängigkeit der Stichproben nicht beeinträchtigen. Es entstand eine Stichprobe mit 370 Fällen. Tabelle 6.2 zeigt die deskriptiven Daten:

Tab. 6.2: Mittlere Fehlergrößen bei der Produktion inzidentell und intentional gelernter Melodien.

	Abweichungsgröße bei der Tonartproduktion	
Lernart	inzidentell (n = 121)	intentional (n = 249)
M (HT)	2,84	2,35
SE (HT)	0,18	0,12

M = Mittelwert, SE = Standardfehler, beide in Halbtönen (HT).

Der Levene-Test zeigte, dass die Varianzen zwischen den beiden Gruppen homogen sind (F = 2,003, p = .158). Der t-Test ergab einen signifikanten Mittelwertunterschied (t (368) =

2,335, p = .020). Somit wurden die intentional gelernten Melodien signifikant besser produziert als die inzidentell gelernten Melodien.

Um den Einfluss der Vertrautheit auf die Genauigkeit der Tonart-Produktion zu überprüfen, wurden die mittleren Abweichungsgrößen der hoch vertrauten Melodien mit denen der mittelmäßig vertrauten Melodien mit einem t-Test für unabhängige Stichproben verglichen. Das Verfahren zur Sicherung unabhängiger Stichproben entsprach dem oben beschriebenen. Da auch hier eine Aufteilung der Schüler-Stichprobe auf eine Telekom-Stichprobe (hoch vertraut) und eine Stichprobe für gelernte Melodien (mittelmäßig vertraut) erfolgte, ergab sich wieder eine Stichprobe mit 370 Fällen. Tabelle 6.3 zeigt die deskriptiven Daten:

Tab. 6.3: Mittlere Fehlergrößen bei der Produktion mittelmäßig und hoch vertrauter Melodien.

	Abweichungsgröße bei der Tonartproduktion	
Vertrautheit	mittelmäßig (n = 157)	hoch (n = 213)
M (HT)	2,88	2,24
SE (HT)	0,15	0,13

M = Mittelwert, SE = Standardfehler, beide in Halbtönen (HT).

Der Levene-Test zeigte, dass die Varianzen zwischen den beiden Gruppen homogen sind (F = 0,453, p = .502). Der t-Test ergab einen signifikanten Mittelwertunterschied (t (368) = 3,208, p = .001). Somit wurden die hoch vertrauten Melodien signifikant besser produziert als die weniger vertrauten Melodien. Dies bestätigt Hypothese 1 (Häufigkeitseffekt).

6.11.3 Einfluss der Melodie-Merkmale

Um zu überprüfen, ob Zusammenhänge zwischen den Melodie-Merkmalen und der Tonarterinnerung bestehen, wurde zunächst eine graphische Darstellung gewählt, in die die Daten aller produzierten Melodien eingingen. Wie aus Abbildung 6.14 ersichtlich ist, besteht ein quasi u-förmiger Zusammenhang zwischen Eingängigkeit und Produktionsgenauigkeit: Am besten wurden Melodien mit mittlerer bis hoher Eingängigkeit produziert (die beiden niedrigsten Skalenstufen wurden bei keiner der Melodien vergeben). Der Zusammenhang zwischen Ausdrucksstärke und Produktionsgenauigkeit ist dagegen eher umgekehrt u-förmig: Am besten wurden Melodien mit sehr geringer Ausdrucksstärke produziert, am zweitbesten Melodien mit mittlerer bis hoher Ausdrucksstärke und am schlechtesten Melodien mit geringer bis mittlerer Ausdrucksstärke. Zwischen dem Tonumfang und der Produktionsgenauigkeit sind dagegen keine systematischen Zusammenhänge erkennbar. Dass bei allen drei Melodie-Merkmalen keine linearen Zusammenhänge mit der Produktionsgenauigkeit vorliegen, ist aus Abbildung 6.14 erkennbar, und es ergaben sich auch keine signifikanten Korrelationen (Produktionsgenauigkeit / Eingängigkeit: r = 0,060, p = .193; Produktionsgenauigkeit / Ausdrucksstärke: r = - 0,038, p = .407; Produktionsgenauigkeit / Tonumfang: r = 0,017, p = .717).

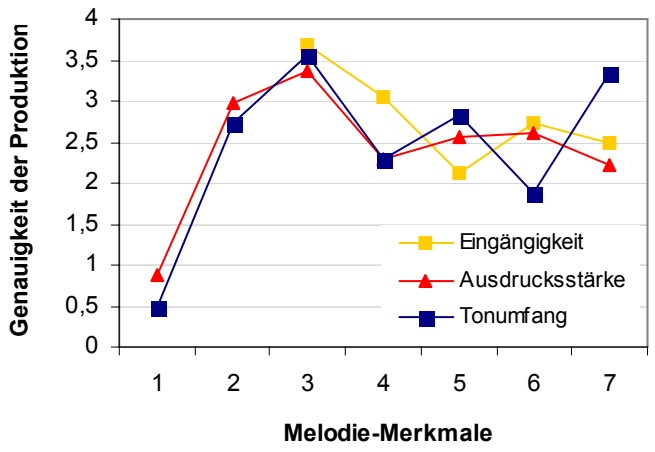

Abb. 6.14: Zusammenhänge zwischen den Melodie-Merkmalen Eingängigkeit, Ausdrucksstärke sowie Tonumfang (Bewertungen auf einer Skala von 1-7) und der Fehlergröße bei der Tonarterinnerung.

Um zu überprüfen, ob die in Abb. 6.14 sichtbaren Effekte der Melodie-Merkmale auf die Produktionsgenauigkeit signifikant sind, wurde eine Varianzanalyse mit den unabhängigen Variablen Eingängigkeit, Ausdrucksstärke und Tonumfang und der Produktionsgenauigkeit als abhängiger Variable berechnet. Die Analyse erfolgte dreifaktoriell, weil die Melodie-Merkmale miteinander korreliert und somit keine unabhängigen Faktoren sind. Die Analyse wurde in zwei Varianten durchgeführt: Zunächst wurde eine Stichprobe ohne Messwiederholungen auf Seiten der Melodien hergestellt. Diese beinhaltete die Merkmalswerte aller 137 Melodien sowie die mittlere Produktionsgenauigkeit für jede Melodie. In den Fällen, in denen eine Melodie mehrmals produziert wurde, wurden alle Produktionsfehler gemittelt. Für diese Stichprobe ergab die dreifaktorielle Varianzanalyse keine signifikanten Haupteffekte sowie keine Interaktionen der Faktoren (Tab. 6.4).

Die Aussagekraft dieser Analyse wird allerdings durch die genannte Mittelungsprozedur eingeschränkt: Dadurch werden Melodien, für die bis zu 77 Produktionen vorliegen auf einen Fall reduziert und mit Melodien verglichen, für die nur eine einzige Produktion vorliegt. Um dieses inhaltliche Problem zu umgehen, wurde eine zweite Variante mit allen 472 Fällen berechnet. Diese Analyse ist aufgrund der vorliegenden teilweisen Messwiederholungen nur explorativ, wird aber durch inhaltliche Aspekte gerechtfertigt. Die Ergebnisse zeigen (Tab. 6.5), dass der Eingängigkeits-Effekt, nicht aber die Effekte von Ausdrucksstärke und Tonumfang signifikant sind. Außerdem traten keine signifikanten Interaktionen auf.

Mit dem Ergebnis der zweiten Varianzanalyse lässt sich vorläufig die Hypothese, dass die Eingängigkeit einer Melodie die Genauigkeit der Tonarterinnerung beeinflusst, bestätigen. Dagegen lässt sich die Hypothese über den Einfluss der Ausdrucksstärke auf die Genauigkeit der Tonarterinnerung mit dieser Analyse nicht bestätigen.

Tab. 6.4: Ergebnisse der dreifaktoriellen Varianzanalyse mit 137 Fällen.

unabhängige Variable	F	df	Signifikanz
Eingängigkeit	1,127	13	.501
Ausdrucksstärke	0,484	15	.864
Tonumfang	0,527	18	.845
Eingängigkeit x Ausdrucksstärke	0,638	1	.469
Eingängigkeit x Tonumfang	1,924	1	.238
Ausdrucksstärke x Tonumfang	0,571	4	.700

Tab. 6.5: Ergebnisse der dreifaktoriellen Varianzanalyse mit 472 Fällen.

unabhängige Variable	F	df	Signifikanz
Eingängigkeit	1,883	13	.031
Ausdrucksstärke	0,627	15	.852
Tonumfang	0,941	17	.526
Eingängigkeit x Ausdrucksstärke	1,001	1	.318
Eingängigkeit x Tonumfang	2,731	1	.099
Ausdrucksstärke x Tonumfang	0,795	4	.529

6.11.4 Interaktionen zwischen den Einflussfaktoren

Da sich signifikante Expertise- und Tonbenennungs-Effekte auf die Genauigkeit der Tonarterinnerung gezeigt haben, erscheint es notwendig, die Effekte der Lernfaktoren sowie der Melodie-Merkmale auf Konfundierungen zu überprüfen. Beispielsweise traten die besten Produktionsleistungen in der Gruppe der Instrumentalmusiker auf, die nicht nur intentional gelernt, sondern auch die längste musikalische Ausbildung genossen haben. Es soll daher überprüft werden, ob die gefundenen Effekte der Lernart, der Vertrautheit und der Eingängigkeit auch dann signifikant sind, wenn die Expertise-Variablen kontrolliert werden. Zu diesem Zweck wurde explorativ eine mehrfaktorielle Varianzanalyse durchgeführt, in die alle dichotomen Variablen als feste Faktoren eingingen und alle stetigen Variablen als Kovariaten. Demnach wurden als feste Faktoren die Lernart (inzidentell vs. intentional) und die Vertrautheit (mittel vs. hoch) festgelegt, während die Dauer der Instrumentalausbildung, die Leistung in der Tonbenennung, die Eingängigkeit, Ausdrucksstärke und der Tonumfang der Melodien als Kovariaten in die Analyse eingingen.

In die Gesamtanalyse gingen alle 472 Fälle ein, da es inhaltlich zu problematisch erschien, einzelne Experimente durch die Mittelung aller Produktionen einer Melodie auf einen Fall zu reduzieren und somit gemittelte Produktionen mit Einzelwerten zu vergleichen. Wiederholte Melodien werden also aus inhaltlichen Gründen eingeschlossen, aber auch aus der Überlegung heraus, dass gleiche Melodien durch die Kennzeichnung mit gleichen Merkmalswerten kenntlich gemacht werden. Vor der Annahme, dass mit den drei Melodie-Merkmalen wichtige Melodie-Eigenschaften bezüglich der abhängigen Variablen erfasst werden, erscheint der mehrfache Einbezug einiger Melodien in die Analyse gerechtfertigt. Aus dem gleichen Grund werden die doppelten Produktionen einiger Vpn in dieser Analyse nicht gemittelt (wie in den Korrelationsanalysen), denn mit den beiden Expertise-Variablen sind wichtige Eigenschaften der Vpn bezüglich der abhängigen Variablen erfasst und doppelte Vpn gekennzeichnet.

Der Levene-Test ergab homogene Fehlervarianzen zwischen den Faktorstufen (F (3, 468) = 1,177, p = .318). Die Varianzanalyse zeigte, dass der Effekt der Instrumentalausbildungsdauer auf die Produktionsgenauigkeit signifikant war (F (1,463) = 8,236, p = .004), und ebenso der Effekt der Tonbenennungsleistung (F (1,463) = 4,280, p = .039). Es ergab sich kein signifikanter Haupteffekt für Lernart (F (1,463) = 1,983, p = .160), während die Vertrautheit signifikant wurde (F (1,463) = 6,350, p = .012) und es zwischen diesen beiden Faktoren keine signifikante Interaktion gab (F (1,463) = 0,378, p = .539). Von den Melodie-Merkmalen hatte die Eingängigkeit einen signifikanten Effekt auf die Produktionsgenauigkeit (F (1,463) = 5,151, p = .024), ebenso der Tonumfang (F (1,463) = 4,975, p = .026), nicht aber die Ausdrucksstärke (F (1,463) = 1,707, p = .192).

Das Ergebnis der Varianzanalyse weist darauf hin, dass der oben berichtete Effekt der Lernart durch die musikalische Expertise konfundiert ist und verschwindet, sobald diese kontrolliert wird. Der berichtete Effekt der Vertrautheit bleibt dagegen auch bei Kontrolle der musikalischen Expertise erhalten. Somit wird Hypothese 1 (Häufigkeitseffekt) durch diese Ergebnisse bestätigt. Auch die Hypothesen 2 (Expertiseeffekt) und 3 (Tonbenennungseffekt) werden durch die Ergebnisse der Gesamtanalyse bestätigt. Unter den Melodie-Merkmalen wurde durch die Varianzanalyse das Ergebnis der Einzelanalyse bezüglich der Eingängigkeit und der Ausdrucksstärke bestätigt, denn es zeigte sich bei ersterer, nicht aber bei letzterer, ein Effekt. Der Effekt des Tonumfangs dagegen wurde erst in der Gesamtanalyse sichtbar.

6.12 Diskussion der zusammenfassenden Analyse

Der Vergleich der Produktionsdaten aller Experimente erbrachte drei Hauptergebnisse. Erstens zeigte sich ein moderater Zusammenhang zwischen der Genauigkeit der Tonarterinnerung und musikalischer Expertise sowie der Leistung in der Tonbenennung. Vpn mit längerer Instrumentalausbildung sowie mit besseren Leistungen in der Tonbenennung trafen die Tonarten insgesamt genauer als Vpn ohne Instrumentalausbildung und mit schlechten Tonbenennungs-Leistungen. Zweitens zeigte sich ein Einfluss der Vertrautheit mit dem abzurufenden Material, aber kein Einfluss der Lernart. Es spielt für die Produktionsleistung einer Vp also keine Rolle, ob sie die Melodie durch häufiges Hören oder durch häufiges eigenes Mu-

sizieren gelernt hat. Dagegen werden Melodien, die schon seit über einem Jahr regelmäßig gehört, gesungen oder gespielt werden, deutlich genauer produziert als Melodien mit kürzerer Bekanntheitsdauer. Drittens zeigte sich ein Einfluss der Eingängigkeit gelernter Melodien, wohingegen sich kein Einfluss der Ausdrucksstärke gelernter Melodien zeigte sowie ein nur eingeschränkt interpretierbarer Einfluss des Tonumfangs gelernter Melodien.

Da sich die gefundenen Expertise- und Tonbenennungs-Effekte sowohl in der Korrelationsanalyse als auch in der Varianzanalyse zeigten, kann davon ausgegangen werden, dass es sich hier um stabile Effekte handelt. Diese sind vor dem Hintergrund der in Kapitel 4 berichteten gedächtnispsychologischen Befunde hypothesenkonform. Die Tatsache, dass sich hier solche Zusammenhänge zeigen ließen, ist sicherlich durch die große und hinsichtlich musikalischer Expertise durchmischte Stichprobe erklärbar, die musikalische Laien ebenso wie professionelle Musiker enthielt. In den Experimenten von Halpern (1989) und Levitin (1994) wurden hinsichtlich musikalischer Expertise homogenere Stichproben untersucht, in denen sich Expertise- und auch Tonbenennungs-Effekte nur schwer nachweisen lassen. Allerdings zeigt sich auch in der hier untersuchten großen Stichprobe, dass die Zusammenhänge nicht sehr hoch sind. Unter den hochtrainierten Musikern, die überzufällig gut Töne benennen können, fanden sich zwar überwiegend gute Produktionsleistungen, aber unter den weniger trainierten Vpn mit zufälligen Tonbenennungsleistungen fand sich das ganze Spektrum an Produktionsleistungen von perfekt bis zufällig. Dies verweist darauf, dass die Fähigkeit, Töne zu kategorisieren und zu benennen, zwar hilfreich für das Tonartgedächtnis ist, es aber nicht alleine ausmacht. Somit ist eine Interpretation der genauen Tonarterinnerung als „latentes“ absolutes Gehör meiner Meinung nach irreführend, weil es nicht sehr hoch mit dem „manifesten“, also durch Tonbenennungsleistungen nachweisbaren, absoluten Gehör korreliert, sondern sich noch mehrere andere wichtige Einflussfaktoren nachweisen lassen.

Einer der weiteren wichtigen Einflussfaktoren auf die Tonarterinnerung ist die Vertrautheit mit den abzurufenden Melodien. Durch den gefundenen Vertrautheits- bzw. Häufigkeitseffekt werden die in Kap. 4 berichteten gedächtnispsychologischen Befunde, die eine Abhängigkeit der Lebensdauer von Gedächtnisinhalten von der Übungsintensität während der Lernphase fanden, bestätigt und auf die Tonarterinnerung erweitert. Die Ergebnisse bezüglich der Lernart-Einflüsse auf die Tonarterinnerung sprechen dafür, dass es für diese weniger entscheidend ist, *wie* eine Melodie gelernt wurde, als vielmehr, *wie häufig* sie geübt wurde.

Einen dritten Einflussfaktor auf die Tonarterinnerung machen verschiedene Merkmale der Melodien aus. Die Effekte der Melodie-Merkmale sind aufgrund ihrer Erfassung am vorsichtigsten zu interpretieren. Schließlich ist anzunehmen, dass nicht z.B. die von den Beurteilern empfundene Eingängigkeit entscheidend ist, sondern die von der individuellen Versuchsperson empfundene. Gleiches gilt für die Ausdrucksstärke. Der Tonumfang kann zwar einigermaßen objektiv durch die Rahmentöne der Melodie beurteilt werden, trotzdem gibt es aber auch für den Tonumfang ein individuelles Charakteristikum, dass hier nicht erfasst werden konnte. Dies betrifft die Relation des Tonumfangs einer Melodie zum Stimmumfang einer Versuchsperson. Es ist durchaus denkbar, dass Melodien, die vollständig innerhalb des

Stimmumfangs liegen, häufiger nachgesungen werden (auch unwillkürlich und nebenbei) und somit über die Hörhäufigkeit die Tonart-Repräsentation beeinflusst wird. Da der Stimmumfang der Vpn in den hier vorgestellten Experimenten ebenso wenig gemessen wurde wie die individuellen Beurteilungen der Melodie-Merkmale, kann die Analyse nur die allgemeineren, „gröberen“ Einschätzungen der externen Beurteiler berücksichtigen.

Die Ergebnisse zeigen einen stabilen Effekt der Eingängigkeit von Melodien, der sowohl in der einfaktoriellen als auch in der mehrfaktoriellen Analyse auftrat. Jedoch ist er nicht linear, sondern u-förmig. Wie vorhergesagt, werden Melodien mit mittlerer bis hoher Eingängigkeit genauer produziert als Melodien mit eher geringer oder extremer Eingängigkeit. Dies ist hypothesenkonform und bestätigt den Befund von Winkelhaus (2004), die eine prägnante Melodie als mittelmäßig eingängig definierte. Dass kein Effekt der Ausdrucksstärke gefunden werden konnte, könnte mit der Beurteilung „im Nachhinein“ erklärt werden. Zwar waren sich die drei Beurteiler relativ einig mit der Beurteilung der Ausdrucksstärke, jedoch könnte gerade dieses Merkmal stark von individuellen Hörgewohnheiten abhängen. So könnte z.B. manche Orchesterstelle oder Chor-Mittelstimme hier als wenig ausdrucksstark beurteilt worden sein, für den Spieler oder Sänger jedoch eine andere Geschichte und dadurch auch einen anderen Ausdruck haben. Ein Effekt des Tonumfangs konnte nur bedingt nachgewiesen werden, denn er zeigte sich nicht in der Einzelanalyse, sondern nur in der Gesamtanalyse. Dies könnte entweder daran liegen, dass der Effekt des Tonumfangs mit den Expertisevariablen oder mit Merkmalen des Lernens konfundiert ist und sich erst nach der Kontrolle dieser Variablen zeigen konnte, oder aber an der größeren Fallzahl aufgrund der Melodiewiederholungen in der Gesamtanalyse. Außerdem lässt sich ein Effekt des Tonumfangs vermutlich eindeutiger zeigen, wenn er in Relation zum individuellen Stimmumfang analysiert wird. Aus den hier vorgelegten Daten wird in jedem Fall deutlich, dass es keine einfache (lineare) Beziehung zwischen Tonumfang und Genauigkeit der Tonart-Produktion gibt. Diese zu untersuchen erfordert ebenso wie die beiden anderen und weitere potenziell einflussreiche Melodie-Merkmale Experimente mit der individuellen Erfassung melodischer Charakteristika.

Ein weiteres wichtiges Merkmal, das hier nicht erhoben wurde, aber einen Einfluss auf das Melodie-Gedächtnis haben könnte, ist die Häufigkeit, mit der Melodien in der Originaltonart vs. in anderen Tonarten gehört werden. Besonders bei Melodien mit „Ohrwurm-Potenzial“ ist es denkbar, dass sie eine Person nicht immer in der Originaltonart verfolgen, sondern in verschiedenen Tonarten. Das Hören einer Melodie in wechselnden Tonarten führt vermutlich zu einer weniger stabilen Tonart-Repräsentation als das Hören der Melodie in immer der gleichen Tonart. Von daher wäre eine Kontrolle dieser Variable in weiteren Untersuchungen wünschenswert, ist allerdings praktisch schwer zu verwirklichen. Notwendig wäre dazu die Mitarbeit der Vpn, die sich selbst hinsichtlich potenzieller Ohrwürmer der im Experiment zu produzierenden Melodien beobachten und dabei auf die Tonarten achten müssten. Außerdem ist eine Analyse des musikalischen Materials hinsichtlich des Vorhandenseins in mehreren Tonarten notwendig. Dies betrifft unter den hier verwendeten Melodien vor allem die Handy-Melodien, die zu großen Teilen aus größeren Werken der U- und E-Musik entnommen sind.

7 Zur Vertrautheit beim intentionalen Melodielernen

Der Vergleich aller Ergebnisse aus den Produktionsaufgaben hat gezeigt, dass die individuelle Vertrautheit mit einer Melodie einflussreich für die Tonarterinnerung ist. Melodien, die über längere Zeit regelmäßig gehört, gesungen oder gespielt wurden, wurden häufiger in der Originaltonart produziert als seltener gehörte bzw. erst seit kürzerer Zeit bekannte Melodien. Da sich außerdem gezeigt hat, dass die Art der Melodie (z.B. ihre Eingängigkeit) einen Einfluss auf die Tonarterinnerung hat, erschien eine gezielte Variation der Vertrautheit innerhalb eines Experiments, also bezogen auf die gleiche Melodie, für eine weitere Stützung des Häufigkeitseffekts angezeigt. Dies war das erste Ziel des in diesem Kapitel beschriebenen Experiments. Ein weiteres Ziel war der Leistungsvergleich bei Produktions- und Wiedererkennungsaufgaben, um zu prüfen, ob sich ein Häufigkeitseffekt auch bei der Wiedererkennung von Melodien zeigt. Folgende spezifische Hypothesen werden also in diesem Experiment überprüft:

1. Bei der Produktion der gleichen Melodie wird die Tonart genauer produziert, wenn die Melodie während der Lernphase häufiger gesungen wurde, als wenn sie seltener gesungen wurde.

2. Die Wiedererkennung transponierter und originaler Versionen einer gelernten Melodie ist genauer, wenn die Melodie während der Lernphase häufiger gesungen wurde, als wenn sie seltener gesungen wurde.

3. Der Vertrautheitseffekt tritt nicht nur innerhalb, sondern auch zwischen Melodien auf: Bei einer schwierigen Melodie, die intensiver geübt werden muss, ist sowohl die Produktion als auch die Wiedererkennung der Originaltonart besser als bei einer extrem eingängigen Melodie, die wenig geübt werden muss.

7.1 Methode

7.1.1 Charakterisierung der Stichprobe

Untersucht wurden Schüler des Droste-Hülshoff-Gymnasiums in Berlin Zehlendorf. Es wurden zwei Gruppen von Schülern einbezogen, nämlich 32 Schüler aus einer „musikbetonten“ (d.h. eine zusätzliche Musikstunde pro Woche) 9. Klasse (13 männlich, 19 weiblich, Altersdurchschnitt 14,7 (SD = 0,8) Jahre) sowie 45 Schüler aus dem Schulchor (11 männlich, 34 weiblich, Altersdurchschnitt 16,8 (SD = 1,9) Jahre). Der Schulchor wird von der gleichen Person geleitet, die auch die Musiklehrerin der 9. Klasse ist. Dies ist wichtig, da die Musiklehrerin einen entscheidenden Anteil an der Aufgabe hatte, die Schüler zur Teilnahme am Experiment zu motivieren. Die Teilnahme am Experiment erfolgte freiwillig, mit Zustimmung der Eltern und ohne Vergütung.

7.1.2 Gelernte Melodien

Es wurden zwei Melodien eingesetzt, nämlich ein sehr eingängiger Kanon, „By the waters of Babylon“ (Melodie 1) sowie ein Kanon aus dem 14. Jahrhundert, „Andray soulet“ (Melodie 2). Die beiden Melodien wurden in Kapitel 6.3.2 genauer beschrieben und hinsichtlich ihrer Schwierigkeit verglichen. Festzuhalten ist, dass Melodie 1 eine den Schülern vertraute Struktur aufweist, während Melodie 2 für die Schüler sehr ungewöhnlich und gewöhnungsbedürftig war. Durch diesen Unterschied in der musikalischen Struktur sollte ein Unterschied in der Lernintensität bewirkt werden, durch den Melodie 2 eine größere Aufmerksamkeit entgegengebracht wird als Melodie 1.

Im Experiment wurde noch eine dritte Melodie erfragt, nämlich der Werbejingle der Telekom. Diese Melodie diente als Vergleichsmelodie, da sie nicht intentional gelernt wurde wie die beiden anderen Melodien, sondern inzidentell durch TV- und Radiokonsum. Dadurch war die Lernphase für den Jingle wesentlich länger als für Melodie 1 und 2, so dass er als stark überlernt angesehen werden kann. Die Schüler, die angaben, den Jingle nicht zu kennen, wurden von der Produktion des Jingles befreit.

7.1.3 Lernphase

Die erste Phase des Experiments war die Lernphase, die 3 Monate dauerte und von September bis Dezember 2002 durchgeführt wurde. In diesem Zeitraum besuchte die Versuchsleiterin wöchentlich die 9. Klasse sowie alle 2-3 Wochen den Schulchor und führte eine Lernsitzung durch. Die Lernsitzungen dauerten jeweils 10-15 Minuten. Die Schüler erhielten beide Melodien als Notentext. Begonnen wurden die Lernsitzungen in beiden Gruppen abwechselnd mit Melodie 1 und Melodie 2. Zu Beginn der Lernphase wurde den Schülern jede Melodie zunächst zeilenweise am Klavier vorgespielt, woraufhin sie jede Zeile so oft nachsangen, bis sie in der Gruppe korrekt gesungen werden konnte. Dabei wurde besonders auf die Einhaltung der Tonart geachtet. Es zeigte sich nämlich bei Melodie 1 eine Tendenz, in der Tonart zu sinken, bei Melodie 2 dagegen eine Tendenz, in der Tonart zu steigen. Dies ist vermutlich durch die Melodieführung zu erklären, die in Melodie 1 überwiegend abwärts geht, in Melodie 2 dagegen eher aufwärts. Da Melodie 1 einem Teil der Schüler bereits vor der Lernphase bekannt und insgesamt sehr eingängig war, konnte sie schon in der ersten Lernsitzung dreistimmig im Kanon gesungen werden. Melodie 2 dagegen konnte erst in der vierten Lernsitzung sicher nach den Noten gesungen werden, so dass sie auch erst ab der vierten Lernsitzung als Kanon geübt wurde. Da es für den Schulchor insgesamt nur 4 Lernsitzungen gab, war das Lernergebnis dort das sichere einstimmige Singen von Melodie 2 nach Noten, sowie das weniger sichere Singen von Melodie 2 im zweistimmigen Kanon. Melodie 1 dagegen konnte vom Schulchor auswendig im dreistimmigen Kanon gesungen werden. In der 9. Klasse konnte Melodie 1 ebenfalls auswendig im dreistimmigen Kanon gesungen werden. Außerdem konnte dort Melodie 2 auswendig einstimmig gesungen werden sowie nach Noten im dreistimmigen Kanon.

7.1.4 Abrufphase: Experiment

Das Experiment, das die Tonarterinnerung überprüfen sollte, fand im Januar und Februar 2003 in den Räumen des Droste-Hülshoff-Gymnasiums statt. Die Schüler wurden gebeten, nach Ende des Unterrichts nochmals zur Schule zu kommen und wurden dann im Einzelversuch getestet. Das Experiment dauerte ca. 30 Minuten und umfasste eine Produktionsaufgabe, eine Wiedererkennungsaufgabe, die Beantwortung einiger Fragen zum musikalischen Hintergrund sowie einen Screening-Test auf absolutes Gehör.

In der *Produktionsaufgabe* wurden die Vpn (wie in Kap. 6.1 beschrieben) angewiesen, sich die beiden gelernten Melodien so genau wie möglich vor ihr inneres Ohr zu rufen. Sobald sie eine Vorstellung der jeweiligen Melodie generiert hatten, sollten sie die vorgestellte Melodie singen. Außerdem wurden die Vpn gebeten, sich den Werbejingle der Telekom auf die gleiche Art vorzustellen wie die beiden Versuchsmelodien und dann ebenso zu produzieren. Besonderer Wert wurde in der Instruktion darauf gelegt, dass es beim Singen nicht auf den Stimmklang ankäme, sondern auf die Genauigkeit, mit der die vorgestellten Töne gesungen werden. Um Reihenfolgeeffekte in den Ergebnissen auszuschließen, wurde die Produktionsreihenfolge der drei Melodien vollständig permutiert. Die produzierten Melodien wurden aufgenommen und wie in Kap. 6.1 beschrieben analysiert. Die Produktionsaufgabe fand bei allen Vpn zu Beginn des Experiments statt, um zu verhindern, dass die verschiedenen Melodieversionen aus der Wiedererkennungsaufgabe einen Einfluss auf die Produktion haben.

In der *Wiedererkennungsaufgabe* wurden den Vpn jeweils 5 verschiedene Versionen der beiden gelernten Melodien sowie des Telekom-Werbejingles vorgespielt: die Originalversion, jeweils eine um einen Halbton nach unten und nach oben transponierte Version sowie jeweils eine um eine große Terz (4 Halbtöne) nach unten und nach oben transponierte Version. Die Auswahl dieser Transpositionsgrößen erfolgte in Anlehnung an Terhardt & Ward (1982) sowie Terhardt & Seewann (1983), die mit der gleichen Auswahl an Transpositionen arbeiteten und zusätzlich noch größere Transpositionen um 6 und 7 Halbtöne nutzten. Bei letzteren hatte sich ein Deckeneffekt sowie eine sehr sichere Beurteilung der Terztranspositionen gezeigt, daher wurde hier auf die Nutzung noch größerer Transpositionen verzichtet.

Die jeweils 5 Versionen der 3 Melodien wurden in fester Reihenfolge dargeboten. Um Reihenfolgeeffekte zu minimieren, wurden drei Reihenfolgen hergestellt und die Vpn gleichmäßig auf diese drei Reihenfolgen aufgeteilt. Bei der Reihenfolgeerstellung der wiederzuerkennenden Melodien wurde darauf geachtet, dass niemals zwei Versionen der gleichen Melodie direkt aufeinander folgten. Dies ist wichtig, da die Vpn keine relativen Urteile (Vergleiche von Versionen) fällen, sondern jede Version mit ihrer eigenen Vorstellung vergleichen sollten. Außerdem wurde darauf geachtet, dass keine extremen „Lagenwechsel“ vorkamen, also z.B. auf eine „+4“ Transposition keine „-4“ Transposition folgte.

Bevor den Vpn eine Melodieversion vorgespielt wurde, bekamen sie einen visuellen Hinweisreiz (Liedtitel) vorgelegt und wurden instruiert, sich die jeweilige Melodie so genau wie möglich vor ihr inneres Ohr zu rufen (analog der Produktionsaufgabe). Dafür hatten sie 5 Sekunden Zeit. Nachdem dann die jeweilige Melodieversion erklungen war, sollten die Vpn auf

einem Antwortblatt ankreuzen, ob es sich ihrer Meinung nach um die Originalversion der entsprechenden Melodie handele oder ob die Version zu hoch oder zu tief war. Für ihr Urteil hatten die Vpn wieder 5 Sekunden Zeit. Dieses enge Zeitschema wurde in einem Vorversuch an einigen Vpn getestet, um sicherzugehen, dass die Zeit zum Urteilen ausreicht. Insgesamt kamen die Vpn mit dem Zeitschema gut zurecht. Signalisierte eine Vp, dass sie mehr Zeit für ihr Urteil brauche, so wurde das Band gestoppt, bis die Vp bereit für das nächste Urteil war. Dies kam aber nur bei wenigen Vpn und nur während der ersten 1-2 Trials der Aufgabe vor.

Um den Vpn das Vergessen einer soeben gehörten Version und das Einstellen auf die nächste Melodie zu erleichtern, wurden zwischen allen Trials Störklänge dargeboten. Diese Störklänge waren 5 Sekunden lang und bestanden aus lauten, zufällig durcheinander gespielten Klaviertönen. Es wurden insgesamt 4 Versionen von Störklängen genutzt, damit sich die Vpn nicht an ein tonales Zentrum aus dem Störklang gewöhnen und dieses als Referenz für die Melodiebeurteilung nutzen konnten. Allerdings ist diese Strategie nicht sehr wahrscheinlich, da die Störklänge kein tonales Zentrum haben, sondern etwa so klingen, als habe der Pianist seine Arme längs auf die Tasten gelegt.

Der Ablauf eines Trials der Wiedererkennungsaufgabe verlief also folgendermaßen:

5 s	Hinweisreiz + Vorstellungszeit
10 s	Darbietung einer Melodieversion (Telekom: 2 s)
5 s	Reaktion (Ankreuzen auf Antwortblatt)
5 s	Störklänge
5 s	Pause (darauf folgt der nächste Hinweisreiz)

Die Durchführung und Ausführung des Screening Tests auf absolutes Gehör entsprach dem in Kap. 6.10.1 beschriebenen Vorgehen.

7.1.5 Herstellung des Versuchsmaterials

Die Originalversionen der beiden gelernten Melodien wurde von einem Tonmeister aufgenommen und mithilfe des Programms Time Factory (Prosoniq) transponiert. Die Originalversion des Telekom-Werbejingles war im Internet verfügbar und wurde mit dem gleichen Programm transponiert. Bei den gesungenen Melodien zeigte sich, dass die digitalen Transpositionen nach oben wesentlich besser klangen als die Transpositionen nach unten. Bei den letzteren traten durch die Änderung der Formanten Vokalfärbungen auf, die sich von der Originalversion deutlich unterschieden. Auf Anraten des Tonmeisters wurden daher die Transpositionen nach unten erneut aufgenommen, anstatt die Originalversion digital zu transponieren. Streng genommen ist somit nicht von ganz identischen Transpositionen auszugehen, da die Singstimme in verschiedenen Lagen unterschiedlich klingt. Letztendlich ist hier aber ein Expertenurteil gefallen, da der Tonmeister für jede Transposition entschieden hat, ob die erneute Aufnahme oder die digital hergestellte Version dem Original am ähnlich-

sten war. Dieses Problem trat beim instrumentalen Telekom-Werbejingle nicht auf, daher wurden hier alle Transpositionen digital hergestellt.

Aus den Melodieversionen und den Störklängen wurde mithilfe des Programms Cool Edit eine CD erstellt, auf der die Trials nach dem oben beschriebenen Zeitschema aneinander geschnitten waren. Diese CD, die die notwendigen Pausen für die Urteile bereits enthielt, konnte dann den Vpn über Kopfhörer vorgespielt werden, wodurch ein gleicher zeitlicher Ablauf für alle Vpn gesichert wurde. Da jede Melodieversion zweimal dargeboten wurde, enthielt die CD 3 (Melodien) x 5 (Versionen) x 2 (Versuche) = 30 Trials. Die Bearbeitung der Wiedererkennungsaufgabe dauerte knapp 14 Minuten.

7.1.6 Besonderheiten bei der Versuchsdurchführung

In das Experiment sollten ursprünglich noch ca. 30 Schüler einer zweiten Schule (Anna-Freud-Oberschule) einbezogen werden. Während der Lernphase stellte sich jedoch heraus, dass sich das Lernklima und die Interessenlage in dieser Schule so stark von der ersten Schule unterschieden, dass nicht von vergleichbaren Lernbedingungen ausgegangen werden kann. Zum Einen traten in der zweiten Schule sehr viele Fehlzeiten auf, etwa die Hälfte der Schüler fehlte in jeder 2. bis 3. Stunde (es wurde wöchentlich geübt). Zum Anderen waren die Schüler nur sehr zögerlich bereit, die geübten Melodien zu singen, viele hörten statt dessen nur zu. Während der Abrufphase erschienen dann nur 17 von 30 Schülern überhaupt zum Experiment, und viele weigerten sich zu singen. Daher wurden die Daten dieser Schüler größtenteils von der weiteren Auswertung ausgeschlossen. Nur die Produktions- und Wiedererkennungsdaten bezüglich des Telekom-Werbejingles wurden ausgewertet, da dieser außerhalb der Schule gelernt wurde.

7.1.7 Datenanalyse

Die in der Produktionsaufgabe erhaltenen Anfangstöne wurden mit den Original-Anfangstönen der jeweiligen Melodien verglichen und die Differenzen berechnet. Die Verteilungen dieser Differenzen wurden auf Gleichverteilung und Normalverteilung geprüft (vgl. Kap. 6.1). Um die Hypothesen bezüglich der Vertrautheit zu überprüfen, wurde über die Differenzen eine zweifaktorielle Varianzanalyse (ANOVA) mit dem Gruppierungsfaktor *Lernart* (oft vs. selten) und dem Messwiederholungsfaktor *Melodie* (1 vs. 2) berechnet. Zuvor wurde mit dem Box-Test geprüft, ob die Kovarianz-Matrizen zwischen den Faktorstufen gleich sind.

Für die Analyse der Wiedererkennungsaufgabe wurde zunächst pro Vp die relative Anzahl von „zu hoch“, „richtig“ und „zu tief“ Antworten pro Version jeder Melodie bestimmt. Zum Beispiel wurde gezählt, wie oft Vp 1 bei der Transposition „+4“ von Melodie 1 das Urteil „zu hoch“ gegeben hatte. Gab sie es einmal, so ist die relative Anzahl an „zu hoch“ Antworten, da diese Transposition zweimal beurteilt wurde, gleich „0,5“. Die über alle Vpn gemittelten Daten dienten zunächst der graphischen Veranschaulichung der Ergebnisse. Des weiteren wurde in Anlehnung an Terhardt & Seewann (1983) eine „guessing-normalized identification rate“ (IR), d.h. eine um die falsch positiven Antworten korrigierte relative Erkennungshäufig-

keit, berechnet. Terhardt & Seewann (1983, S. 70) berechnen die IR, indem von der relativen Anzahl an „richtig“ Antworten (= Anzahl „richtig“ / Anzahl Originalversionen) die relative Anzahl falsch positiver Antworten (= Anzahl „richtig“ / Anzahl Transpositionen) abgezogen wird. Die IR wird für jede Kombination Vp x Melodie bestimmt. Sie kann maximal den Wert 1 erreichen, wenn von einer Vp alle Originale als richtig beurteilt werden und alle Transpositionen als entweder zu hoch oder zu tief (2/2 – 0/8 = 1). Wird dagegen von einer Vp kein Original als richtig beurteilt, sondern alle Transpositionen, so erreicht sie den Wert -1 (0/2 – 8/8 = -1). Wird von einer Vp immer mit „richtig“ geantwortet, egal welche Version vorgespielt wurde, so ist die IR gleich 0 (2/2 – 8/8 = 0).

Um zunächst zu überprüfen, ob die Daten der Wiedererkennungsaufgabe eine signifikante Tonarterinnerung nahe legen, wurde (in Anlehnung an Terhardt & Seewann, 1983, S. 74) ein Histogramm der IRs erstellt, indem für jedes 10 % Intervall der IRs (0-10 %, 10-20 %, etc.) die relative Häufigkeit bestimmt wurde. Terhardt & Seewann schließen aus ihren Daten: „it would thus seem to be safe to regard general IR’s of less than about 10 % and semitone IR’s of less than about 15 % as insignificant, i.e., not indicative of key-identification ability” (S. 74). Mit “semitone IR’s” sind IRs gemeint, die in einem Experiment auftreten, in dem die Originalversion nur mit Halbtontranspositionen verglichen wird. Da dies im vorliegenden Experiment nicht der Fall war, wurde die 10 % Marke als Signifikanzkriterium gewählt und berechnet, wie viele IRs über dieser Marke lagen.

Um weiterhin die Hypothesen bezüglich der Vertrautheit zu überprüfen, wurden die über alle Vpn gemittelten IRs einer zweifaktoriellen ANOVA mit dem Gruppierungsfaktor *Lernart* (oft vs. selten) und dem Messwiederholungsfaktor *Melodie* (1 vs. 2) zugeführt. Zuvor wurden die abhängige Variable mit dem Kolmogorov-Smirnov-Test auf Normalverteilung geprüft und der Box-Test durchgeführt. Während zunächst die Wiedererkennungsdaten für beide Melodien und den Telekom-Werbejingle dargestellt und auf Signifikanz bezüglich der Tonarterinnerung überprüft wurden, gingen in die hypothesenbezogene Auswertung nur die beiden gelernten Melodien ein, weil der Faktor *Lernart* bezüglich des Telekom-Jingles nicht gestuft wurde.

7.2 Ergebnisse

7.2.1 Produktionsaufgabe

Das Ergebnis der Verteilungsanalysen für die Differenzen wurde in der Übersicht über alle Produktionsaufgaben beschrieben (vgl. Kap. 6.3.4). Hier soll es daher nur um die hypothesenbezogene Auswertung der Produktionsdaten gehen. In Tabelle 7.1 sind die über alle Vpn gemittelten Differenzen getrennt nach den beiden Faktoren *Lernart* und *Melodie* aufgeführt. Dargestellt sind sowohl die mittleren absoluten Differenzen als auch die Mittelpunkte der Fehlerverteilungen mit Berücksichtigung der Fehlerrichtung.

Die zweifaktorielle ANOVA mit den Faktoren *Lernart* (oft vs. selten) und *Melodie* (1 vs. 2) wurde sowohl für die mittleren absoluten Differenzen als auch für die Mittelpunkte der Fehlerverteilungen durchgeführt. Der Box-Test ergab für beide abhängigen Variablen kein signi-

fikantes Ergebnis, d.h. die Varianzen zwischen den Gruppen sind homogen (Mittelpunkte der Fehlerverteilungen: F (3,459532) = 1,876, p = .131; mittlere absolute Differenzen: F (3,459532) = 2,422, p = .064). Für die mittleren absoluten Differenzen ergab die ANOVA einen signifikanten Haupteffekt für *Melodie* (F(1,68) = 14,732, p < .001), jedoch nicht für *Lernart* (F(1,68) = 1,224, p = .272), sowie keine Interaktion der Faktoren (F(1,68) = 0,521, p = .473). Für die Mittelpunkte der Fehlerverteilungen ergab sich das gleiche Ergebnismuster, also ein signifikanter Haupteffekt für *Melodie* (F(1,68) = 14,518, p < .001), nicht aber für *Lernart* (F(1,68) = 2,690, p = .106) sowie keine Interaktion der Faktoren (F(1,68) = 0,391, p = .534). Um den Einfluss der Lernhäufigkeit für die beiden Melodien getrennt zu untersuchen, wurden als nächstes einfaktorielle ANOVAs mit dem Gruppierungsfaktor *Lernart* (oft vs. selten) berechnet. Diese ergaben für Melodie 2 einen tendenziellen *Lernart* Effekt für die Mittelpunkte der Fehlerverteilungen (F(1,68) = 3,393, p = .070), jedoch nicht für die mittleren absoluten Differenzen (F(1,68) = 2,110, p = .151). Für Melodie 1 wurde der Faktor *Lernart* weder in den Mittelpunkten der Fehlerverteilungen (F(1,75) = 1,428, p = .236) noch in den mittleren absoluten Differenzen signifikant (F(1,75) = 0,557, p = .458).

Tab. 7.1: Mittlere produzierte Fehler in den vier Versuchsbedingungen

Lernart	oft		selten	
Melodie	Melodie 1	Melodie 2	Melodie 1	Melodie 2
	mittlere absolute Differenzen			
M (HT)	3,72	2,13	4,16	2,88
SE (HT)	0,49	0,29	0,35	0,38
	Mittelpunkte der Fehlerverteilungen			
M (HT)	- 3,41	- 1,73	- 4,16	- 2,78
SE (HT)	0,56	0,37	0,35	0,40

Mittlere absolute Differenzen und Mittelpunkte der Fehlerverteilungen in Halbtönen (HT), getrennt nach den Faktoren *Lernart* und *Melodie*. M = Mittelwert, SE = Standardfehler.

Demnach kann Hypothese 1 aufgrund der vorliegenden Daten nicht bestätigt werden. Obgleich die Daten in Tabelle 7.1 in die vorhergesagte Richtung weisen, finden sich keine Gruppenunterschiede, die groß genug sind, um statistisch bedeutsam zu sein. Lediglich bei Melodie 2 lässt sich ein tendenziell signifikanter Leistungsvorteil in der oft-Gruppe erkennen. Hypothese 3 dagegen kann für die Produktionsaufgabe als bestätigt gelten, da sich zwischen den beiden Melodien signifikante Unterschiede in der Produktionsgenauigkeit gezeigt haben: Die intensiver geübte Melodie 2 wird signifikant besser produziert als die weniger intensiv geübte Melodie 1.

7.2.2 Wiedererkennungsaufgabe

Abbildung 7.1 stellt zunächst die Häufigkeiten, mit denen die Vpn bei den je fünf Versionen der beiden gelernten Melodien sowie des Telekom-Jingles die drei möglichen Antworten („zu hoch“, „richtig“, „zu tief“) gegeben haben, gegenüber.

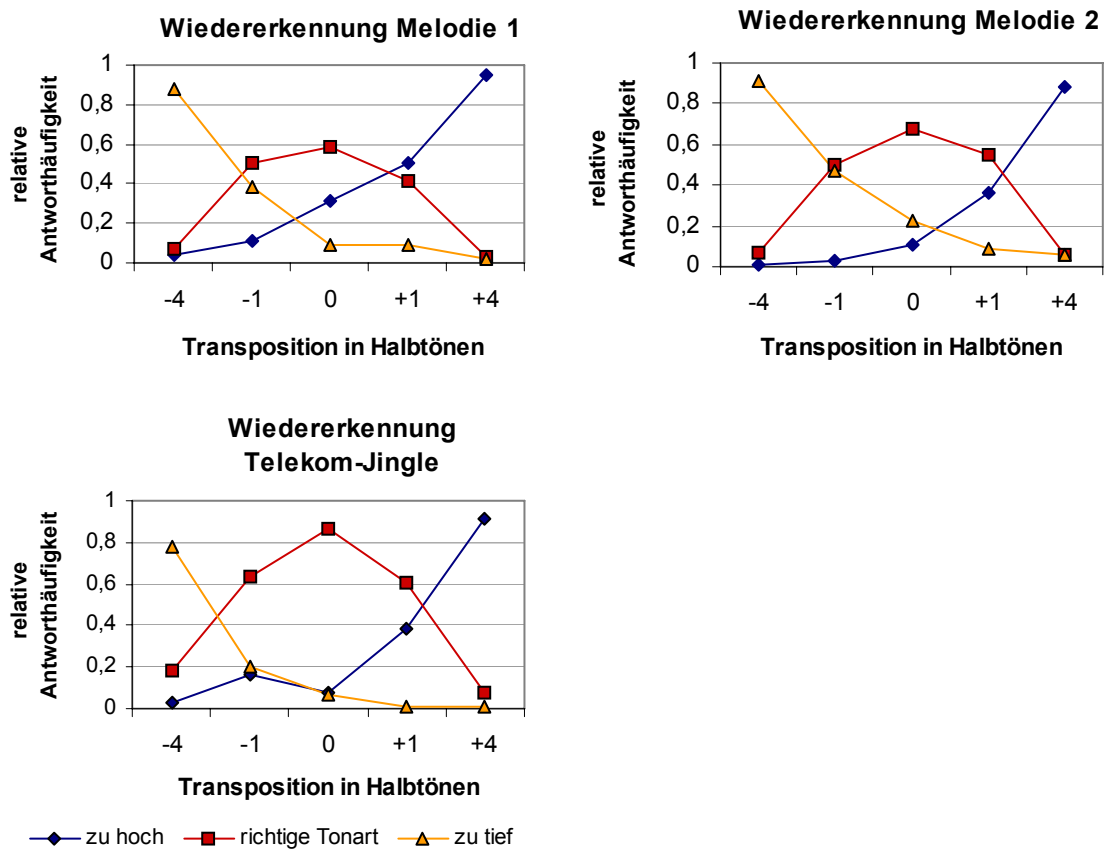

Abb. 7.1: Wiedererkennung der drei Melodien. Auf der Kategorienachse sind die 5 Versionen jeder Melodie abgetragen. Die y-Achse zeigt die relative Häufigkeit der drei möglichen Antworten „zu hoch“, „richtig“ und „zu tief“.

„Per Augenmaß“ kann man sehen, dass die beiden gelernten Melodien etwa gleich gut wiedererkannt wurden. Die jeweils richtige Version (Transposition = 0) wurde in etwa 60 % der Fälle erkannt, während die großen Transpositionen (+ / - 4 Halbtöne) in über 80 % der Fälle erkannt wurden. Schwierigkeiten gab es offensichtlich bei der Beurteilung der Halbtontranspositionen, die bei beiden Melodien häufiger mit „richtig“ als mit „zu hoch“ oder „zu tief“ bewertet wurden. Bei der Wiedererkennung des Telekom-Jingles zeigte sich mit über 80 % eine noch bessere Wiedererkennung der Originalversion sowie eine vergleichbare Wiedererkennung der großen Transpositionen, während die Halbtontranspositionen ähnlich stark wie bei den beiden Melodien mit der Originalversion verwechselt wurden.

Um die Wiedererkennungsleistung auf Signifikanz zu überprüfen, wurde wie oben beschrieben für jede Vp und jede Melodie die IR berechnet. Abb. 7.2 zeigt das Histogramm der über alle Vpn gemittelten IRs, getrennt nach den drei Melodien. Es zeigt sich, dass für Melodie 1 knapp 25 % der Fälle in die Kategorie 0-10 % fallen (wobei diese Kategorie, in Anlehnung an Terhardt & Seewann, 1983, auch die negativen IRs enthält) und somit als zufällige Leistung zu bewerten sind. Für Melodie 2 ist dieser Wert mit 22 % vergleichbar hoch, während er für den Telekom-Jingle nur 8,5 % beträgt. Demzufolge liegt die Rate signifikanter Wiedererkennungsleistungen bei Melodie 1 bei 75 % der Vpn, bei Melodie 2 bei 78 % der Vpn, und beim Telekom-Jingle bei 92,5% der Vpn.

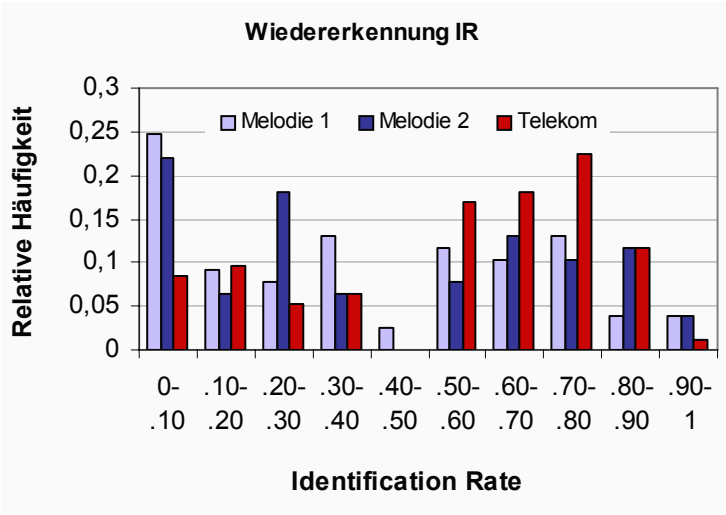

Abb. 7.2: Über alle Vpn gemittelte IRs (relative Erkennungshäufigkeit, korrigiert um falsch positive Urteile), getrennt nach den drei Melodien.

Terhardt & Seewann (1983) nehmen außer der Erstellung von IR-Histogrammen keine weitere Signifikanzprüfung vor. In dieser Arbeit wird noch ein weiteres Signifikanzkriterium für die absolute Tonart-Erinnerung vorgeschlagen, das sich mehr an den Antwortverteilungen selbst orientiert. Während man an den Verteilungen (Abb. 7.1) „per Augenmaß“ sehen kann, dass die Terztranspositionen eindeutig als Transpositionen erkannt wurden (die jeweils korrekte Antwort wird fast überall in über 80 % der Fälle gegeben, während die Häufigkeit der falschen Antworten sich nahe 0 bewegt), ist das Antwortmuster bei den Halbtontranspositionen nicht so eindeutig als überzufällig zu erkennen. Da die IR-Histogramme auf einer Berechnung über alle Transpositionen beruhen, soll hier noch ein direkter Vergleich zwischen den Antworten bei Originalversionen vs. Halbtontranspositionen erfolgen. Dies geschieht über einen Vergleich der korrekten Urteile und den falsch positiven Urteilen, also über einen t-Test mit Messwiederholungen, der die relative Häufigkeit an „richtig“-Urteilen bei Originalversionen und bei Halbtontranspositionen vergleicht. Dieser wird getrennt für die drei Melodien sowie getrennt für die beiden Vpn-Gruppen durchgeführt. In Tab. 7.2 sind die relativen Häufigkeiten an „richtig“-Urteilen, getrennt nach Melodien und Gruppen, dargestellt. Die Darstellung für den Telekom-Jingle ist nicht nach Gruppen getrennt, da hier die Lernhäufigkeit nicht variiert wurde. Als Voraussetzung für den t-Test mit Messwiederholungen wurde geprüft, ob die jeweils verglichenen Testreihen positiv korreliert sind. Bei negativen Korrelationen wurde ein nichtparametrisches Verfahren (Wilcoxon-Test) angewandt. Aufgrund der großen Stichprobe wurde auf eine Prüfung der Normalverteilung verzichtet.

Der t-Test für verbundene Stichproben ergab für Melodie 1 ein signifikantes Ergebnis in der oft-Gruppe (r = 0,276, p = .126; t (31) = 2,533, p = .017), nicht aber in der selten-Gruppe (r = 0,080, p = .601; t (44) = 1,484, p = .145). Für Melodie 2 wurde der Wilcoxon-Test weder für die oft-Gruppe (r = - 0,320, p = .074; Z = - 1,355, p = .175) noch für die selten-Gruppe signifikant (r = - 0,015, p = .920; Z = - 1,303, p = .193). Beim Telekom-Werbejingle wurde der t-Test über beide Gruppen hinweg signifikant (r = 0,076, p = .465; t (93) = 7,006, p < .001).

Demnach werden bei Melodie 1 nur in der Gruppe, die die Melodien intensiver gelernt hat, signifikant häufiger die Originale als „richtig“ bewertet als die Halbtontranspositionen. Da bei Melodie 2 die Unterschiede in der oft-Gruppe sogar noch etwas größer sind als bei Melodie 1, kann das nicht signifikante Ergebnis auf den schwächeren Wilcoxon-Test zurückgeführt und insgesamt festgestellt werden, dass sich die Lernintensität deutlich auf die Erkennung von Halbtontranspositionen ausgewirkt hat. Beim Telekom-Werbejingle liegt in beiden Gruppen eine gute Erkennung von Halbtontranspositionen vor.

Tab. 7.2: Relative Häufigkeit an „richtig“-Antworten bei Originalversionen und Halbtontranspositionen, getrennt nach Melodien und Gruppenzugehörigkeit

Lernart	oft		selten	
Version	Original	+ / - 1 Halbton	Original	+ / - 1 Halbton
	Relative Häufigkeit an „richtig“-Antworten			
	Melodie 1			
M	0,656	0,484	0,544	0,441
SE	0,065	0,044	0,057	0,046
	Melodie 2			
M	0,719	0,469	0,644	0,567
SE	0,059	0,042	0,047	0,041
Lernart	Original		+ / - 1 Halbton	
	Relative Häufigkeit an „richtig“-Antworten			
	Telekom			
M	0,862		0,620	
SE	0,026		0,025	

M = Mittelwert, SE = Standardfehler.

Für die hypothesenbezogene Signifikanzprüfung wurden die IRs der zweifaktoriellen ANOVA mit den Faktoren *Lernart* (oft vs. selten) und *Melodie* (1 vs. 2) zugeführt. Tabelle 7.3 zeigt zunächst die mittleren IRs, getrennt nach den beiden Faktoren.

Tab. 7.3: Mittlere IRs in den vier Versuchsbedingungen

Lernart	oft		selten	
Melodie	Melodie 1	Melodie 2	Melodie 1	Melodie 2
	mittlere IRs			
M	0,399	0,461	0,313	0,328
SE	0,065	0,069	0,059	0,055

M = Mittelwert, SE = Standardfehler.

Die IRs von Melodie 1 genügen einer Normalverteilung (Z = 1,051, p = .219), ebenso die IRs von Melodie 2 (Z = 1,110, p = .170). Der Box-Test wurde nicht signifikant (F(3,341786) = 0,177, p = .912). Die zweifaktorielle ANOVA ergab einen tendenziellen *Lernart* Effekt

(F(1,75) = 3,114, p = .082), aber keinen Effekt der *Melodie* (F(1,75) = 0,381, p = .539) sowie keine Interaktion der Faktoren (F(1,75) = 0,145, p = .704). Um den Einfluss der Lernhäufigkeit für die beiden Melodien getrennt zu untersuchen, wurden als nächstes einfaktorielle ANOVAs mit dem Gruppierungsfaktor *Lernart* (oft vs. selten) berechnet. Diese ergaben für Melodie 1 keinen Effekt der *Lernart* (F(1,75) = 0,930, p = .338), und auch nicht für Melodie 2 (F(1,75) = 2,315, p = .132).

Demnach kann Hypothese 2 aufgrund der vorliegenden Daten nur eingeschränkt bestätigt werden. Wie in der Produktionsaufgabe weisen auch hier die Daten in Tabelle 7.4 in die vorhergesagte Richtung, jedoch sind die Unterschiede zwischen den beiden Gruppen nur auf dem 10 % Niveau signifikant. Eine indirekte Bestätigung erhält die Hypothese jedoch durch den Vergleich der „richtig"-Antwort-Häufigkeit bei Originalversionen vs. Halbtontranspositionen, der nur in der oft-Gruppe signifikant wurde. Hypothese 3 kann für die Wiedererkennungsaufgabe nicht bestätigt werden, da die Unterschiede zwischen den beiden Melodien viel kleiner sind als in der Produktionsaufgabe und somit nicht signifikant werden.

7.2.3 Tonarterinnerung, Tonbenennung & musikalische Expertise

Im Test auf absolutes Gehör wurden durchschnittlich 0,97 (SD = 1,29) Töne richtig benannt. Bei 2,57 (SD = 2,04) Tönen lagen die Vpn durchschnittlich um maximal einen Halbton daneben. Damit ist die Tonbenennungsleistung in dieser Stichprobe nur als zufällige Leistung zu betrachten, denn die Anzahl korrekt benannter Töne liegt im Mittel etwas unter dem Zufallswert 1 (= 12 x 1/12). Wie Abb. 7.3 zeigt, gab es zwar einige Vpn, die einige Töne richtig benennen konnten, jedoch kann man wohl nur bei der Vp, die 8 Töne richtig benannte, davon ausgehen, dass sie über ein absolutes Gehör verfügt (in Übereinstimmung mit ihrer Selbsteinschätzung).

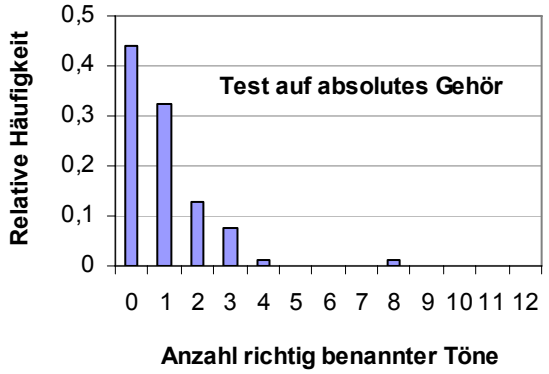

Abb. 7.3: Mittlere Anzahl korrekt benannter Töne über 77 Vpn.

Um einen Einfluss der Fähigkeit, Töne zu benennen, auf die Wiedererkennung von Tonarten zu überprüfen, wurden die IRs mit der Tonbenennungsleistung korreliert (die Korrelation mit der Produktionsaufgabe erfolgte im Gesamtvergleich, vgl. Kap. 6.11). Die Ergebnisse sind in Tabelle 7.4 dargestellt. Es wird deutlich, dass die Leistung in der Tonbenennung nicht signifikant mit der in der Wiedererkennungsaufgabe korreliert.

Tab. 7.4: Korrelationen zwischen den Leistungen in der Wiedererkennungs- und Tonbenennungsaufgabe

		Trefferquote absolutes Gehör	Trefferquote absolutes Gehör mit Halbtontoleranz
IR Melodie 1	Pearson Correlation	,080	-,002
	Sig. (2-tailed)	,491	,986
	N	77	77
IR Melodie 2	Pearson Correlation	-,067	,017
	Sig. (2-tailed)	,565	,882
	N	77	77
IR Telekom	Pearson Correlation	-,015	-,077
	Sig. (2-tailed)	,883	,459
	N	94	94

Um weiterhin zu überprüfen, ob sich musikalische Expertise, hier gemessen durch die Dauer der Instrumentalausbildung und Chorerfahrung, auf die Tonarterinnerung auswirkt, wurden zwischen diesen Variablen weitere Korrelationen berechnet, die in Tab. 7.5 dargestellt sind.

Tab. 7.5: Korrelationen zwischen den Leistungen in der Produktions-, Wiedererkennungs- und Tonbenennungsaufgabe und der musikalischen Ausbildung

		Instrumentalausbildung in Jahren	Beginn des Instrumentalunterrichts	Chorerfahrung
Produktion Melodie 1	Pearson Correlation	**-,322**	,103	-,072
	Sig. (2-tailed)	,004	,394	,535
	N	77	71	77
IR Melodie 1	Pearson Correlation	-,018	-,067	,136
	Sig. (2-tailed)	,876	,579	,240
	N	77	71	77
Produktion Melodie 2	Pearson Correlation	-,125	,082	-,044
	Sig. (2-tailed)	,301	,522	,717
	N	70	64	70
IR Melodie 2	Pearson Correlation	,179	***-,218***	-,045
	Sig. (2-tailed)	,119	,068	,695
	N	77	71	77
Produktion Telekom	Pearson Correlation	-,088	-,018	,108
	Sig. (2-tailed)	,463	,887	,370
	N	71	64	71
IR Telekom	Pearson Correlation	-,035	-,081	,053
	Sig. (2-tailed)	,738	,466	,609
	N	94	84	94
Absolutes Gehör Trefferquote	Pearson Correlation	***,201***	,014	**,307**
	Sig. (2-tailed)	,052	,900	,003
	N	94	84	94
Absolutes Gehör Trefferquote mit Halbtontoleranz	Pearson Correlation	**,318**	-,088	**,225**
	Sig. (2-tailed)	,002	,426	,029
	N	94	84	94

Signifikante Korrelationen auf dem Niveau α =5 % sind fett gedruckt, Tendenzen (α = 10 %) sind kursiv.

Es zeigt sich, dass die Zusammenhänge, wenn überhaupt vorhanden, sehr klein sind. Bezüglich Melodie 1 wird nur die Korrelation zwischen Produktion und der Dauer der Instrumentalerfahrung signifikant, während bezüglich Melodie 2 die Korrelation zwischen Wiedererkennung und Beginn des Instrumentalunterrichts tendenziell signifikant wird. Bei dem Telekom-Jingle gibt es weder in der Produktions-, noch in der Wiedererkennungsaufgabe signifikante Zusammenhänge mit der musikalischen Ausbildung.

7.3 Diskussion der Ergebnisse

Das wichtigste Ziel des beschriebenen Experiments war nachzuweisen, dass die Vertrautheit mit musikalischem Material die Erinnerungsleistung bezüglich der Tonart beeinflusst. Es wurde angenommen, dass die Tonart von sehr vertrauten (überlernten) Melodien besser abrufbar ist als die von weniger vertrauten Melodien. Um dies zu zeigen, wurde die Lernintensität von zwei unterschiedlichen Melodien bei zwei Gruppen von Schülern über ein Zeitintervall von 3 Monaten variiert, so dass eine Gruppe etwa drei mal so viele Lernsitzungen hatte wie die andere Gruppe. Außerdem wurde die Vertrautheit durch die Auswahl der Melodien variiert, denn die eine ist eingängig und daher ohne großen Probenaufwand lernbar, während die andere schwieriger zu lernen und damit probenintensiver war.

Bei der Produktionsaufgabe hat sich gezeigt, dass die Produktionsgenauigkeit von der gesungenen Melodie abhängig ist. Wie in Kap. 6.3.4 beschrieben wurde, konnte Melodie 1 nicht überzufällig gut produziert werden, während man bei Melodie 2 von signifikanter Tonarterinnerung ausgehen kann. Außerdem ergab sich in der zweifaktoriellen ANOVA ein signifikanter Haupteffekt für den Faktor *Melodie*. Da die Melodien den gleichen Tonumfang haben, ist dieser als Ursache für die bessere Produktion von Melodie 2 auszuschließen. Die Ursache liegt vermutlich in der musikalischen Struktur von Melodie 2, die, verglichen mit der Musik, die von dieser Vpn-Population im Allgemeinen gehört wird, sehr ungewöhnlich ist. Sie war zunächst sehr wenig einprägsam, von daher wurde ihr während der Lernphase eine hohe Aufmerksamkeit entgegengebracht und sie wurde intensiver geübt als Melodie 1. Neben diesem Effekt der Lernintensität spielt vermutlich noch ein weiterer Effekt eine Rolle, nämlich die sehr große Eingängigkeit von Melodie 1. Diese ist eine Melodie, die man öfter vor sich hinsummt und die das Potenzial zu einem Ohrwurm hat. Bei Ohrwürmern ist es aber durchaus nicht sicher, dass sie immer in der Originaltonart gesummt oder gesungen werden, sondern möglicherweise von jeder Vp in ihrer bevorzugten Tonlage. Durch das Singen der Melodie in beliebigen Tonarten wird die während des Unterrichts aufgebaute Tonart-Repräsentation der Melodie quasi wieder verlernt.

Des weiteren hat sich in der Produktionsaufgabe ein tendenzieller Gruppenunterschied gezeigt, der aber deutlich kleiner ist als der Unterschied zwischen den beiden Melodien: Während der Leistungsunterschied zwischen den Melodien in beiden Gruppen mehr als einen Halbton betrug, unterschied sich die Produktionsgenauigkeit zwischen den beiden Gruppen um deutlich weniger als einen Halbton. Vergleicht man die Gruppenleistungen für beide Melodien einzeln, so zeigt sich, dass nur bei Melodie 2 ein tendenzieller Gruppenunterschied

vorliegt, nicht aber bei Melodie 1. Dies ist vermutlich dadurch erklärbar, dass Melodie 1 einigen Vpn aus beiden Gruppen schon bekannt war und somit nach kurzer Lernphase in der Gruppe gesungen werden konnte. Das Üben bestand also im Wesentlichen darin, in jeder Lernsitzung die Tonartrepräsentation zu stabilisieren, während Melodie 2 wirklich intensiv geübt wurde. Dies führte schon zum Ende der Lernphase zu einem beobachtbaren Leistungsunterschied zwischen den beiden Gruppen, denn während die oft-Gruppe die Melodie auswendig im Kanon singen konnte, gelang dies der selten-Gruppe nicht. Bei Melodie 1 gab es keine solchen Unterschiede.

Der *Melodie*-Effekt und der tendenzielle *Lernart*-Effekt, die sich in der Produktionsaufgabe gezeigt haben, deuten darauf hin, dass die Vertrautheit mit einer Melodie, die durch Übungshäufigkeit bzw. Lernintensität gekennzeichnet wird, sich auf die Genauigkeit der Tonarterinnerung auswirkt.

Bezüglich des Einflusses von *Melodie* und *Lernart* haben sich die in der Wiedererkennungsanforderung erhaltenen Unterschiede als wesentlich kleiner herausgestellt als die in der Produktionsaufgabe gefundenen. Der *Melodie*-Effekt ließ sich in der Wiedererkennungsaufgabe anhand der gemittelten IRs nicht nachweisen, und auch der *Lernart*-Effekt war sehr klein und dadurch nur tendenziell signifikant. Einen indirekten *Lernart*-Effekt kann man aber durch den Vergleich der korrekten Urteile mit den falsch positiven Urteilen nachweisen, der nur in der oft-Gruppe signifikant wurde. Der Grund für die fehlenden bzw. kleinen Effekte könnte in der Wahl der Transpositions-Größen liegen. Offenbar waren Transpositionen um eine große Terz auch für die Vpn, die die Melodien nicht sehr häufig geübt hatten, eindeutig als Transpositionen erkennbar, so dass es zu einem Deckeneffekt bei der Wiedererkennung kam. Halbtontranspositionen dagegen schienen auch für die oft-Gruppe sehr schwer als solche erkennbar zu sein, und sie wurden oft mit dem Original verwechselt. Das bedeutet, dass auch Transpositionen um 2 und 3 Halbtöne sinnvoll wären, um die Tonarterinnerung besser abbilden zu können. Bei Terhardt & Seewann (1983) zeigte sich eine Art „Erinnerungsgrenze“ für die Tonart bei $\pm$ 1,5 Halbtönen, ab der Transpositionen überzufällig zurückgewiesen werden. Die Wiedererkennungsaufgabe aus dem hier beschriebenen Experiment bestätigt dieses Ergebnis insofern, als die Grenze jenseits der Transposition von $\pm$ 1 Halbton zu liegen scheint.

Betrachtet man die Ergebnisse von Produktions- und Wiedererkennungsaufgabe im Vergleich, so zeigt sich, dass sie unterschiedliche Aspekte der Tonarterinnerung sensibler abbilden können. So konnte die Produktionsaufgabe die genauere Tonarterinnerung von Melodie 2 abbilden, während die Wiedererkennungsaufgabe auch die Richtung der Fehler abbilden konnte. Beim Vergleich der Ergebnisse in Abb. 7.1 fällt nämlich auf, dass sich der Punkt, an dem sich „zu tief“ und „zu hoch“ Antworten kreuzen, zwischen Melodie 1 und Melodie 2 unterscheidet: Bei Melodie 1 liegt er etwas tiefer als die Originalversion, während er bei Melodie 2 etwas höher liegt. Ebenso wurde bei Melodie 1 die Originalversion außer als „richtig“ auch oft als „zu hoch“ bezeichnet, während bei Melodie 2 die Originalversion außer als „richtig“ auch oft als „zu tief“ bezeichnet wurde. In diesem Antwortmuster spiegelt sich die schon

während der Lernphase beobachtete Tendenz zu sinken bei Melodie 1, dagegen die Tendenz zu steigen bei Melodie 2 wider. Interessant ist die Tendenz zu steigen bei Melodie 2, die im Gegensatz zu der Tendenz, in der Produktionsaufgabe zu tief zu singen, steht und ein Hinweis darauf sein könnte, dass in diesem Fall die Wiedererkennungsaufgabe die Tonartrepräsentation besser abbilden konnte als die Produktionsaufgabe.

Der Zusammenhang zwischen Tonarterinnerung (gemessen durch Wiedererkennungsleistungen) und Tonbenennung (gemessen durch den Test auf absolutes Gehör) hat sich als sehr gering herausgestellt. Während sich die in der Literatur zum absoluten Gehör bekannten Befunde zum positiven Zusammenhang zwischen absolutem Gehör und Dauer der musikalischen Ausbildung hier auch für eine einzelne Stichprobe bestätigen ließen, ließen sich nur wenige und insgesamt niedrige Korrelationen zwischen Tonarterinnerung und musikalischer Ausbildung nachweisen. Dies bestätigt die Ergebnisse aus dem Gesamtvergleich und verweist darauf, dass es außer der musikalischen Expertise und der Fähigkeit der Tonidentifikation noch wichtige andere Einflussfaktoren auf die Qualität der Tonarterinnerung geben muss, von denen einer die Vertrautheit mit den gelernten Melodien ist. Zwar sind die nachgewiesenen *Lernart*-Effekte relativ klein, jedoch ist zu bedenken, dass sie während einer insgesamt sehr kurzen Lernphase von nur 3 Monaten erzielt wurden, in der die Variation der Lernintensität nicht sehr extrem ausfallen konnte. Insofern bietet die Variation der Lernintensität einen lohnenden Ansatzpunkt für Folgestudien mit möglichst extremerer Variation, um die vorgestellte Interpretation zu erhärten.

8 Zur Vertrautheit bei der Tonbenennung

Um zu zeigen, dass die berichteten Effekte der Hör-Häufigkeit bzw. Vertrautheit mit musikalischem Material nicht auf die Tonart-Erinnerung beschränkt sind, sondern auch für die Spezialpopulation Absoluthörer bei der Tonbenennung wirksam werden, wurde ein Experiment mit Absoluthörern und Nichtabsoluthörern durchgeführt. Ausgangspunkt für dieses Experiment, das methodisch den Charakter einer Pilot-Untersuchung hat, waren die in der Literatur berichteten Experimente mit Absoluthörern, die zeigen, dass deren Leistung bei der Tonbenennung von Tonklasse, Klangfarbe und Oktavlage der vorgespielten Töne abhängt. Töne weißer Klaviertasten sowie Töne in vertrauten Klangfarben und in mittlerer Lage werden besser und schneller benannt als Töne schwarzer Klaviertasten und Töne in fremden Klangfarben und extremen Lagen (vgl. Kap. 2.2.1).

Der schwarz-weiß Effekt wurde bisher mit zwei theoretischen Ansätzen erklärt. Miyazaki (1990) und Heyde (1987) erklären ihn mithilfe der Early-Learning Theorie des absoluten Gehörs, die davon ausgeht, dass absolutes Gehör in einer kritischen Periode in der Kindheit erlernt wird, die spätestens mit 5-6 Jahren beendet ist. Da viele Kinder ihren Musikunterricht mit einfachen Stücken, meist nur auf den weißen Klaviertasten, beginnen, hat sich das Fenster für den Erwerb absoluten Gehörs möglicherweise schon geschlossen, wenn sie zu schwierigerer Literatur mit vorzeichenreichen Tonarten fortschreiten. Das Ergebnis ist nach Miyazaki (1990) „partielles“ absolutes Gehör, bei dem die Namen weißer Klaviertöne direkt abrufbar sind, wohingegen die Namen schwarzer Klaviertöne von denen der benachbarten weißen abgeleitet werden müssen. Das Benennen schwarzer Klaviertöne wird also als relationaler und nicht als absoluter Prozess beschrieben, denn es beinhaltet im Prinzip einen Vergleichsprozess zwischen zwei Tönen. Dadurch wird die Benennung sowohl langsamer als auch fehleranfälliger. Relativierend muss man allerdings bemerken, dass sowohl die Verlangsamung als auch die Erhöhung der Fehlerquote nur sehr geringfügig sind, so dass die Tonbenennungen sowohl vom Absoluthörer selbst als auch vom beobachtenden Forscher als ebenso absolut wahrgenommen werden wie die Benennungen weißer Klaviertöne.

Einen zweiten Erklärungsansatz schlagen Takeuchi & Hulse (1991) vor, nämlich einen Effekt der Basisrate der Tonklassen in der Musikliteratur. Da vorzeichenarme Tonarten in der westlichen Musikliteratur sowie in zeitgenössischer Unterhaltungsmusik dominieren, hören wir im Laufe unseres Lebens viel mehr weiße als schwarze Klaviertöne. Dies wirkt sich nach Takeuchi & Hulse (1991) auf die Assoziationsstärke zwischen Tönen und Tonnamen aus, die für häufiger gehörte Töne höher ist als für seltener gehörte Töne. Dadurch wird die Benennung, also der Abruf des korrespondierenden Tonnamens, von häufiger gehörten Tönen im Vergleich zu seltener gehörten Tönen erleichtert. Gestützt wird dieser Erklärungsansatz durch einen Befund von Simpson & Huron (1994). Sie prüften Miyazakis (1990) Reaktionszeit-Ergebnisse daraufhin, ob sie durch die Hick-Hyman-Regel erklärbar seien. Diese Regel besagt, dass die Reaktionszeit für einen Reiz umso kürzer ist, je größer seine erwartete Auftrittshäufigkeit ist (Hick, 1952; Hyman, 1953). Nachdem sie mithilfe einer Stichprobe aus der

Musikliteratur (15 zweistimmige Inventionen von J.S. Bach sowie 10 Streichquartette von J. Haydn) die relative Auftrittshäufigkeit aller 12 Tonkategorien einer Oktave bestimmt hatten, konnten sie zwischen dem berechneten Informationsgehalt der Tonklassen und Miyazakis (1990) Reaktionszeit-Ergebnissen für die Tonklassen eine Korrelation von $r = .67$ ($p = .017$) nachweisen. Das bedeutet, dass die Hick-Hyman-Regel 45 % der Varianz in Miyazakis Reaktionszeit-Daten erklären kann und wird von den Autoren als Hinweis dafür interpretiert, dass „the mechanism by which absolute pitch is learned is the same as for innumerable other forms of learning investigated by psychologists“ (Simpson & Huron, 1994, S. 269). Leider stellen die Autoren keine Vermutungen an, durch welche Faktoren die andere Hälfte der Reaktionszeit-Varianz erklärt werden könnte.

Marvin & Brinkman (2000) nutzten eine andere Methode, um den Basisraten Ansatz zu stützen. Sie ließen sowohl Absolut- als auch Nichtabsoluthörer Töne benennen und konnten bei beiden Gruppen eine Verschlechterung der Leistung bei der Benennung schwarzer im Vergleich zu weißen Tönen nachweisen, obgleich sich das Leistungsniveau zwischen den Gruppen natürlich deutlich unterschied. Das spricht gegen eine ausschließliche Erklärung des schwarz-weiß Effekts durch die Early-Learning Theorie, weil Nichtabsoluthörer meistens deutlich später, also nach dem Ende der kritischen Periode, mit dem Musikunterricht beginnen als Absoluthörer. Nimmt man an, dass der schwarz-weiß Effekt bei beiden Gruppen durch ähnliche Lernprozesse verursacht wird, so muss man nach Lernprozessen suchen, die auch nach dem Ende der kritischen Periode wirksam werden. Ein solcher Prozess wäre das Lernen anhand der Basisrate der Töne.

Der Basisraten Ansatz bietet auch eine Erklärung für die bisher nachgewiesenen Klangfarbeneffekte, denn verglichen wurden dort bisher vor allem Sinustöne (die in der Musikliteratur nicht vorkommen) mit „echten“ instrumentalen Klangfarben. In dieser Arbeit wird vermutet, dass die individuelle Vertrautheit mit musikalischem Material – im Sinne einer „persönlichen“ Basisrate – eine wichtige Rolle für den Gedächtnisabruf spielt. Daher wurde die Klangfarbe hier nicht nur unter dem Aspekt „kommt in der Natur vor vs. kommt nicht vor“, sondern auch unter dem der individuellen Lerngeschichte einer Vp (vertraut vs. weniger vertraut) untersucht. Prinzipiell ist aber anzumerken, dass der Early-Learning Ansatz und der Basisraten Ansatz als Facetten des gleichen Mechanismus betrachtet werden können, nämlich dem der Vertrautheit mit bestimmten Tönen. Vertrautheit kann sowohl sehr früh durch das Lernen eines Instruments entstehen, aber auch viel später, und dann sowohl aktiv über das Erlernen von Musikinstrumenten als auch passiv durch bloßes Hören gängiger Musikliteratur. Da hier in Anlehnung an Marvin & Brinkman (2000) angenommen wird, dass die Vertrautheit mit musikalischem Material bei Absolut- und Nichtabsoluthörern ähnliche Effekte hat, werden beide Gruppen mit der gleichen Aufgabe, nämlich der Tonbenennung, konfrontiert und die Effekte von Klangfarbe und Tonklasse geprüft.

Bisher wurden diese Effekte nur auf der Verhaltensebene untersucht (Fehlerhäufigkeit und Reaktionszeiten). Das Anliegen dieses Experiments war die Absicherung der Effekte auf psychophysiologischer Ebene anhand der Pupillenreaktion. Obwohl die primäre Funktion der

Pupille die Kontrolle des Lichteinfalls auf die Retina ist (die „funktionale“ Ebene der Pupillenreaktion, vgl. Hoeks & Levelt, 1993), hat sich die Pupillenreaktion bei einer Vielzahl von Aufgaben als Indikator für den Ressourcenverbrauch sowohl bei der Informationsverarbeitung als auch bei emotionaler Verarbeitung erwiesen (Granholm & Steinhauer, 2004). Dies wird als „nicht-funktionale“ Ebene der Pupillenreaktion bezeichnet, für die der laterale Hypothalamus, der locus coeruleus und der Cortex eine wichtige Rolle spielen (Hoeks & Ellenbroek, 1993). Erweiterungstendenzen der Pupille bilden unter sonst gleichen Umständen unterschiedliche Grade kognitiven Verarbeitungsaufwandes ab (Klix, van der Meer & Preuß, 1984): Je schwieriger die Aufgabe, desto stärker die Tendenz zur Pupillendilatation. Dies wurde mit verschiedenen Aufgabentypen nachgewiesen, z.B. bei der Sprachverarbeitung (Hyönä, Tommola & Alaja, 1995; Just & Carpenter, 1993), bei Wahrnehmungsaufgaben (Verney, Granholm & Dionisio, 2001) und Gedächtnisaufgaben (Granholm, Asarnow, Sarkin & Dykes, 1996; van der Meer, Friedrich, Nuthmann, Stelzel & Kuchinke, 2003). Beispielsweise fanden Kahneman & Beatty (1967) in einer Untersuchung, bei der Vpn Tonvergleiche durchführen mussten (Diskriminationsexperiment mit sehr kleinen Tonabständen) einen signifikanten (negativen) Zusammenhang zwischen der Größe der Pupillendilatation und dem Abstand zwischen den zu unterscheidenden Tönen.

Die Befundlage zu emotionalen Einflüssen auf die Pupillenreaktion ist weniger eindeutig. Mudd, Conway & Schindler (1990) verglichen die Pupillenreaktion mit verbalen Präferenzurteilen für sechs 30 s lange Musikstücke. Sie fanden signifikante Pupillendilatationen bei Musikstücken, die positiv beurteilt wurden, während bei negativ beurteilten Musikstücken signifikante Pupillenkonstriktionen auftraten. Dies ist eines der wenigen Experimente, das die umstrittene Hypothese bidirektionaler Pupillenreaktionen von Hess (1965) bestätigt. Andererseits liegen methodisch verfeinerte Untersuchungen von Steinhauer, Boller, Zubin & Pearlman (1983) und Libby, Lacey & Lacey (1973) vor, die die unmittelbar auf visuelle Reize folgende Pupillenkonstriktion (Lichtreaktion) kontrollierten. Dort wurden Pupillendilatationen sowohl bei als angenehm als auch bei als unangenehm empfundenen Bildern im Vergleich zu neutralen Bildern gefunden. Diese Ergebnisse verweisen darauf, dass sowohl positive als auch negative emotionale Reaktionen Ressourcen beanspruchen. Entscheidend für das Ausmaß der Pupillendilatation scheint demnach eher die Intensität der emotionalen Reaktion als deren Richtung zu sein.

In der Pupillometrie hat sich die Analyse von mehreren Parametern der Pupillenreaktion bewährt. Meist wird entweder der mittlere oder der maximale Pupillendurchmesser in einem interessierenden Zeitintervall berechnet. Nach Nuthmann (2002) ist die Verwendung des mittleren Pupillendurchmessers dann problematisch, wenn die Dauer der Trials innerhalb oder zwischen den Probanden variiert, da in jeden Mittelwert unterschiedlich viele Datenpunkte eingehen. Für diesen Fall empfehlen Beatty & Lucero-Wagoner (2000) die Berechnung des maximalen Pupillendurchmessers (Dilatationspeak), der von der Länge des interessierenden Zeitintervalls unabhängig ist. Da in dem hier vorgestellten Experiment die Trialdauer innerhalb und zwischen den Probanden variiert, wird der Dilatationspeak und nicht die mittlere Pupillendilatation berechnet.

Der Vorteil der Nutzung der Pupillenreaktion bei der Untersuchung von tonspezifischen Effekten auf die Benennungsleistung liegt in der Möglichkeit, den kognitiven Aufwand bei der Tonbenennung abzubilden. Die Fehlerrate zeigt nur die Genauigkeit der Anforderungsbewältigung an, nicht aber, wie viele Ressourcen aufgewandt werden mussten, um zu einem bestimmten Ergebnis zu kommen. Ebenso ist die Reaktionszeit kein Maß für den Ressourcenverbrauch, denn sie bildet nur die Geschwindigkeit bei der Anforderungsbewältigung ab. Eine lange Reaktionszeit kann somit auch ein sequenzielles Abarbeiten von Routineprozessen bedeuten. Außerdem gibt es folgende Probleme bei der Interpretation der Verhaltensdaten: Bei dem durch die Trefferquote gemessenen schwarz-weiß Effekt könnte es sich auch um einen Antwortbias handeln, denn viele Vpn nennen wesentlich häufiger weiße als schwarze Tonnamen und haben bei den weißen somit eine größere Chance, den richtigen Tonnamen zu treffen (vgl. Takeuchi & Hulse, 1991, für eine Diskussion des Antwortbias). Die verlängerten Reaktionszeiten sind zumindest in einigen Experimenten (Miyazaki 1989, 1990 Exp. 1, Marvin & Brinkman, 2000) durch methodische Aspekte mit bedingt, z.B. die Nutzung einer Klaviatur zur Antworterfassung, so dass der Weg zur schwarzen Taste länger ist als der zur weißen Taste. Findet man also bei der Tonbenennung tonspezifische Unterschiede in der Pupillenreaktion, so wäre das ein Hinweis darauf, dass es Absoluthörern tatsächlich schwerer fällt, Töne schwarzer Tasten (und analog Töne fremder Klangfarben) zu benennen und der Effekt kein Artefakt des Versuchsaufbaus ist. Das zweite Ziel dieses Experiments ist also, zusammengefasst, die Absicherung der tonspezifischen Effekte bei der Tonbenennung durch die Pupillenreaktion, wobei hier die Effekte von Tonklasse und Klangfarbe geprüft werden. Folgende spezifische Hypothesen werden in diesem Experiment geprüft:

1. Auf der Verhaltensebene werden in Übereinstimmung mit Miyazaki (1989, 1990), Marvin & Brinkman (2000) sowie Takeuchi & Hulse (1991) eine höhere Fehlerrate und längere Reaktionszeiten bei der Benennung von Tönen schwarzer Klaviertasten (im Vergleich zu weißen Klaviertasten) sowie bei der Benennung von Tönen in fremden (im Vergleich zu vertrauten) Klangfarben erwartet.

2. Es wird angenommen, dass die Benennung von Tönen schwarzer Klaviertasten und Tönen in fremden Klangfarben einen höheren kognitiven Aufwand erfordert als die Benennung von Tönen weißer Klaviertasten und Tönen in vertrauten Klangfarben. Daher werden für die Benennung von Tönen schwarzer (im Vergleich zu weißen) Klaviertasten und Töne in fremden (im Vergleich zu vertrauten) Klangfarben höhere Dilatationspeaks der Pupille erwartet.

3. Aufgrund vorliegender Befunde von Marvin & Brinkman (2000) wird angenommen, dass die erwarteten tonspezifischen Unterschiede bei Absoluthörern und bei Nichtabsoluthörern auftreten. Auf der Verhaltensebene werden bei Nichtabsoluthörern insgesamt eine höhere Fehlerrate und längere Reaktionszeiten erwartet als bei Absoluthörern. Auf psychophysiologischer Ebene werden bei Nichtabsoluthörern insgesamt höhere Dilatationspeaks der Pupille erwartet als bei Absoluthörern, da erstere bei der Tonbenennung auf die Nutzung aufwändiger relationaler Strategien angewiesen sind.

8.1 Methode

8.1.1 Charakterisierung der Stichprobe

Am Experiment nahmen insgesamt 21 Versuchspersonen teil, davon 9 männlich und 12 weiblich. Das Durchschnittsalter betrug 28 (SD = 5,8) Jahre. Dreizehn davon sind Berufsmusiker, Tonmeister, Musiklehrer oder Musikstudenten, die übrigen acht sind Hobbymusiker. Alle Vpn haben mindestens ein Musikinstrument erlernt, 14 Vpn spielen mehr als ein Instrument, und die durchschnittliche Dauer des Instrumentalunterrichts betrug 14,6 (SD = 6,0) Jahre. Folgende Instrumente wurden von den Vpn gespielt: Klavier (alle Vpn), Orgel (1 Vp), Geige (8 Vpn), Bratsche (2 Vpn), Cello (3 Vpn), Gitarre (5 Vpn), Blockflöte (5 Vpn), Querflöte (1 Vp), Klarinette (2 Vpn), Saxophon (1 Vp), Fagott (1 Vp), Horn (1 Vp), E-Bass (1 Vp) und Schlagzeug (2 Vpn). Neun Vpn gaben an, Absoluthörer zu sein, die endgültige Klassifikation der Vpn in Absolut- und Nichtabsoluthörer wurde aber erst bei der Auswertung der Tonidentifikationsleistung vorgenommen. In Übereinstimmung mit ihrer Selbsteinschätzung wurden 9 Vpn als Absoluthörer klassifiziert und 12 als Nichtabsoluthörer. Auf eine weitere Einteilung der Vpn, etwa in „gute“ (> 90 % Richtigurteile) und „schlechte“ (70-90 % Richtigurteile) Absoluthörer, wie von Miyazaki (1990) vorgenommen, wurde hier verzichtet, da die Gruppengrößen dann zu klein werden (< 5), um noch Effekte nachweisen zu können. Die Vpn wurden aus dem Bekanntenkreis der Versuchsleiter rekrutiert, wobei sich vor allem für das Auffinden der Absoluthörer eine Art Schneeballsystem bewährt hat, bei dem ein Absoluthörer weitere ihm z.B. aus der Musikhochschule bekannte Absoluthörer anwarb. Alle Vpn füllten einen an Nuthmann (2002) entlehnten Fragebogen zu ihrer aktuellen gesundheitlichen Situation und Medikamenteneinnahme aus. Dieser ergab, dass keine der Vpn aktuell unter gesundheitlichen Problemen litt oder Medikamente einnahm, die die Pupillenreaktion beeinflussen könnten. Weiterhin wurden mit Ausnahme von korrigierter Kurz- oder Weitsichtigkeit keine Augenprobleme berichtet. Die Vpn nahmen am Experiment freiwillig und ohne Bezahlung teil.

8.1.2 Reizmaterial

Das Reizmaterial für dieses Experiment bestand aus 88 Tönen aus den Oktaven *c* bis *c*'', wobei der Kammerton *a* mit 440 Hz festgelegt wurde. Jeder Ton wurde für 2 Sekunden dargeboten (inklusive Fade-In und Fade-Out von jeweils 200 ms).

Als unabhängige Variablen wurden in der Tonauswahl die Tonklasse (Töne schwarzer vs. weißer Klaviertasten) und die Klangfarbe (den Vpn fremde vs. den Vpn vertraute Klangfarben) variiert. Um Effekte der Tonklasse testen zu können, wurden 44 Töne schwarzer Tasten und 44 Töne weißer Tasten verwendet. Dabei wurde versucht, alle 12 Töne einer Oktave mit etwa gleicher Häufigkeit zu verwenden. Eine gleiche Häufigkeit ist bei der Verwendung gleich vieler schwarzer und weißer Töne nicht möglich, da eine Oktave 5 schwarze und 7 weiße Töne enthält. Somit kam jeder Ton mindestens 6 mal und höchstens 9 mal vor. Um zweitens den Effekt der Klangfarbe testen zu können, wurden insgesamt 10 verschiedene Klangfarben genutzt. Da das Klavier allen Vpn zumindest als Zweit- oder Drittinstrument be-

kannt ist, wurden 24 Klaviertöne als vertraute Töne genutzt. Als fremde Töne wurden ihnen 24 Dreieckstöne gegenübergestellt. In diesen beiden Klangfarben trat jeder Ton der zwei genutzten Oktaven einmal auf. Außerdem wurden 8 Klangfarben von Orchesterinstrumenten genutzt. Die Auswahl der Instrumente erfolgte so, dass aus jeder Instrumentenklasse (Streich-, Holzblas- und Blechblasinstrumente) Instrumente in hoher und in tiefer Lage vertreten waren. Von zwei Instrumenten mit sehr ähnlicher Klangfarbe wurde nur eines ausgewählt, und zwar das, welches tendenziell häufiger gespielt wird. Daher wurden als Streichinstrumente die Geige und das Cello ausgewählt, als Holzblasinstrumente die Querflöte, die Klarinette und das Fagott und als Blechblasinstrumente die Trompete, die Posaune und das Horn. In jeder dieser Klangfarben wurden 5 Töne dargeboten.

Die Töne in instrumentalen Klangfarben wurden live mit einem DAT-Recorder aufgenommen und mit der Software Cool Edit (Syntrillium) nachbearbeitet. Die Aufnahmen wurden mit Hilfe von professionellen Musikern hergestellt, wobei deren Töne während des Spielens mit einem digitalen Stimmgerät überprüft wurden, so dass die Intonation als sehr genau angesehen werden kann. Die Nachbearbeitung beinhaltete ein Zuschneiden der Töne auf die gleiche Dauer von 2 Sekunden, das Einfügen von Fade-In und Fade-Out sowie die Normalisierung der Lautstärke. Die Dreieckstöne wurden ebenfalls mit Cool Edit generiert und auf die gleiche Dauer und Lautstärke gebracht wie alle anderen Töne.

Der Faktor Tonklasse ist objektiv festgelegt: Jeder Ton ist entweder ein schwarzer oder ein weißer Ton. Dagegen ist der Faktor Klangfarbe von jeder Vp abhängig, da hier nicht der Einfluss der 10 Klangfarben an sich überprüft wurde, sondern der Einfluss der Vertrautheit einer Klangfarbe für eine Vp. Daher wurde jede Vp gebeten, alle Instrumente zu nennen, auf denen sie jemals unterrichtet wurde. Die von einer Vp genannten Instrumente wurden für diese Vp dann als vertraut klassifiziert, alle anderen instrumentalen Klangfarben als fremd. Der Nachteil dieser Klassifikation von vertrauten und fremden Klangfarben ist, dass für jede Vp ein etwas anderes Zahlenverhältnis hinsichtlich vertrauter und fremder Klangfarben entstand. Dadurch wird bei der Auswertung eine Analyse über die Items unmöglich. Der Vorteil, der dann zur Entscheidung für dieses Vorgehen führte, ist aber, dass jede Vp zumindest eine ihr sehr vertraute Klangfarbe hörte. Denn allen Vpn war zwar das Klavier vertraut, aber nicht jede Vp hatte Klavier als Hauptinstrument.

Die Töne wurden in einer festen Reihenfolge präsentiert, die so konstruiert war, dass zwischen zwei aufeinander folgenden Tönen immer mindestens 10 Halbtöne lagen, so dass die Berechnung von Intervallen (also die Nutzung relationaler Information) erschwert wurde. Außerdem wurden die Klangfarben randomisiert, so dass sich aufeinander folgende Töne immer in der Klangfarbe unterschieden.

8.1.3 Ablauf des Experiments

Das Experiment fand in einem ruhigen Labor mit einer Hintergrundbeleuchtung von ca. 300 Lux statt. Die Vpn saßen auf einem bequemen Schreibtischstuhl, ohne ihren Kopf anzulehnen oder aufzustützen. Der Kopf der Vp war etwa 80 cm vom Bildschirm, der sich auf Au-

genhöhe befand, entfernt, sowie etwa 1 m von den mit dem Computer verbundenen Lautsprechern.

Zu Beginn des Experiments wurden die Vpn mit dem Ablauf vertraut gemacht, indem ihnen eine schriftliche Instruktion vorgelegt wurde. Während sie die Instruktion lasen und den medizinischen Fragebogen sowie einen Fragebogen zu ihrem musikalischen Ausbildungsstand ausfüllten, konnten ihre Augen an die Beleuchtung im Labor adaptieren. Das Experiment begann, sobald die Vp mit einem Mouseclick signalisiert hatte, dass sie startbereit ist, woraufhin ein Fixationskreuz für 1000 ms auf dem Bildschirm erschien. Dann ertönte der erste Ton aus den Lautsprechern. Die Vpn wurden angewiesen, einen Mouseclick durchzuführen, sobald sie den Namen dieses Tons wussten. Sie wurden instruiert, so schnell wie möglich zu reagieren, aber erst dann, wenn sie sich für einen Tonnamen entschieden hatten. Durch diese Instruktion wurde verhindert, dass die Vpn die Blinzelpause nach dem Mouseclick noch zum Nachdenken nutzten. Da den Vpn diese Instruktion sehr nachdrücklich gegeben und auch betont wurde, wie wichtig dieses Vorgehen für die Ergebnisse sei, kann von einer Compliance der Vpn ausgegangen werden. Als Reaktionszeit wurde die Zeit zwischen Ton-Onset und Mouseclick gemessen. Nach dem Mouseclick blieb das Fixationskreuz noch für 1000 ms auf dem Bildschirm, um eine Weitermessung der Pupillenweite zu ermöglichen. Dies ist wichtig, weil sich in verschiedenen Experimenten gezeigt hat, dass der Dilatationspeak der Pupille häufig erst 300-500 ms nach der Verhaltensantwort auftritt (Beatty, 1982; Hoeks & Levelt, 1993). Danach erschien statt des Kreuzes ein Smiley, der den Vpn signalisierte, dass sie nun erstens der Versuchsleiterin den Namen des soeben gehörten Tons ansagen sollten und zweitens blinzeln oder die Augen zur Entspannung schließen durften. Es wurde kein Feedback gegeben, um die Berechnung von Intervallen von einem richtig benannten Ton aus möglichst zu verhindern.

Durch die doppelte Reaktion (1. Mouseclick, 2. verbale Antwort) wurde sichergestellt, dass nur in der Blinzelpause gesprochen wurde, dass während der Trials keine Wechsel der visuellen Reize notwendig wurden und nur der Finger auf der Mouse bewegt werden musste. Die Vermeidung visueller Reizwechsel während der Trials ist für die Ableitung der Pupillenreaktion vorteilhaft, weil jeder visuelle Reizwechsel eine Lichtreaktion, also eine Konstriktion der Pupille, auslöst. Auch die Vermeidung von Bewegungen während der Trials ist vorteilhaft für die Ableitung der Pupillenreaktion, da jede Bewegung der Augen Artefakte in der Pupillenreaktion hervorrufen kann, besonders wenn der Kopf nicht fixiert ist.

Als maximale Reaktionszeit wurden 5 Sekunden festgelegt. Wenn die Vp in dieser Zeit nicht per Mouseclick reagierte, erschien nach den 5 s (+ 1 s Nachmessung) der Smiley auf dem Bildschirm und der Ton wurde als nicht benannt registriert. Die Vpn waren über dieses Vorgehen informiert, so dass sie keinen Tonnamen ansagten, wenn ohne eine Reaktion ihrerseits nach 6 s ein Smiley erschien.

Die Vpn wurden angewiesen, ihren Blick stets auf das Fixationskreuz zu richten, wann immer es auf dem Bildschirm präsent war, außerdem nicht zu blinzeln und ihre Augen möglichst nicht zu bewegen. Sie wurden gebeten, nur zu blinzeln, wenn der Smiley auf dem Bildschirm

sichtbar war und diese Blinzelpause so lange auszudehnen, wie es für sie angenehm war, meist einige Sekunden. Die probandengesteuerte Blinzelpause wird von Beatty & Lucero-Wagoner (2000) empfohlen, um das Auftreten von Blinzlern in den interessierenden Zeitintervallen zu minimieren. Sobald die Vpn sich für den nächsten Ton bereit fühlten, sollten sie wieder mit der Mouse klicken, woraufhin das Fixationskreuz erschien und nach 1000 ms der nächste Ton (vgl. Abb. 8.1 für das Ablaufschema des Experiments). Das Experiment wurde im Einzelversuch durchgeführt. Es wurden zunächst 4 Übungstrials durchgeführt, dann folgten zwei Blöcke à 44 Tönen, so dass für jede Vp 88 Tonbenennungen vorlagen. Das Experiment dauerte etwa 45 Minuten.

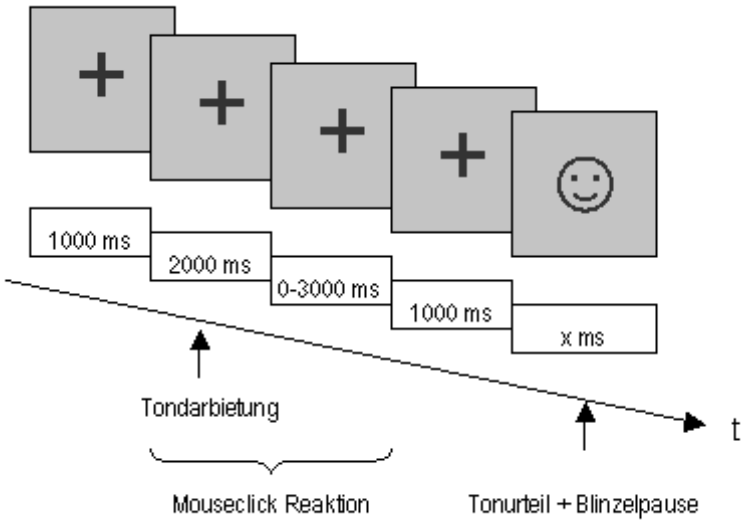

Abb. 8.1: Ablauf eines experimentellen Trials im Pupillenexperiment (t = Zeit)

8.1.4 Geräte

Das Experiment wurde mit zwei PCs durchgeführt, die über den Parallelport vernetzt waren. Der erste PC (Stimulus-PC) diente der Reizpräsentation und war an die Lautsprecher angeschlossen. Das Fixationskreuz und der Smiley wurden auf einem 14 Zoll Bildschirm dargeboten. Der graue Bildschirmhintergrund hatte eine Leuchtdichte von 39.65 cd/m², während die beiden dunkelgrauen Symbole eine vergleichbare Leuchtdichte von etwa 30 cd/m² hatten (Fixationskreuz: 30,58 cd/m², Smiley: 29,56 cd/m²). Das Experiment wurde in Matlab (Version 6.5) unter Nutzung der Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) programmiert. Bei jedem Mouseclick der Vpn sandte der Stimulus-PC Trigger-Signale über das parallele Kabel an den zweiten PC (Pupillen-PC). Der Pupillen-PC diente der Registrierung der Pupillenwerte mit Hilfe eines iView-Systems von Sensomotoric Instruments. Eine Infrarot-Videokamera wurde auf dem Kopf der Vp befestigt, um die Pupillenreaktion des rechten Auges abzuleiten. Eine an der Kamera befestigte Infrarot-Lichtquelle (mit λ = 700-1.049 nm) beleuchtete das rechte Auge und erzeugte dadurch eine Reflektion der Hornhaut. Die Reflektion der Hornhaut wurde dann mit der Videokamera aufgenommen und der horizontale Pupillendurchmesser mit einer Frequenz von 50 Hz (also einem Messpunkt alle 20 s) auf dem Pupillen-PC gespeichert. Außerdem wurden auf dem Pupillen-PC die Trigger-Signale

gespeichert, damit bei der späteren Analyse die Pupillenwerte den Trial-Anfängen und Reaktionszeiten der jeweiligen Vp zugeordnet werden konnten.

8.1.5 Analyse der Pupillenrohwerte

Zunächst wurde ein Computer-Algorithmus angewandt, um Trials mit großen Artefakten (durch Augenbewegungen oder exzessives Blinzeln) zu entfernen sowie kleinere Blinzel-Artefakte durch lineare Interpolation zu korrigieren. Der Algorithmus erkannte Artefakte an abrupten Veränderungen des Pupillendurchmessers, die aufgrund der Trägheit der Pupille nur durch einen Lidschluss (und folglich die Messung einer Pupillenweite von 0 mm) hervorgerufen worden sein konnten. Dann wurden Trials mit falsch benannten Tönen sowie Trials, bei denen die Vp nicht innerhalb von 5 s reagiert hatte, von der weiteren Analyse ausgeschlossen. Da die Reaktionszeiten sehr stark streuten, wurde dann ein Ausreißertest nach der 2-Sigma-Regel durchgeführt, bevor die Daten zwecks Glättung gefiltert wurden. Insgesamt wurden 55,6 % der Trials aufgrund von Artefakten, fehlender Mouseclick-Reaktion oder zu langsamen Antworten von weiteren Analysen ausgeschlossen. Weitere 9,7 % der Trials wurden falsch beantwortet und daher ausgeschlossen, so dass für die statistische Analyse der Pupillenwerte noch 640 Trials (34,7 %) blieben. Der Datenverlust war systematisch auf die Versuchsgruppen verteilt: Bei den Absoluthörern waren 564 Trials (71,3 %) gültig und richtig beantwortet, während 20,6 % der Trials aufgrund von Artefakten, fehlenden oder langsamen Antworten sowie 8,1 % der Trials aufgrund von falschen Antworten ausgeschlossen wurden. Bei den Nichtabsoluthörern waren nur 76 Trials (7,2 %) gültig und richtig beantwortet, während 81,8 % der Trials aufgrund von Artefakten, fehlenden oder langsamen Antworten sowie 11,0 % der Trials aufgrund von falschen Antworten ausgeschlossen wurden.

Das iView System misst den Pupillendurchmesser in Pixeln, so dass als nächstes eine Kalibrierung durchgeführt wurde. Dazu wurde eine künstliche Pupille genutzt, also ein schwarzer Punkt mit dem Durchmesser von 5 mm, der zu Beginn und am Schluss des Experiments jeder Vp auf das geschlossene Augenlid gelegt und mit der Pupillen-Videokamera gefilmt wurde. Nachdem die Größe der künstlichen Pupille in Pixeln bestimmt war, konnte man die Pixelwerte jeder Vp in mm umrechnen. Außerdem konnte durch den Vergleich der Anfangs- und End-Pupillenweite festgestellt werden, ob sich bei einer Vp die Pupillenweite im Verlauf des Experiments ändert. Ein solcher Drift wurde z.B. von Nuthmann (2002), Hyönä et al. (1995) und Rössger (1997) festgestellt, ist aber laut Nuthmann (2002) bei der Berechnung einer Baseline für jedes Trial unproblematisch für die Interpretation der Pupillendaten.

Nach der Kalibrierung wurde für jedes Trial jeder Vp die Baseline des Pupillendurchmessers bestimmt. Als Baseline wurde der durchschnittliche Pupillendurchmesser der während der letzten 200 ms vor Ton-Onset gemessenen Werte definiert. Der durchschnittliche Baseline Pupillendurchmesser, gemittelt über alle gültigen Werte aller Vpn betrug 4,43 mm (2,77 – 6,79 mm). Die große Varianz in den Baselines ist vor dem Hintergrund eines Befundes von Beatty (1982, S. 284), dass die Pupillenreaktion „independent of baseline pupillary diameter over a physiologically reasonable but not extreme range of values“ ist, unproblematisch. Zu

dem gleichen Schluss kommen Hyönä et al. (1995) und Hoeks & Ellenbroek (1993), die eine Unabhängigkeit der Pupillenreaktion vom Baseline Pupillendurchmesser berichten, wenn dieser zwischen 2 und 7 mm liegt. Um auszuschließen, dass die berichteten Effekte in der Pupillenreaktion durch Baseline-Unterschiede zwischen den sechs Bedingungen mit bedingt sind, wurden für jede Vp die Baselines jeder der sechs Bedingungen gemittelt (vgl. Tab. 8.4) und die Unterschiede auf Signifikanz geprüft (vgl. Ergebnisteil). Auch die Dilatationspeaks der Pupille wurden pro Vp für jede der sechs experimentellen Bedingungen gemittelt. Als Dilatationspeak wurde die um die Baseline korrigierte maximale Pupillendilatation im Zeitintervall zwischen Ton-Onset und Reaktionszeit + 500 ms (aufgrund der oben berichteten Trägheit der Pupille) definiert.

8.2 Ergebnisse

8.2.1 Klassifikation der Vpn anhand der Verhaltensdaten

Sowohl unter Absolut- als auch unter Nichtabsoluthörern (zunächst nach der Selbsteinschätzung der Vpn klassifiziert) gab es große interindividuelle Unterschiede in der Trefferrate: Bei Absoluthörern streut die Trefferrate zwischen 73 und 100 %, während sie bei Nichtabsoluthörern zwischen 2 und 43 % variiert. Abbildung 8.2 zeigt die Trefferrate aller 21 Vpn sowie den Cut-off-Punkt für die Klassifikation der Vpn in Absolut- und Nichtabsoluthörer. Als Cut-off-Punkt wurde die 70 % Marke gewählt, da zwischen 43 und 73 % der größte Sprung in der Verteilung liegt.

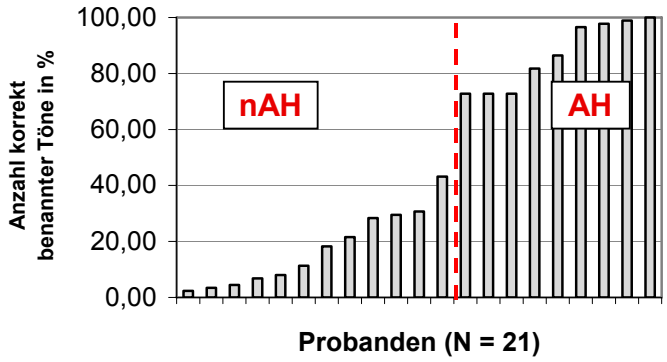

Abb. 8.2: Verteilung der Trefferrate (in %) aller 21 Vpn. Die Trennlinie für die Klassifikation der Vpn in 9 Absoluthörer (AH) und 12 Nichtabsoluthörer (nAH) wurde bei 70 % Richtigurteilen gezogen.

Um zu überprüfen, ob sich Trefferrate und Reaktionszeit zwischen Absolut- und Nichtabsoluthörern unterscheiden, wurden t-Tests für unabhängige Stichproben durchgeführt. Zusätzlich wurde geprüft, ob sich die mittlere Fehlergröße (der Abstand zwischen den genannten Tonnamen und den richtigen Tonnamen, gemessen in Halbtönen) zwischen den beiden Gruppen unterscheidet, wie es von Ward (1999) beschrieben wurde. Zuvor wurden die abhängigen Variablen auf Normalverteilung (Kolmogorov-Smirnov-Test) und Varianzhomogeni-

tät (Levene-Test) überprüft. Tabelle 8.1 zeigt die mittlere Trefferrate, Reaktionszeit und Fehlergröße beider Gruppen.

Tab. 8.1: mittlere Trefferrate, Reaktionszeit und Fehlergröße der Absolut- und Nichtabsoluthörer

Gruppe	Absoluthörer	Nichtabsoluthörer
	Trefferrate	
M (%)	86,62	17,33
SE (%)	4,01	3,85
	Reaktionszeit	
M (ms)	2088	4281
SE (ms)	379	274
	Fehlergröße	
M (HT)	0,17	2,19
SE (HT)	0,05	0,25

M = Mittelwert, SE = Standardfehler, HT = Halbtöne.

Der Kolmogorov-Smirnov-Test ergab, dass die Verteilungen der Trefferrate (Z = 0,885, p = .414), Reaktionszeit (Z = 0,842, p = .477) und Fehlergröße (Z = 0,926, p = .358) nicht signifikant von einer Normalverteilung abweichen. Der Levene-Test ergab homogene Varianzen zwischen den Gruppen bezüglich der Fehlerrate (F = 0,118, p = .735) und Reaktionszeit (F = 0,908, p = .353), nicht aber bezüglich der Fehlergröße (F = 8,857, p = .008), so dass für diese die Freiheitsgrade im t-Test angepasst wurden. Es ergaben sich signifikante Gruppenunterschiede in den Trefferraten (t (19) = - 12,273, p < .001) und Reaktionszeiten (t (19) = 4,819, p < 001). Absoluthörer benennen also deutlich mehr Töne richtig und sind dabei viel schneller als Nichtabsoluthörer. Dadurch wird Hypothese 3 auf der Verhaltensebene bestätigt. Weiterhin findet man in den Daten die von Ward (1999) beschriebenen typischen Fehlerverteilungen, die sich in den beiden Gruppen unterscheiden (Abb. 8.3):

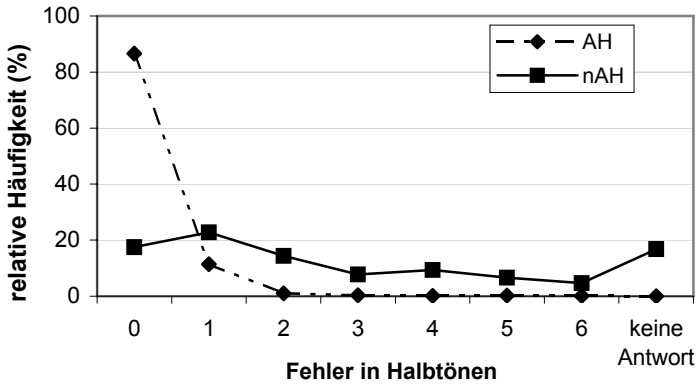

Abb. 8.3: Relative Häufigkeit der Fehler (in %), abhängig von der Zugehörigkeit der Vpn zur Gruppe der Absoluthörer (AH) oder Nichtabsoluthörer (nAH).

Während Absoluthörer fast nur Halbtonfehler machen (die auch typischen Oktavfehler wurden in diesem Experiment nicht erfasst), sind bei Nichtabsoluthörern Fehler von 1 bis 6 Halbtönen vorhanden (da die Oktavlage nicht erfragt wurde, war der größtmögliche Fehler in dieser Aufgabe eine halbe Oktave = 6 Halbtöne). Die mittlere Fehlergröße der Absoluthörer ist signifikant kleiner als die der Nichtabsoluthörer (t (12) = 8,030, p < .001).

Zusammenfassend lässt sich also feststellen, dass die Unterschiede in der Leistung beim Tönebenennen zwischen den zunächst anhand ihrer Selbsteinschätzung klassifizierten Vpn groß genug sind, um die Einteilung in die zwei Gruppen Absoluthörer und Nichtabsoluthörer zu rechtfertigen. Außerdem ist festzuhalten, dass die Selbsteinschätzung der Vpn hinsichtlich ihrer Fähigkeit absolut zu hören, durch die Verhaltensdaten bestätigt wurde. Es scheint also eine realistische Einschätzung dieser Fähigkeit vorzuliegen. Interessanterweise scheint eine gewisse Mindest-Trefferrate (nach diesen Daten: > 43 %) erforderlich zu sein, damit eine Vp sich als Absoluthörer erlebt: Immerhin 7 der 12 Nichtabsoluthörer erzielten eine Trefferrate oberhalb der Ratewahrscheinlichkeit von 8,33 %, aber nicht einmal die Vp mit der für Nichtabsoluthörer hohen Trefferrate von 43 % nahm sich selbst als Absoluthörerin wahr.

8.2.2 Hypothesenbezogene Auswertung der Verhaltensdaten

Tabelle 8.2 zeigt die deskriptiven Daten der beider Gruppen auf der Verhaltensebene:

Tab. 8.2: Mittlere Trefferrate und Reaktionszeit beider Gruppen in den 4 Versuchsbedingungen

Absoluthörer				
Tonklasse	Schwarz		Weiß	
Klangfarbe	Fremd	Vertraut	Fremd	Vertraut
	Trefferate			
M (%)	81,87	91,54	87,64	88,27
SE (%)	6,67	3,88	3,49	4,63
	Reaktionszeit			
M (ms)	1973	1711	1852	1651
SE (ms)	348	269	283	284
Nichtabsoluthörer				
Tonklasse	Schwarz		Weiß	
Klangfarbe	Fremd	Vertraut	Fremd	Vertraut
	Trefferrate			
M (%)	14,21	12,88	15,73	32,94
SE (%)	3,40	4,48	4,13	8,56
	Reaktionszeit			
M (ms)	2869	2522	3013	3327
SE (ms)	265	488	480	428

M = Mittelwert, SE = Standardfehler.

Um zu prüfen, ob die Effekte von Tonklasse und Klangfarbe auf die Trefferrate signifikant sind, wurde eine dreifaktorielle ANOVA mit den Messwiederholungs-Faktoren *Tonklasse* (schwarz vs. weiß) und *Klangfarbe* (vertraut vs. fremd) und einem Gruppierungsfaktor (Absoluthörer vs. Nichtabsoluthörer) berechnet. Diese ergab einen signifikanten Haupteffekt für die *Tonklasse* ($F(1,19) = 4{,}53$, $p = .047$) und die *Klangfarbe* ($F(1,19) = 7{,}50$, $p = .013$) sowie einen signifikanten Gruppenunterschied ($F(1,19) = 129{,}57$, $p < .001$). Die Interaktionen waren mit Ausnahme der dreifachen Interaktion Tonklasse x Klangfarbe x Gruppe ($F(1,19) = 5{,}57$, $p = .029$) nicht signifikant.

In einer nächsten Analyse wurde mit einem Einstichproben t-Test überprüft, ob sich die mittlere Trefferrate der Nichtabsoluthörer in den vier Bedingungen signifikant vom Zufallsniveau (1/12 bzw. 8,3 %) unterschied. Es zeigte sich, dass nur die Trefferrate in der Bedingung weiß-vertraut das Zufallsniveau signifikant überstieg ($t(11) = 2{,}88$, $p = .015$). Dieses Ergebnis stellt für die weitere Hypothesenprüfung der Nichtabsoluthörer-Daten ein Problem dar, da nur Reaktionszeiten aus vergleichbaren Trials, also korrekt benannten Tönen, verglichen werden können. Aufgrund der Zufallsleistung in der Tonbenennung von Nichtabsoluthörern in drei von vier Bedingungen können Effekte von Tonklasse und Klangfarbe auf die Reaktionszeiten für diese Gruppe nicht differenziert werden.

Daher wurden nur die Reaktionszeiten der Absoluthörer einer zweifaktoriellen ANOVA mit den Messwiederholungs-Faktoren *Tonklasse* (schwarz vs. weiß) und *Klangfarbe* (vertraut vs. fremd) zugeführt. Diese ergab einen nur auf dem 10 % Niveau signifikanten Haupteffekt der *Klangfarbe* ($F(1,8) = 4{,}31$, $p = .071$), während es keinen signifikanten Haupteffekt der *Tonklasse* gab ($F(1,8) = 2{,}127$, $p = .183$) sowie keine signifikante Interaktion.

Während Hypothese 1 für die Trefferrate also als bestätigt gelten kann, wurde sie für die Reaktionszeiten nicht bestätigt.

8.2.3 Hypothesenbezogene Auswertung der Pupillendaten

Da nur 7,2 % der Trials von Nichtabsoluthörern für die weitere Auswertung zur Verfügung standen und da ihre Trefferrate nur in einer von vier experimentellen Bedingungen das Zufallsniveau überstieg, konnten Effekte von Tonklasse und Klangfarbe auf die Dilatationspeaks der Pupille nur für die Gruppe der Absoluthörer analysiert werden. Abbildungen 8.4 und 8.5 zeigen die gemittelten, baseline-korrigierten Pupillenerweiterungen aller 9 Absoluthörer, getrennt nach den Faktoren Tonklasse und Klangfarbe. Eine explorative Analyse der Dilatationspeaks ergab, dass diese vor allem in den Bedingungen mit fremden Klangfarben sehr stark streuten. Der Grund hierfür wird aus Abbildung 8.5 ersichtlich, in der die Reaktionen nach künstlichen Dreieckstönen und fremden instrumentalen Klangfarben differenziert dargestellt sind. Um für die Analyse des Klangfarben-Effekts nur Töne mit vergleichbarer Klang-Komplexität zu vergleichen und emotionale Einflüsse auf die Pupillenreaktion zu minimieren, gingen in die folgende Analyse nur noch Trials mit Tönen in instrumentalen Klangfarben ein. In Tabelle 8.4 sind die mittleren Dilatationspeaks und die mittleren Baselines getrennt nach den Faktoren Tonklasse und Klangfarbe zusammengefasst.

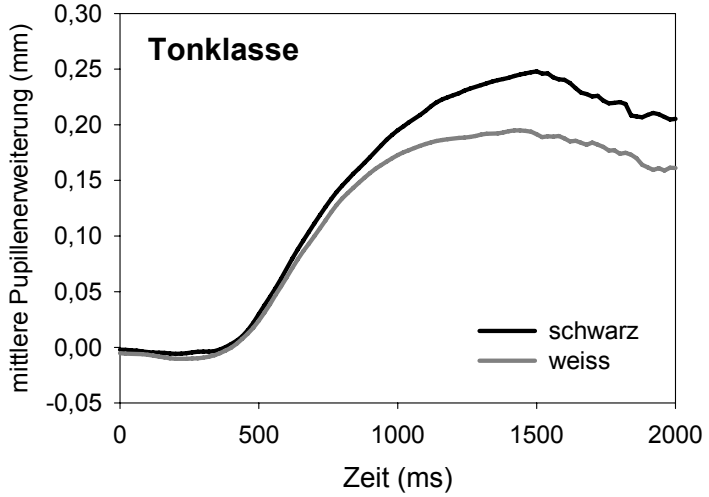

Abb. 8.4: Mittlere baseline-korrigierte Pupillenerweiterung in Abhängigkeit von dem Faktor Tonklasse (schwarz vs. weiss), nur Absoluthörer.

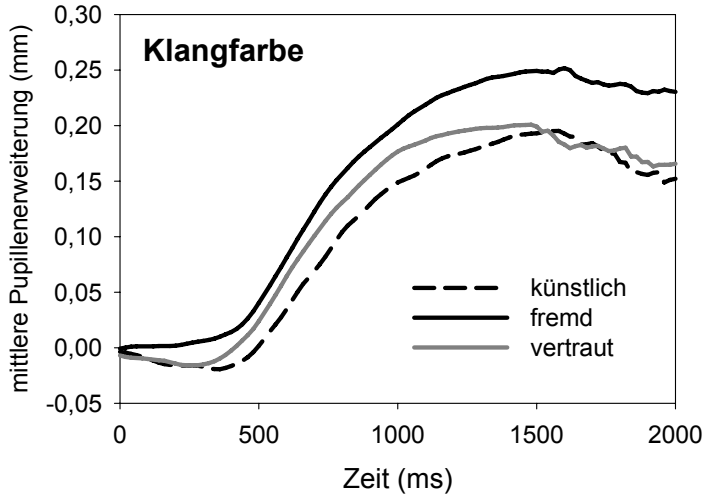

Abb. 8.5: Mittlere baseline-korrigierte Pupillenerweiterung in Abhängigkeit von dem Faktor Klangfarbe (vertraut vs. fremd vs. künstlich), nur Absoluthörer

Tabelle 8.4: Mittlere Dilatationspeaks und Baselines in den 4 Versuchsbedingungen

Tonklasse	Schwarz		Weiß	
Klangfarbe	Fremd	Vertraut	Fremd	Vertraut
		Baseline		
M (mm)	4,385	4,353	4,367	4,440
SE (mm)	0,238	0,245	0,244	0,260
		Dilatationspeak		
M (mm)	0,383	0,354	0,358	0,296
SE (mm)	0,056	0,056	0,056	0,045

M = Mittelwert, SE = Standardfehler.

Der Kolmogorov-Smirnov-Test ergab, dass die Verteilungen der Dilatationspeaks (Z = 0,628, p = .825) und Baselines (Z = 0,334, p = 1,0) nicht signifikant von einer Normalverteilung abweichen. Um auszuschließen, dass sich die Baselines zwischen den vier Bedingungen signifikant unterscheiden, wurde zunächst eine zweifaktorielle ANOVA mit den Messwiederholungs-Faktoren *Tonklasse* (schwarz vs. weiß) und *Klangfarbe* (vertraut vs. fremd) und der Baseline als abhängiger Variablen durchgeführt. Es ergaben sich keine signifikanten Haupteffekte der Faktoren *Tonklasse* (F(1,8) = 1,507, p = .254) oder *Klangfarbe* (F(1,8) = 1,096, p = .326) und keine Interaktion. Somit kann ein Einfluss des Baseline Pupillendurchmessers auf die Dilatationspeaks ausgeschlossen werden. Des weiteren wurde mit einem t-Test für verbundene Stichproben überprüft, ob es einen Drift in den Baselines gibt. Dabei wurden für jede Vp die Baselines der ersten 10 Trials (m = 4,635 mm, SE = 0,303 mm) sowie der letzten 10 Trials (m = 4,234 mm, SE = 0,259 mm) gemittelt und miteinander verglichen (r = 0,697, p = .037; t (8) = 1,802, p = .109). Da der Unterschied nicht signifikant ist, kann ein Drift in der Pupillenweite für dieses Experiment ausgeschlossen werden.

Um die Effekte von Tonklasse und Klangfarbe auf die Dilatationspeaks der Pupille auf Signifikanz zu prüfen, wurde eine zweifaktorielle ANOVA mit den Messwiederholungs-Faktoren *Tonklasse* (schwarz vs. weiß) und *Klangfarbe* (vertraut vs. fremd) berechnet. Diese ergab einen signifikanten Haupteffekt für die *Tonklasse* (F(1,8) = 7,930, p = .023) sowie einen auf dem 10 % Niveau signifikanten Haupteffekte für die *Klangfarbe* (F(1,8) = 4,080, p = .078). Die Interaktion Tonklasse x Klangfarbe wurde nicht signifikant. Dieses Ergebnis bestätigt Hypothese 2.

Um die Pupillenreaktion von Absoluthörern und Nichtabsoluthörern zu vergleichen und Hypothese 3 auf psychophysiologischer Ebene zu prüfen, wurden die Dilatationspeaks in der einzigen Bedingung, in der Nichtabsoluthörer signifikante Tonbenennungsleistungen zeigten (weiß-vertraut), verglichen. Ein t-Test für unabhängige Stichproben ergab einen signifikanten Unterschied (t(14) = 2,175, p = .047), wobei die Nichtabsoluthörer höhere Dilatationspeaks der Pupille (m = 0,536, SE = 0,112) aufwiesen als die Absoluthörer (vgl. Abb. 8.6). Um zu zeigen, dass die Gruppenunterschiede in den Dilatationspeaks einen erhöhten kognitiven Aufwand bei der korrekten Tonbenennung und keine allgemeine Überforderung der Nichtabsoluthörer widerspiegeln, werden in Abbildung 8.6 auch die mittleren Pupillendurchmesser bei falsch benannten Tönen in allen Bedingungen dargestellt. Hier ergaben sich keine Gruppenunterschiede. Demnach kann Hypothese 3 auch auf psychophysiologischer Ebene als bestätigt angesehen werden.

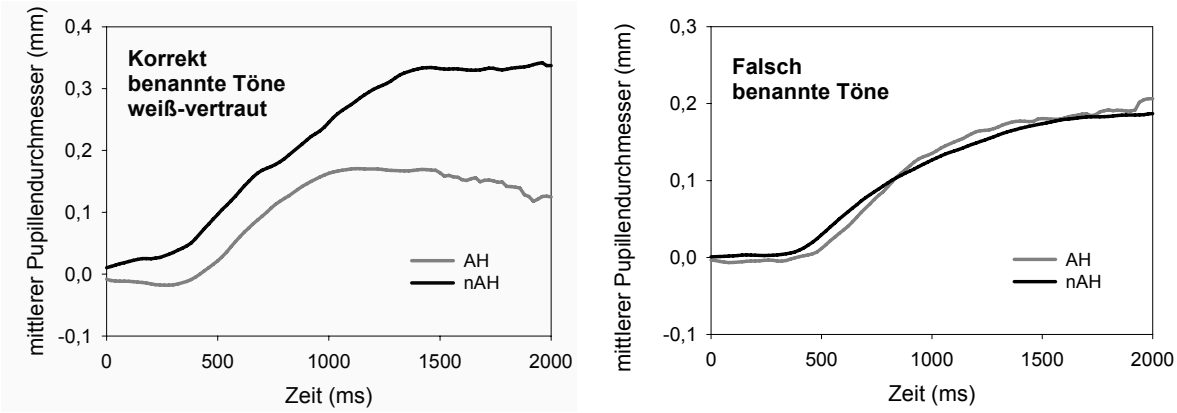

Abb. 8.6: Mittlere baseline-korrigierte Pupillenerweiterung für korrekt benannte Töne in der Bedingung weiß-vertraut (links) und für falsch benannte Töne in allen Bedingungen (rechts), in Abhängigkeit von der Gruppenzugehörigkeit (AH = Absoluthörer, nAH = Nichtabsoluthörer).

8.3 Diskussion

Mit dem vorliegenden Experiment sollte geprüft werden, ob sich in der Pupillenreaktion tonspezifische Effekte auf die Tonbenennung nachweisen lassen. Dabei wurden folgende Ergebnisse erzielt: Auf der Verhaltensebene konnten bei Absoluthörern und Nichtabsoluthörern signifikante Effekte von Tonklasse und Klangfarbe auf die Trefferrate, jedoch nicht auf die Reaktionszeit bei der Tonbenennung nachgewiesen werden. Auf psychophysiologischer Ebene konnte bei Absoluthörern ein signifikanter Effekt der Tonklasse und ein tendenziell signifikanter Effekt der Klangfarbe auf die Dilatationspeaks der Pupille nachgewiesen werden. In der Bedingung weiß-vertraut ergab sich außerdem ein signifikanter Unterschied zwischen den Dilatationspeaks der Pupille von Absolut- und Nichtabsoluthörern.

Der nachgewiesene Effekt der Tonklasse auf die Trefferrate bestätigt vorhandene Befunde von Heyde (1987), Marvin & Brinkman (2000), Miyazaki (1989, 1990) und Takeuchi & Hulse (1991). Jedoch sind sowohl die berichteten Tonklassen-Unterschiede in der Trefferrate als auch in der Reaktionszeit relativ klein, verglichen mit anderen Studien. Miyazaki (1990) fand in Experiment 1 bei den sehr guten Absoluthörern (> 90 % Richtigurteile) zwar schwarz-weiß Unterschiede in der gleichen Größenordnung (einige Prozentpunkte) hinsichtlich der Trefferrate, jedoch bei den mittelguten Absoluthörern (70-90 % Richtigurteile) Unterschiede von etwa 20 % und bei seinen schlechtesten Absoluthörern (70 % Richtigurteile) sogar Unterschiede von rund 40 %. Bei den Reaktionszeiten betrug der schwarz-weiß Unterschied in allen drei Gruppen etwa 200 ms. Die großen Differenzen in den Trefferraten sind allerdings z.T. auf seinen Versuchsaufbau, nämlich die Antwortregistrierung mit einer Klaviatur, zurückzuführen. In seinem zweiten Experiment, bei dem ein Tonklassenkreis zur Antworterfassung genutzt wurde, fand Miyazaki (1990) über alle drei Gruppen gemittelt lediglich schwarz-weiß Unterschiede in der Trefferrate von 10 %. Bei der Aufgabe des dritten Experiments, die Tonnamen einfach anzusagen, sanken die schwarz-weiß Unterschiede auf 7 % in der Trefferquote, während der Reaktionszeit-Unterschied in beiden Experimenten mit 200 ms stabil blieb. Während Heyde (1987) keine Reaktionszeiten berichtet, sind ihre schwarz-weiß Effek-

te hinsichtlich der Trefferrate mit denen von Miyazaki (1990) vergleichbar. In einem weiteren Experiment fand Miyazaki (1989) allerdings mit den hier berichteten Ergebnissen vergleichbare Reaktionszeit-Unterschiede von nur rund 100 ms, die aber aufgrund seiner höheren Trialanzahl signifikant werden. Bei Marvin & Brinkman (2000) wurden die schwarz-weiß Unterschiede nur auf der Ebene der Reaktionszeiten gefunden und betrugen wie bei Miyazaki (1990) etwa 200 ms. Methodisch ist an der Studie von Marvin & Brinkman (2000) zu bemängeln, dass sich die Vpn bei der Entscheidung z.B. für ein *cis* zwischen den Kategorien *cis* und *des* entscheiden mussten. Es gab also zwei Kategorien für den gleichen Ton, was die Entscheidung für einen schwarzen Ton sicherlich verlangsamt hat. Während das Ergebnis des vorliegenden Experiments hinsichtlich der Trefferraten mit Marvin & Brinkmans (2000) Ergebnissen vereinbar ist, ist der Reaktionszeit-Unterschied mit Miyazakis (1989) Befunden vergleichbar.

Ein Grund für den vergleichsweise kleinen Effekt der Tonklasse in den Verhaltensdaten des vorliegenden Experiments könnte sein, dass die Anforderung schwieriger war als in den übrigen berichteten Experimenten. Zum Einen wurden mit 10 verschiedenen Klangfarben deutlich mehr Klangfarben eingesetzt als bei Miyazaki (1989, 1990), der nur zwei künstliche Klangfarben sowie Klaviertöne einsetzte. Auch Marvin & Brinkman (2000) nutzten nur zwei Klangfarben, während Heyde (1987) und Takeuchi & Hulse (1991) nur Klaviertöne nutzten. Die Konzentration auf so viele verschiedene Klangfarben könnte die schwarz-weiß Unterschiede nivelliert haben. Zum Anderen wurde die Tonbenennung im Vergleich zu den übrigen berichteten Experimenten durch die gleichzeitige Ableitung der Pupillenreaktion erschwert. Ein Grund für die nur geringen Unterschiede in der Reaktionszeit könnte in der hier verwendeten Methode zur Antworterfassung liegen. Da Töne nicht auf einer Tastatur gesucht werden mussten und keine doppelten Tonnamen für Töne schwarzer Tasten vorgegeben waren, war der Einfluss methodischer Artefakte auf die Reaktionszeiten geringer als z.B. bei Miyazaki (1989, 1990 Exp. 1) und Marvin & Brinkman (2000).

Interessanterweise erwies sich die Pupillenreaktion als sensiblerer Indikator für die Abbildung von Effekten der Tonklasse als die Reaktionszeit, denn Dilatationspeaks der Pupille waren bei Tönen schwarzer Tasten signifikant größer als bei Tönen weißer Tasten. Dieses Ergebnis weist darauf hin, dass der Effekt der Tonklasse auf die Tonbenennungsleistung nicht nur auf einem Antwortbias oder anderen Artefakten in der Versuchsdurchführung beruhen kann. Die Benennung von Tönen schwarzer Tasten beansprucht tatsächlich mehr Ressourcen als die Benennung von Tönen weißer Tasten. Dieser Befund stützt die Annahme, dass durch das häufigere Hören weißer Töne (sowohl in früher Kindheit als auch später) jene so gut mit den entsprechenden Tonnamen assoziiert sind, dass der Abruf der Tonnamen aufwandsärmer zu bewältigen ist als der für schwarze Töne. Schwarze Töne dagegen scheinen weniger fest mit den zugehörigen Tonnamen assoziiert zu sein.

Der nachgewiesene Effekt der Klangfarbe auf die Trefferrate sowie der marginal signifikante Klangfarben-Effekt in der Reaktionszeit bestätigt Befunde von Marvin & Brinkman (2000) und Miyazaki (1989) und erweitert sie um den Aspekt der individuellen Vertrautheit mit Klangfar-

ben. Denn während Miyazaki nur künstliche Töne mit Klaviertönen verglich, und Marvin & Brinkman (2000) zwar Klavier- und Geigentöne verglichen, jedoch ohne die jeweilige Vertrautheit mit dem Instrument bei ihren Vpn zu erfragen, wurden hier explizit den Vpn vertraute mit ihnen weniger vertrauten Klangfarben verglichen. Dieses Ergebnis bedeutet, dass der beobachtete Klangfarbeneffekt tatsächlich auf die Lerngeschichte der Vpn zurückzuführen ist und nicht nur auf einen Effekt der Basisrate von Klangfarben in der Musikliteratur.

Der Effekt der Klangfarbe wurde durch die psychophysiologischen Daten zumindest tendenziell bestätigt, denn die Dilatationspeaks der Pupille sind für Töne in fremden (instrumentalen) Klangfarben größer als für Töne in vertrauten Klangfarben. Aus Befragungsdaten der Absoluthörer lässt sich ablesen, dass einige Vpn die Töne in fremden Klangfarben benannt haben, indem sie Intervalle zu vorher benannten Tönen in einer vertrauten Klangfarbe gebildet haben. Eine mögliche Erklärung für die höhere Ressourcenkonsumption bei der Benennung von Tönen in fremden (verglichen mit vertrauten) Klangfarben könnte demnach die aufwändige Nutzung relationaler Strategien sein, auch wenn diese von den Vpn als automatisch ablaufender Prozess empfunden wurde. Um diese Interpretation zu erhärten, wäre in einer weiteren Untersuchung eine Strategiebefragung nach jedem benannten Ton nötig, um Trials mit unterschiedlichen Strategien (z.B. absolut vs. relativ) vergleichen zu können.

Interessant ist das Ergebnis, dass sich die Pupillenreaktionen bei Tönen in fremden instrumentalen und Tönen in künstlichen Klangfarben stark unterscheiden. Entgegen der Annahme, dass die Benennung von künstlichen Tönen am schwierigsten ist, zeigten sich dort nicht die größten, sondern die kleinsten Pupillendilatationen. Dies könnte einerseits ein Überlastungseffekt sein, wie er auch von Peavler (1974) gefunden wurde. In Peavlers Experiment stieg die Pupillendilatation bei der Aufgabe, Ziffernfolgen zu memorieren, mit der Anzahl an Ziffern an. Die höchste Pupillendilatation trat bei 7 Ziffern auf, wohingegen die Pupillendilatation ab 8 Ziffern auf etwas niedrigerem Niveau stagnierte. Bei Granholm (1996), der die gleiche Aufgabe verwandte, die Vpn aber im Unterschied zu Peavler nicht über die Länge der auftretenden Ziffernfolgen informiert hatte, kam es ab dem Zeitpunkt der Ressourcenüberlastung sogar zur Pupillenkonstriktion. Aus diesen beiden Experimenten lässt sich schließen, dass die Pupille bei schwieriger werdenden Anforderungen nicht mit immer größerer Dilatation reagiert, sondern dass es eine Grenze gibt, ab der es zu keiner weiteren Dilatation kommt. Es sind jedoch weitergehende Untersuchungen notwendig, um entscheiden zu können, ob die vergleichsweise kleine Pupillendilatation bei der Benennung von künstlichen Tönen, die ja im Unterschied zu Peavlers und Granholms KZG-Aufgaben ein Effekt des Abrufs aus dem LZG ist, auf einer Ressourcenüberlastung beruht.

Eine alternative Interpretation der vergleichsweise kleinen Pupillendilatation bei der Benennung von künstlichen Tönen ist ein Einfluss emotionaler Reaktionen auf die Pupillenreaktion. Möglicherweise werden künstliche Töne aufgrund des weniger komplexen Obertonspektrums als neutraler empfunden als Töne in instrumentalen Klangfarben. In Übereinstimmung mit Steinhauer et al. (1983) und Libby et al. (1973) wäre bei als neutral empfundenen Tönen eine geringere Pupillendilatation zu erwarten als bei emotional „ansprechenderen“ Tönen.

Um diese Interpretation zu erhärten, wäre eine Untersuchung in Anlehnung an Mudd et al. (1990) erforderlich, bei der verbale Präferenzurteile bezogen auf Töne in unterschiedlichen Klangfarben mit Pupillenreaktionen verglichen werden. In ähnlicher Weise könnte man auch emotionale Einflüsse auf den Effekt der Tonklasse überprüfen.

Schließlich konnte Hypothese 3, in der postuliert wird, dass sich in den Verhaltensdaten der Nichtabsoluthörer ebenfalls Tonklassen- und Klangfarben-Effekte ergeben sollten, bestätigt werden, denn den Nichtabsoluthörern gelang nur bei weißen Tönen und vertrauten Klangfarben eine überzufällige Tonbenennung. Methodisch hat dies die Konsequenz, dass die Effekte von Tonklasse und Klangfarbe weder hinsichtlich der Reaktionszeiten noch hinsichtlich der Pupillenreaktionen differenziert werden können. Inhaltlich dagegen ist dieses Ergebnis interessant, da es zeigt, dass Nichtabsoluthörer durchaus eine überzufällige Leistung bei der Tonbenennung erzielen können, wenn sie mit sehr vertrautem Tonmaterial konfrontiert werden – vertraut sowohl hinsichtlich der Klangfarbe als auch hinsichtlich der Tonklasse. Die nachgewiesenen Unterschiede in der Pupillenreaktion zwischen Absoluthörern und Nichtabsoluthörern sind ein Hinweis darauf, dass die Tonbenennung für Nichtabsoluthörer aufwändiger ist als für Nichtabsoluthörer. Der Grund hierfür könnte die Nutzung relationaler Strategien bei Nichtabsoluthörern sein.

Ursache für die in beiden Gruppen beobachteten tonspezifischen Effekte könnte eine größere Repräsentation von vertrauten Tönen im auditorischen Cortex sein, wie sie von Pantev und Mitarbeitern berichtet wurde (Pantev, Roberts, Schulz, Engelien & Ross, 2001; Pantev, Oostenveld, Engelien, Ross, Roberts & Hoke, 1998). In dieser Studie hatten Vpn die Aufgabe, sich Töne anzuhören, während ein Magnet-Enzephalogramm (MEG) abgeleitet wurde. Es zeigte sich, dass die kortikale Repräsentation von Tönen in instrumentalen Klangfarben gegenüber der von künstlichen Sinustönen um rund 25 % vergrößert ist (Pantev et al., 1998) und dass es eine spezifische Vergrößerung für die Töne des eigenen Hauptinstruments im Vergleich mit einer anderen instrumentalen Klangfarbe gibt (Pantev et al., 2001). Demnach wurde sowohl ein Effekt der Basisrate (seltene Sinustöne vs. häufige Instrumententöne) als auch ein Effekt der individuellen Lerngeschichte (Eigeninstrument vs. Fremdinstrument) nachgewiesen. Interessanterweise traten die Effekte sowohl bei Absoluthörern als auch bei Nichtabsoluthörern auf, waren aber bei Absoluthörern etwas stärker ausgeprägt. Betrachtet man die Ergebnisse von Pantev et al. (1998, 2001) im Zusammenhang mit den hier berichteten Verhaltensdaten, so lässt sich annehmen, dass das häufige Hören von bestimmten Tönen die kortikale Repräsentation dieser Töne vergrößert und sowohl bei Absoluthörern als auch bei Nichtabsoluthörern die Tonbenennung erleichtert.

Zusammenfassend kann man festhalten, dass die psychophysiologischen Ergebnisse die Verhaltensdaten bestätigen und zum Teil erweitern. Sie liefern wichtige Hinweise für die Interpretation der Verhaltensdaten, da mithilfe der Pupillenreaktion ein Unterschied in der Ressourcenkonsumption nachweisbar ist, der in den Verhaltensdaten so klar nicht sichtbar ist. Angesichts des Vorteils, der aus der Nutzung der Pupillenreaktion für die Interpretation der tonspezifischen Effekte bei Absoluthörern entstanden ist, stellt sich die Frage, wie man

auch die Gedächtnisprozesse von Nichtabsoluthörern mit der Pupillenreaktion untersuchen könnte. In einer Tonbenennungsaufgabe erscheint dies nicht möglich, da die hier verwendete Anzahl an Trials die Kapazität der Nichtabsoluthörer bereits erschöpfte. Eine Verlängerung des Experiments zwecks Erhöhung der Anzahl auswertbarer Trials ist demnach nicht sinnvoll. Als Lösung käme demnach nur eine Aufgabe in Frage, die sowohl für Absoluthörer als auch für Nichtabsoluthörer lösbar ist. Denkbar wäre eine Gedächtnisaufgabe im Sinne von Rakowski & Morawska-Büngeler (1987), bei der beiden Gruppen Töne vorgespielt werden, die sie sich über einen bestimmten Zeitraum merken müssen, um dann eine Aufgabe mit diesen Tönen zu lösen, beispielsweise einen Tonvergleich durchzuführen oder ein Intervall zu benennen. Problematisch dabei könnte allerdings sein, dass Rakowski & Morawska-Büngeler (1987) erst ab Zeiträumen von über einer Minute Gruppenunterschiede fanden, Zeitintervalle, die in der bisherigen Literatur zur Pupillometrie aufgrund der dabei auftretenden Blinzeltendenz unüblich sind.

Dass in diesem Experiment der Nachweis gelungen ist, dass die individuelle Vertrautheit mit Tönen unterschiedlicher Tonklasse und Klangfarbe sich sowohl auf die Tonbenennungs-Leistung von Absoluthörern als auch auf die von Nichtabsoluthörern auswirkt, ist ein Hinweis darauf, dass auch bei so „exotischen“ und im Alltag der meisten Menschen nicht nützlichen Aufgaben wie der Tonbenennung allgemeingültige, d.h. über den Rahmen dieser Aufgabe hinausgehende, Gedächtnisprinzipien wie z.B. das Lernen anhand von Auftrittshäufigkeiten wirksam werden.

9 Motorische Gedächtnisinhalte bei der Tonarterinnerung

Die theoretische Betrachtung der Rolle motorischer Erinnerungsprozesse beim Abrufen von vertrauten Melodien (vgl. Kap. 4.4.5) hat gezeigt, dass singmotorische Gedächtnisinhalte von motorischen Kontextinformationen getrennt betrachtet werden sollten. Aufschlussreich für die Rolle singmotorischer Gedächtnisinhalte beim Melodieabruf ist die berichtete PET-Untersuchung von Halpern & Zatorre (1999), deren Ergebnis einen Melodieabruf ohne die Beteiligung von Subvokalisationsprozessen nahe legt. Andererseits sprechen die in Kapitel 3.1.2 berichteten Ergebnisse zum leichteren Abruf von Tonarten durch Produktions- im Vergleich zur Wiedererkennungsmethoden (Halpern, 1989) dafür, dass singmotorische Gedächtnisinhalte nicht ganz unwichtig für den Tonartabruf sein könnten. Daneben könnten motorische Kontextinformationen, z.B. vorgestellte Spielbewegungen bei den in Kap. 6.8 beschriebenen Instrumentalmusikern, den Tonartabruf erleichtern.

Um die Rolle sowohl singmotorischer Gedächtnisinhalte als auch motorischer Kontextinformationen zu untersuchen, wurden zwei Experimente durchgeführt. Zunächst wurde in einem Experiment mit Instrumentalmusikern überprüft, ob das Fehlen singmotorischer Gedächtnisinhalte zu einem (im Vergleich zu den Ergebnissen von Chorsängern) schlechteren Tonartabruf führt. Die Logik dieses Experiments ist folgende: Spieler von Holzblasinstrumenten (Querflöte, Klarinette, Oboe, Fagott) können beim Spielen ihres Instruments, anders als z.B. Pianisten oder Spieler von Streichinstrumenten, nicht gleichzeitig singen. Eine informelle Befragung (Danko, 2005) von Professoren der Berliner Musikhochschulen „Hanns Eisler“ und „Universität der Künste“ hat ergeben, dass einige Professoren zwar manchmal schwierige Passagen aus zu spielenden Stücken singen lassen, damit ihre Studenten ein Gefühl für die richtige Phrasierung bekommen. Im Vergleich zu der Häufigkeit, mit der sie am Instrument geübt werden, werden diese Stücke jedoch nur verschwindend selten gesungen. Dies bedeutet, dass Instrumentalmusiker bei der Aufgabe, gelernte Melodien aus dem LZG abzurufen und dann zu singen, weder auf eine singmotorische Erinnerung zurückgreifen können noch auf das gewohnte instrumentenspezifische motorische Programm. Statt dessen sind sie vor allem auf das Abrufen der auditiven Information angewiesen. Zusätzlich können sie sich jedoch die gewohnte Spielsituation vorstellen. Um herauszufinden, welche Rolle solche Vorstellungen spielen, wurden die Instrumentalmusiker nach dem Abruf der Melodien nach ihren Erinnerungsstrategien bei der Anforderungsbewältigung befragt. Des weiteren wurde, ähnlich wie bei den in Kap. 7 beschriebenen Schülern, eine Wiedererkennungsaufgabe durchgeführt, in der den Instrumentalmusikern verschiedene transponierte Versionen einer Melodie vorgelegt wurden, die sie beurteilen sollten. Dadurch sollte überprüft werden, ob sich ähnlich wie bei Halpern (1989) ein Leistungsvorteil in der Produktionsaufgabe zeigt.

Das zweite Experiment greift auf das Doppelaufgaben-Paradigma zurück (vgl. Kap. 1 und 4.4.5). Es wurden Sekundäraufgaben vorgegeben, die selektiv das „innere Hören“ im Sinne der auditiven Vorstellung einer Melodie und das „innere Singen“ im Sinne der singmotorischen Vorstellung einer Melodie beanspruchen sollten. Untersucht wurde der Einfluss der Sekundäraufgaben auf die Leistung in der Primäraufgabe, die Produktionsgenauigkeit, sowie

auf die für den Abruf benötigte Zeit. In diesem Experiment wurde also nicht nur der (weitgehende) Wegfall der motorischen Information untersucht, wie bei den Instrumentalmusikern, sondern umgekehrt auch der (weitgehende) Wegfall der auditiven und somit eine Aufgabenbewältigung überwiegend auf Basis der motorischen Information. Die grundsätzliche Annahme beider Experimente ist die, dass singmotorische Gedächtnisinhalte für den Tonartabruf aus dem LZG nicht zwingend erforderlich sind (Hypothesen 4 a und b).

9.1 Abruf von Instrumentalmusik

9.1.1 Methode

9.1.1.1 Stichprobe

Die Stichprobe bestand aus 40 Instrumentalmusikern, die ein Holzblasinstrument im Hauptfach „Orchestermusik“ an einer der beiden Berliner Musikhochschulen studierten. 12 Vpn waren Flötisten, 6 Vpn waren Klarinettisten, 13 Vpn spielten Oboe und 9 Vpn Fagott. Sie waren im Durchschnitt 24,02 (SD = 2,70) Jahre alt, spielten seit dem Alter von 7,44 (SD = 2,93) Jahren ein Instrument und hatten durchschnittlich 16,60 (SD = 3,35) Jahre Instrumentalunterricht, 0,87 (SD = 1,67) Jahre Gesangsunterricht und 5,40 (SD = 3,61) Jahre Gehörbildungsunterricht erhalten. Sie verfügten im Durchschnitt über 9,08 (SD = 4,17) Jahre Orchestererfahrung. Die Versuchsteilnehmer wurden zum Teil über ihre Hauptfach-Lehrer und zum Teil durch Weiterempfehlung anderer Vpn rekrutiert. Die Versuchsteilnahme erfolgte freiwillig und ohne Bezahlung.

9.1.1.2 Versuchsdurchführung

Das Experiment bestand aus drei Teilaufgaben: einer Produktionsaufgabe, einer Wiedererkennungsaufgabe und einem Screening-Test auf absolutes Gehör. Zuletzt wurde den Vpn ein Fragebogen zu ihrem musikalischen Hintergrund sowie zu ihrer Lerngeschichte bezüglich des Mozart-Konzerts und der Orchesterstellen vorgelegt. Die Durchführung und Auswertung der Produktionsaufgabe wurde in Kap. 6.8 beschrieben. Dort wurden auch die zu produzierenden Melodien charakterisiert. Die Durchführung und Auswertung des Screening-Tests auf absolutes Gehör wurde in Kap. 6.10.1 beschrieben.

In der Wiedererkennungsaufgabe wurden den Vpn jeweils 7 Versionen des Mozart-Konzerts des eigenen Instruments (vgl. Kap. 6.8.2) vorgespielt: die Originalversion sowie Transpositionen um ± 1, 2 und 3 Halbtöne. Vorgespielt wurde immer die erste Melodiephrase des Konzerts, die für Fagott- und Klarinettenkonzert 4 Takte umfasste, für das Flöten- und Oboenkonzert nur 2 Takte. Die Auswahl der Transpositionsgrößen erfolgte nach den Erfahrungen mit dem Schüler-Experiment so, dass die Transposition um ± 4 Halbtöne, deren Wiedererkennung quasi zu einem Deckeneffekt geführt hatte, hier nicht mehr genutzt wurde. Um die Tonarterinnerung genauer abbilden zu können, wurden statt dessen die kleineren Transpositionen um eine große Sekunde und kleine Terz genutzt. Die 7 Versionen des „eigenen“

Mozart-Konzerts wurden in fester Reihenfolge dargeboten. Um Reihenfolgeeffekten entgegenzuwirken, wurden zwei unterschiedliche Reihenfolgen hergestellt. Zwischen jeweils zwei Versionen des Mozart-Konzerts wurde eine der 5 Versionen des Telekom-Jingles (aus dem Schüler-Experiment, vgl. Kap. 7.1) vorgespielt, um den direkten Vergleich zweier Versionen des Mozart-Konzerts zu verhindern. Dies ist wichtig, da die Vpn keine relativen Urteile (Vergleiche von Versionen) fällen, sondern jede Version mit ihrer eigenen Vorstellung vergleichen sollten. Außerdem wurde darauf geachtet, dass keine extremen Lagenwechsel vorkamen, also z.B. auf eine „+3“ Transposition keine „-3“ Transposition folgte.

Bevor den Vpn über Kopfhörer eine Version des Mozart-Konzerts oder des Telekom-Jingles vorgespielt wurde, bekamen sie einen visuellen Hinweisreiz (Melodie-Titel) vorgelegt und wurden instruiert, sich die jeweilige Melodie so genau wie möglich vor ihr inneres Ohr zu rufen (analog der Produktionsaufgabe). Dafür hatten sie 15 Sekunden Zeit. Nachdem die jeweilige Melodieversion erklungen war, sollten die Vpn auf einem Antwortblatt angeben, ob es sich ihrer Meinung nach um die Originalversion der entsprechenden Melodie handele oder ob die Version zu hoch oder zu tief war. Für ihr Urteil hatten die Vpn 5 Sekunden Zeit. Das Zeitschema aus dem Schüler-Experiment wurde nach der Durchführung eines Probe-Versuchsdurchgangs mit einer Vp, die am Hauptexperiment nicht beteiligt war, dahingehend verändert, dass den Instrumentalmusikern etwas mehr Zeit zur Vorstellung der Melodie gegeben wurde. Um den Vpn das Vergessen einer soeben gehörten Version und das Einstellen auf die nächste Melodie zu erleichtern, wurden zwischen allen Trials Störklänge (die gleichen wie in Kap. 7.1.4 beschrieben) dargeboten. Der Ablauf eines Trials der Wiedererkennungsaufgabe verlief also folgendermaßen:

15 s Vorstellungszeit für Mozart-Konzert

10 s Mozart-Konzert-Version

5 s Zeit für Reaktion (Ankreuzen auf Antwortblatt)

5 s Störgeräusche

10 s Vorstellungszeit für Telekom-Jingle

2 s Telekom-Jingle

5 s Zeit für Reaktion (Ankreuzen auf Antwortblatt)

5 s Störgeräusche (darauf folgt Mozart-Konzert nächste Version)

9.1.1.3 Herstellung der Versuchsmaterialien

Die Versionen für die Wiedererkennungsaufgabe wurden in diesem Experiment in einer einheitlichen Klangfarbe, nämlich Klavier, dargeboten. Dieses Vorgehen hat zwar den Nachteil, dass die Melodien etwas weniger authentisch klingen. Angesichts des Vorteils höherer Vergleichbarkeit der Daten verschiedener Instrumentalisten fiel die Entscheidung aber zugunsten einer stärkeren Standardisierung der Melodien. Für die Flötisten wurde eines der beiden Solokonzerte ausgewählt, da die meisten Flötisten (und alle Vpn) beide Konzerte regelmäßig

üben. Die Wahl fiel auf das Konzert in D-Dur, das nach Aussage einiger Flötisten das häufiger gespielte ist, eine Aussage, die sich insofern bestätigt hat, als mehr Flötisten in der Produktionsaufgabe dieses Konzert als das G-Dur Konzert wiedergaben.

Die Originalversionen sowie Transpositionen aller Mozart-Konzerte wurden mithilfe eines Digitalklaviers (Yamaha Clavinova CLP-950) hergestellt. Ein professioneller Pianist spielte die Originalversionen auf dem Clavinova, das diese Version auf seinem internen Speicher speicherte. Für die Herstellung der Transpositionen wurde der Transpositionsmodus des Clavinovas genutzt, der die eingespielten Melodien in beliebig großen Transpositionen nach oben oder unten wiedergeben kann. Ein Minidisc Rekorder wurde mit einem Digitalkabel an das Clavinova angeschlossen, um die Originalversion und alle Transpositionen der vier Mozart-Konzerte aufzunehmen. Diese Vorgehensweise resultiert in Versionen, die hinsichtlich des Tempos, der Lautstärke, der Klangfarbe und der Spieltechnik identisch sind und sich nur in der Tonart unterscheiden. Im Vergleich zur separaten Einspielung aller Versionen auf einem Klavier (wie von Terhardt & Mitarbeitern vorgenommen) hat dieses Vorgehen den Vorteil, dass das Clavinova in allen Tonarten gleich „künstlich“ klingt, sich also keine bei manuell gestimmten Klavieren teilweise noch rudimentär vorhandenen Tonartencharakteristika ausmachen lassen. Die Vpn sind bei ihrem Urteil, ob es sich bei einer vorgespielten Version um das Original oder eine Transposition handelt, also alleine auf die Beurteilung der absoluten Tonhöhen angewiesen (dieses Vorgehen entspricht dem von Vitouch & Gaugusch, 2000).

Aus den Melodieversionen und den Störklängen wurde mithilfe des Programms Cool Edit eine CD erstellt, auf der die Trials nach dem oben beschriebenen Zeitschema aneinander geschnitten waren. Diese CD, die die notwendigen Pausen für die Urteile bereits enthielt, konnte dann den Vpn vorgespielt werden, wodurch ein gleicher zeitlicher Ablauf für alle Vpn gesichert wurde. Da jede Melodieversion einmal dargeboten wurde, enthielt die CD 7 (Mozart-Konzert) + 6 (Telekom-Jingle) = 13 Trials. Die Bearbeitung der Wiedererkennungsaufgabe dauerte ca. 7 Minuten.

9.1.2 Ergebnisse

9.1.2.1 Produktionsaufgabe

Um zu überprüfen, ob sich die Produktionsleistungen zwischen den vier Gruppen von Instrumentalisten unterscheiden, wurden die Abweichungen einer einfachen ANOVA mit dem Gruppierungsfaktor *Instrument* (Klarinette, Fagott, Flöte, Oboe) zugeführt. Tabelle 9.1 zeigt die Gruppenmittelwerte der Abweichungen sowie die Ergebnisse der ANOVAs (inklusive Levene-Test zur Prüfung der Varianzhomogenität; über die Prüfung der Normalverteilung berichtet Kap. 6.8). Die ANOVAs ergaben keinen signifikanten Haupteffekt für *Instrument*, so dass die gemeinsame Analyse der Produktionsdaten aller Vpn gerechtfertigt erscheint.

Die Verteilung und Signifikanzprüfung der Produktionsdaten wurde in Kap. 6.8.4 dargestellt. Die Daten weisen auf eine hochsignifikante Tonarterinnerung sowohl beim Mozart-Konzert als auch bei den Orchesterstellen hin. Vergleicht man die mittleren Fehler der Instrumental-

musiker (m = 1,85, SD = 1,81 Halbtöne) mit denen der Chorsänger (m = 2,71, SD = 1,86 Halbtöne), so ist erkennbar, dass die Instrumentalmelodien nicht schlechter produziert wurden als die Vokalmelodien. Dies bestätigt Hypothese 4 b für die Produktionsaufgabe.

Tab. 9.1: Produktionsergebnisse der vier Instrumentengruppen und Signifikanzprüfung der Unterschiede

	Instrument	**N**	**Mittelwert**	**SE**	**Ergebnisse Levene-Test, ANOVA**
	Flöte	12	1,25	0,51	F (3,36) = 0,221
Produktion Mozartkonzert	Klarinette	6	2,17	0,75	p = .881
mittlere absolute	Oboe	13	1,46	0,48	F (3,36) = 0,522
Differenzen	Fagott	9	2,00	0,65	p = .670
	Flöte	12	1,67	0,51	F (3,36) = 0,529
Produktion Orchesterstelle	Klarinette	6	2,67	0,67	p = .665
mittlere absolute	Oboe	13	1,77	0,51	F (3,36) = 0,804
Differenzen	Fagott	9	2,67	0,73	p = .500

9.1.2.2 Wiedererkennungsaufgabe

Für die Analyse der Wiedererkennungsaufgabe wurde zunächst pro Vp die relative Anzahl von „zu hoch“, „richtig“ und „zu tief“ Antworten pro Version jeder Melodie bestimmt und über alle Vpn gemittelt. Dann wurde analog dem Vorgehen, das in Kap. 7.1.7 beschrieben wurde, für jede Vp und jede Melodie eine IR berechnet. Als nächstes wurde geprüft, ob die IRs normalverteilt sind. Der Kolmogorov-Smirnov-Test ergab eine signifikante Abweichung der Mozart-Konzert IRs von der Normalverteilung (Z = 1,552, p = .016), während die Telekom IRs einer Normalverteilung genügten (Z = 0,955, p = .321). Aufgrund dieses Ergebnisses wurde für die Mozart-Konzert IRs ein nichtparametrisches Verfahren (Kruskal-Wallis-Test) angewandt, während die Telekom IRs mit einer einfaktoriellen ANOVA (Gruppierungsfaktor *Instrument*) analysiert wurden, um herauszufinden, ob es signifikante Leistungsunterschiede zwischen den Instrumentengruppen gibt (vgl. Tab. 9.2). Es zeigte sich ein signifikanter Haupteffekt für *Instrument* bei den IRs des Mozart-Konzerts. Dass dieser Effekt stückspezifisch ist, zeigt sich daran, dass bei den IRs des Telekom-Jingles keine signifikanten Gruppenunterschiede auftraten. Für die weitere Signifikanzprüfung wurden die IRs daher beim Mozart-Konzert für jede Instrumentengruppe getrennt berechnet und beim Telekom-Jingle über alle Vpn zusammen.

Abbildung 9.1 stellt die Häufigkeiten, mit denen die Vpn bei den sieben Versionen des Mozartkonzerts sowie bei den fünf Versionen des Telekom-Werbejingles die drei möglichen Antworten („zu hoch“, „richtig“, „zu tief“) gegeben haben, gegenüber. Die Darstellung erfolgt beim Mozart-Konzert getrennt nach Instrumentengruppen (wobei Klarinette und Oboe aufgrund vergleichbarer IRs zusammengefasst wurden, da die Gruppengrößen sonst sehr klein werden), beim Telekom-Jingle für alle Vpn zusammengefasst.

Tab. 9.2: Wiedererkennungsleistungen in den vier Instrumentengruppen und Signifikanzprüfung der Unterschiede

	Instrument	N	Mittelwert	SE	Ergebnisse Kruskal-Wallis-Test, Levene-Test, ANOVA
IR Mozart-Konzert	Flöte	12	0,57	0,13	chi² = 9,897
	Klarinette	6	0,23	0,29	df = 3
	Oboe	13	0,22	0,17	p = .019
	Fagott	9	- 0,28	0,04	
IR Telekom-Jingle	Flöte	12	0,35	0,13	F (3,36) = 1,950
	Klarinette	6	0,54	0,15	p = .139
	Oboe	13	0,21	0,16	F (3,36) = 0,976
	Fagott	9	0,50	0,13	p = .415

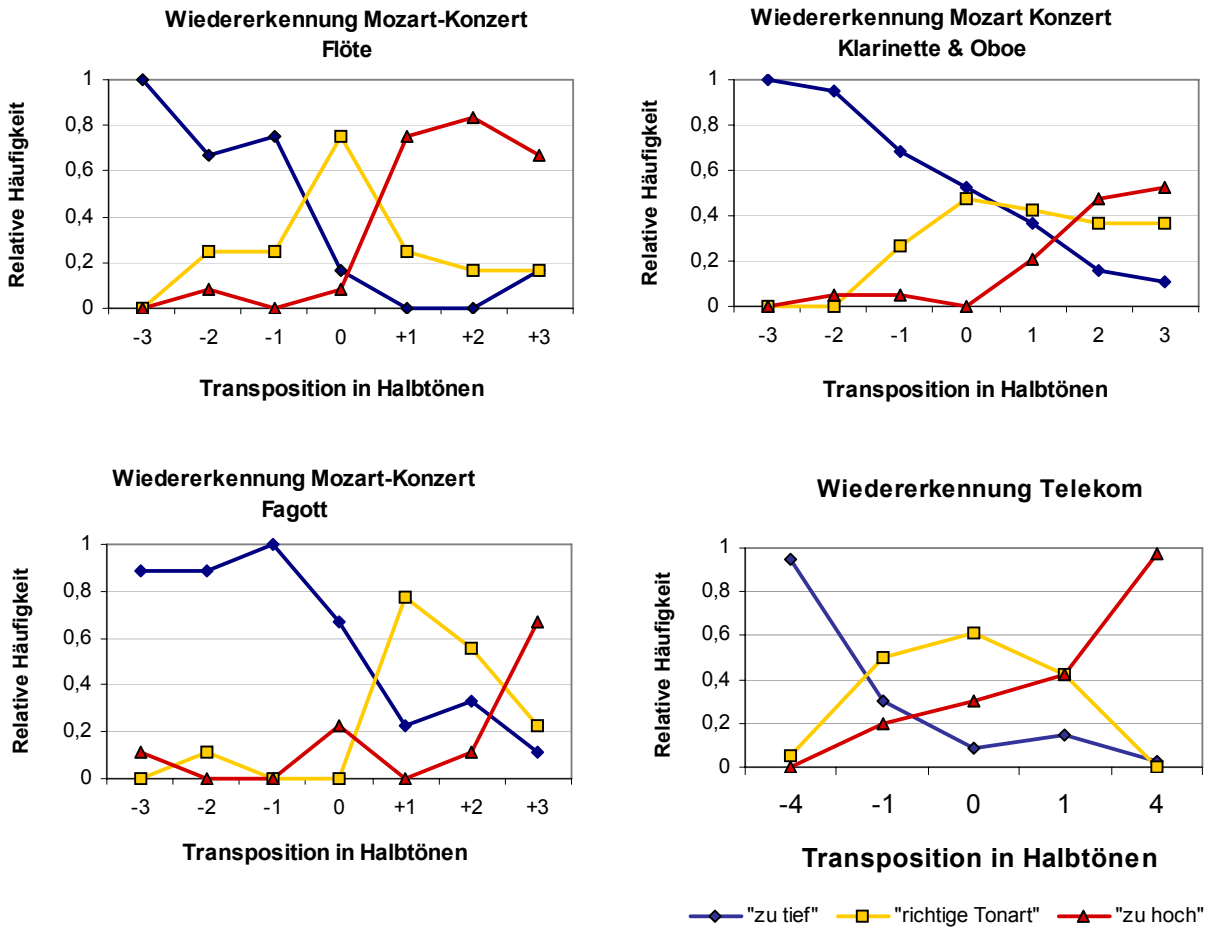

Abb. 9.1: Wiedererkennung des Mozart-Konzerts (getrennt nach Instrumentengruppen) und Telekom-Jingles (alle Vpn). Auf der Kategorienachse sind die 7 bzw. 5 Versionen der Melodien abgetragen. Die y-Achse zeigt die relative Häufigkeit der drei möglichen Antworten „zu hoch“, „richtig“ und „zu tief“.

Aus der Abbildung wird ersichtlich, dass bei den Flötisten eine sehr gute Wiedererkennung vorlag, denn schon bei Halbtontranspositionen wird deutlich häufiger die Antwort „zu hoch“ / „zu tief“ gegeben als die Antwort „richtig“. Bei Klarinette und Oboe zeigt sich eine gute Wiedererkennung der Transpositionen nach unten, wohingegen Transpositionen nach oben nicht gut erkannt werden: Die Linien liegen viel dichter beieinander und selbst die Transposition um eine kleine Terz nach oben wird nur in ca. 50 % der Fälle als „zu hoch“ beurteilt, dagegen in fast 40 % der Fälle als „richtig“. Beim Fagott zeigt sich eine gute Wiedererkennung der Transpositionen nach unten und eine sehr schlechte Wiedererkennung der Transpositionen nach oben. Kein einziger Fagottist beurteilte die Originalversion als „richtig“, statt dessen wurde die Version „+1“ am häufigsten als „richtig“ beurteilt und nur die Transposition um eine kleine Terz nach oben häufiger „zu hoch“ als „richtig“. Beim Telekom-Jingle wurde die Originalversion am häufigsten als „richtig“ beurteilt, jedoch wurden die Halbtontranspositionen häufig mit dem Original verwechselt und nur die Terztranspositionen sicher wiedererkannt. Die Abbildungen zeigen, dass von allen Vpn außer den Flötisten eine Tonart als „richtig“ empfunden wird, die eher etwas höher liegt als die Originaltonart.

Um die Wiedererkennungsleistung auf Signifikanz zu überprüfen, wurde ein Histogramm der IRs erstellt. Abb. 9.2 zeigt das Histogramm der gemittelten IRs, getrennt nach den beiden Melodien und für das Mozart-Konzert getrennt nach Instrumentengruppen. Über alle Vpn gemittelt fielen für das Mozart-Konzert 55 % der Fälle in die Kategorie 0-10 % (wobei diese Kategorie, in Anlehnung an Terhardt & Seewann, 1983, auch die negativen IRs enthält) und sind somit als zufällige Leistung zu bewerten. Betrachtet man die Instrumentengruppen einzeln, so zeigt sich, dass die Leistung aller Fagottisten als zufällige Leistung zu betrachten ist, während diese Kategorie bei den Oboisten 54 %, bei den Klarinettisten 50 % und bei den Flötisten nur 25 % betrug. Für den Telekom-Jingle umfasst diese Kategorie ebenfalls 25 % der Fälle, so dass hier die Leistung von 75 % der Vpn als überzufällig zu bewerten ist. Auf eine weitere Signifikanzprüfung, wie sie für die Schüler-Stichprobe erfolgte (vgl. Kap. 7.2.2), wurde hier verzichtet, da bei einer nach Instrumentengruppen getrennten Analyse die Gruppengrößen zu klein werden, um einen t-Test durchführen zu können.

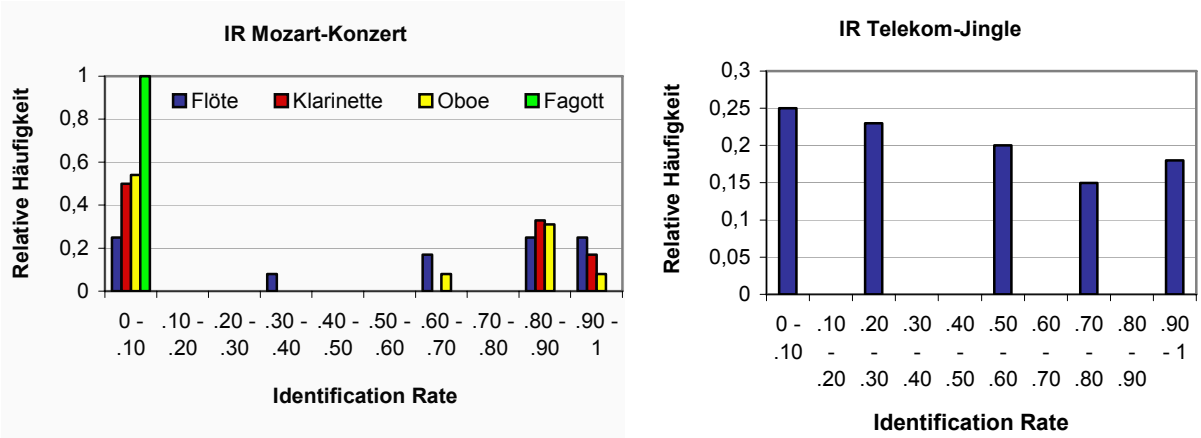

Abb. 9.2: Über alle Vpn gemittelte IRs (relative Erkennungshäufigkeit, korrigiert um falsch positive Urteile), getrennt nach den beiden Melodien und für das Mozart-Konzert getrennt nach den 4 Instrumentengruppen.

Zusammengefasst bedeuten diese Ergebnisse, dass die Leistung bei der Wiedererkennung des Mozart-Konzerts zwar zu überzufälligen Ergebnissen bei 45 % der Vpn führte. Dieses Ergebnis steht aber erstens in deutlichem Gegensatz zu dem der Produktionsaufgabe, denn dort hatten 67,5 % der Vpn die Tonart des Mozart-Konzerts mit höchstens einem Halbton Abweichung getroffen. Zweitens ist die Wiedererkennung hier deutlich schlechter als die im Schüler-Experiment (Kap. 7.2.2), wo 75 bzw. 78 % der Vpn überzufällige Leistungen beim Abruf der Vokalmelodien erbrachten. Hypothese 4 b kann demnach für die Wiedererkennungsaufgabe nicht bestätigt werden.

9.1.2.3 Tonarterinnerung & Absolutes Gehör

Im Test auf absolutes Gehör wurden durchschnittlich 2,93 (SD = 3,03) Töne richtig benannt. Bei 5,08 (SD = 3,60) Tönen betrug der Fehler durchschnittlich maximal einen Halbton. Um zu überprüfen, ob die Tonbenennungsleistung dieser Stichprobe überzufällig ist, wurde die Anzahl korrekt benannter Töne im Einstichproben t-Test gegen den zufällig zu erwartenden Wert 1 (= 12 x 1/12) getestet. Der t-Test ergab ein signifikantes Ergebnis (t (39) = 4,014, p < .001), so dass man insgesamt von einer überzufälligen Leistung in der Tonbenennung ausgehen muss. Wie Abb. 9.3 zeigt, liegt ein Großteil der Stichprobe im Leistungsbereich 0 bis 4 richtig benannte Töne, aber es gibt auch eine Reihe von Vpn, die über 50 % der Töne richtig benennen konnten und somit in den Leistungsbereich von „echten“ Absoluthörern fallen.

Um innerhalb dieser Stichprobe, die eine größere Bandbreite an Tonbenennungsleistungen umfasst als alle anderen getesteten Stichproben, einen Einfluss der Fähigkeit, Töne zu benennen, auf die Tonarterinnerung zu überprüfen, wurden die absoluten Differenzen aus der Produktionsaufgabe sowie die IRs aus der Wiedererkennungsaufgabe mit der Tonbenennungsleistung korreliert. Die Ergebnisse sind in Tabelle 9.3 dargestellt.

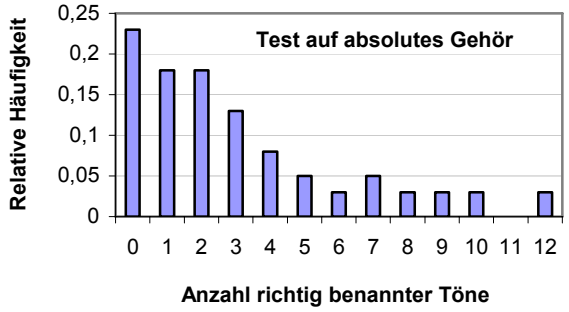

Abb. 9.3: Relative Häufigkeit korrekt benannter Töne über 40 Vpn.

Die Leistung in der Tonbenennungsaufgabe korrelierte signifikant mit der in der Produktionsaufgabe sowie in der Wiedererkennungsaufgabe. Der Zusammenhang zwischen Produktionsgenauigkeit und Tonbenennungsleistung innerhalb dieser Stichprobe ist deutlich größer als der in der Gesamt-Stichprobe gefundene Zusammenhang (vgl. Kap. 6.11.1). Außerdem findet sich ein tendenzieller Zusammenhang zwischen der Wiedererkennung des Telekom-Jingles und der Tonbenennung. Um zu überprüfen, ob die in dieser Stichprobe erzielten her-

vorragenden Leistungen in der Produktionsaufgabe allein durch die höhere Anzahl an Absoluthörern bzw. die bessere Fähigkeit der Tonbenennung erklärbar sind, wurden die durchschnittlichen Fehlergrößen bei der Produktion des Mozartkonzerts (m = 1,63) sowie der Orchesterstelle (m = 2,08) mit der durchschnittlichen Fehlergröße im Test auf absolutes Gehör (m = 2,43) verglichen. Ein t-Test für abhängige Stichproben ergab einen signifikanten Unterschied zwischen den Fehlergrößen von Tonbenennung und Mozartkonzert (r = 0,078, p = .634; t (39) = -2,300, p = .027), aber nicht von Tonbenennung und Orchesterstelle (r = 0,116, p = .475; t (39) = -1,010, p = .319). Somit haben die Vpn die Anfangstöne des Mozartkonzerts, aber nicht die der Orchesterstelle signifikant besser getroffen als die Einzeltöne in der Tonbenennungsaufgabe.

Tab. 9.3: Korrelationen zwischen den Leistungen in der Produktions-, Wiedererkennungs- und Tonbenennungsaufgabe.

		Trefferquote absolutes Gehör	Trefferquote absolutes Gehör mit Halbtontoleranz
Produktion Mozartkonzert	Pearson Correlation	**-,331**	***-,302***
	Sig. (2-tailed)	,037	,058
	N	40	40
IR Mozartkonzert	Pearson Correlation	**,446**	**,534**
	Sig. (2-tailed)	,004	,000
	N	40	40
Produktion Orchesterstelle	Pearson Correlation	-,149	-,250
	Sig. (2-tailed)	,359	,120
	N	40	40
IR Telekom	Pearson Correlation	***,274***	,258
	Sig. (2-tailed)	,087	,108
	N	40	40

Signifikante Korrelationen auf dem Niveau α =5 % sind fett gedruckt, Tendenzen (α = 10 %) sind kursiv.

Um weiterhin zu überprüfen, ob sich musikalische Expertise, hier gemessen durch die Dauer der Instrumentalausbildung und Orchestererfahrung, auf die Tonarterinnerung auswirkt, wurden zwischen diesen Variablen weitere Korrelationen berechnet, die in Tab. 9.4 dargestellt sind. Es ergaben sich innerhalb dieser Stichprobe keinerlei Zusammenhänge zwischen den Leistungen in Produktions-, Wiedererkennungs- und Tonbenennungsaufgabe und der musikalischen Ausbildung.

Tab. 9.4: Korrelationen zwischen den Leistungen in der Produktions-, Wiedererkennungs- und Tonbenennungsaufgabe und der musikalischen Ausbildung.

		Instrumentalausbildung in Jahren	Beginn des Instrumentalunterrichts	Orchestererfahrung in Jahren
Produktion Mozartkonzert	Pearson Correlation	,074	-,178	,112
	Sig. (2-tailed)	,651	,272	,491
	N	40	40	40
IR Mozartkonzert	Pearson Correlation	,082	-,123	-,230
	Sig. (2-tailed)	,616	,448	,153
	N	40	40	40
Produktion Orchesterstelle	Pearson Correlation	,042	-,199	,099
	Sig. (2-tailed)	,797	,218	,545
	N	40	40	40
IR Telekom	Pearson Correlation	,099	,116	-,021
	Sig. (2-tailed)	,542	,475	,898
	N	40	40	40
Absolutes Gehör Treffer-quote	Pearson Correlation	,131	-,168	-,170
	Sig. (2-tailed)	,421	,301	,294
	N	40	40	40
Absolutes Gehör Trefferquote mit Halbtontoleranz	Pearson Correlation	,145	-,133	-,197
	Sig. (2-tailed)	,372	,413	,223
	N	40	40	40

9.1.2.4 Erinnerungsstrategien

Nach dem Versuch wurden die Vpn nach den Strategien befragt, die sie bei der Bewältigung der Produktions- und Wiedererkennungsaufgabe nutzten, um sich die jeweilige Melodie vor ihr inneres Ohr zu rufen. Sie wurden gefragt, ob sie

- versuchten, sich die Melodie klanglich vorzustellen,
- versuchten, sich den Notentext der Melodie vorzustellen,
- sich ein bestimmtes Spielgefühl vorstellten,
- sich an eine bestimmte Unterrichtssituation erinnerten,
- und ob sie sich bewusst an eine bestimmte Tonart erinnert haben.

Tab. 9.5 stellt die Anzahl an Vpn, die angaben, die jeweilige Strategie genutzt zu haben (es waren Mehrfachnennungen möglich) gegenüber. Außerdem sind die jeweiligen Gruppenmittelwerte bezüglich der Differenzen aus der Produktionsaufgabe (hier wurden die Differenzen von Mozart-Konzert und Orchesterstelle gemittelt) sowie der IRs aus der Wiedererkennungsaufgabe (nur für Mozart-Konzert) gegenübergestellt. Die Unterschiede bezüglich der Produktionsgenauigkeit wurden mit einem t-Test für unabhängige Stichproben verglichen, nachdem mit dem Levene-Test die Varianzhomogenität geprüft wurde. Die Unterschiede bezüglich der nicht normalverteilten IRs wurden mit dem U-Test verglichen.

Tab. 9.5: Nutzung verschiedener Erinnerungsstrategien und Signifikanzprüfung der Leistungsunterschiede

Erinnerungs-Strategie	N	M (SE) Produktion	M (SE) Wieder-erkennung	Levene-Test Ergebnisse Produktion	Ergebnisse t-Test Produktion / U-Test Wiedererkennung
Klang				F = 2,789	t (38) = - 0,437, p = .664
ja	33	1,80 (0,24)	0,24 (0,10)	p = .103	
nein	7	2,07 (0,76)	0,07 (0,20)		U = 115, p = .986
Notentext				F = 0,138	t (38) = - 0,570, p = .572
ja	16	1,69 (0,34)	0,22 (0,15)	p = .712	
nein	24	1,96 (0,31)	0,21 (0,12)		U = 176,5, p = .665
Spielgefühl				F = 0,169	t (38) = - 1,159, p = .254
ja	18	1,56 (0,34)	0,17 (0,13)	p = .684	
nein	22	2,09 (0,31)	0,25 (0,13)		U = 154,5, p = .231
Unterrichts-Situation				F = 0,196	t (38) = - 0,409, p = .685
ja	7	1,64 (0,56)	0,07 (0,20)	p = .660	
nein	33	1,89 (0,26)	0,25 (0,10)		U = 104, p = .678
Tonart				F = 1,403	t (38) = - 0,204, p = .840
ja	11	1,77 (0,49)	0,26 (0,21)	p = .244	
nein	29	1,88 (0,26)	0,20 (0,10)		U = 150, p = .771

Gegenüberstellung der Anzahl an Vpn (n = 40), die angaben, eine bestimmte Erinnerungsstrategie genutzt zu haben, sowie mittlere Differenzen aus der Produktionsaufgabe (in Halbtönen) und mittlere IRs aus der Wiedererkennungsaufgabe für Strategie-Nutzer und -Nichtnutzer im Vergleich, inklusive Standardfehler (SE), sowie Ergebnisse der Levene-Tests und t-Tests für unabhängige Stichproben für die Produktionsaufgabe und Ergebnisse der U-Tests für die Wiedererkennungsaufgabe.

Es zeigt sich, dass fast alle Vpn versucht haben, sich an den Klang der jeweiligen Melodie zu erinnern, während weniger als die Hälfte sich den Notentext vorstellte. Mehr als die Hälfte der Vpn nutzte als zusätzliche Erinnerungsstrategie die Erinnerung an ein bestimmtes Spielgefühl, während die Erinnerung an die Unterrichtssituation für nur wenige Vpn eine nutzbare Strategie darstellte. Nur etwa ein Viertel der Vpn stellte sich die Melodien explizit in einer bestimmten Tonart vor. Vergleicht man für jede Erinnerungsstrategie die Produktions- und Wiedererkennungsleistungen der „Nutzer“ und „Nicht-Nutzer“, so zeigen sich in der Produktionsaufgabe immer etwas kleinere Abweichungen bei Strategie-„Nutzern“ als bei „Nicht-Nutzern“. In der Wiedererkennungsaufgabe ist dies nicht immer der Fall: Hier sind die mittleren IRs *größer* bei „Nutzern“ der Strategien Klang, Notentext und Tonart, jedoch *kleiner* bei „Nutzern“ der Strategien Spielgefühl und Unterrichtssituation. Alle diese Unterschiede sind jedoch nicht signifikant und somit nicht weiter interpretierbar.

9.1.3 Diskussion

Mit dem beschriebenen Experiment sollte geprüft werden, inwieweit das Vorhandensein motorischer Gedächtnisinhalte bezüglich eines Sing- und Spielgefühls notwendig ist für die genaue Produktion und Wiedererkennung häufig geübter Instrumentalwerke. Um diese Frage

zu untersuchen, wurden Spieler von Holzblasinstrumenten sowohl mit einer Produktions- als auch mit einer Wiedererkennungsaufgabe untersucht und die Genauigkeit der Tonarterinnerung geprüft. Holzbläser sind insofern geeignete Probanden für ein solches Experiment, weil sie während des Spielens nicht singen können und somit kaum eine singmotorische Erinnerung an geübte Melodien haben können.

In der Produktionsaufgabe erzielten die Instrumentalmusiker sowohl beim Abruf des Solo-Konzerts von W.A. Mozart für das eigene Instrument als auch beim Abruf einer häufig geübten Orchesterstelle überzufällige Leistungen: 30 % der Vpn trafen die Originaltonart des Solo-Konzerts genau, und bei zwei Dritteln der Vpn betrug die produzierte Abweichung von der Originaltonart maximal einen Halbton. Dagegen war die Leistung in der Wiedererkennungsaufgabe nur bei knapp der Hälfte der Vpn überzufällig gut, während viele Vpn nur die zu tiefen Versionen als Transpositionen erkannten, nicht aber die zu hohen.

Die Ergebnisse aus der Produktionsaufgabe zeigen, dass es Instrumentalmusikern gelingt, sich an die Tonarten oft geübter Instrumentalwerke zu erinnern. Da ihnen weder spielmotorische noch singmotorische Gedächtnisinhalte zur Verfügung standen, spricht dies für eine gute Erinnerungsfähigkeit an Tonarten durch das Abrufen klanglicher Informationen aus dem LZG. Bei dieser Interpretation ist aber das in Kap. 4.4.5 erwähnte Ergebnis der Ähnlichkeit von Vorstellung und Ausführung von Bewegungen zu beachten. Somit könnte den Instrumentalmusikern auch die Vorstellung ihres Instruments, die bei ca. der Hälfte der Vpn vorlag, spielmotorische Informationen geliefert haben. Jedoch sprechen die Befunde eindeutig gegen eine Interpretation des Tonartgedächtnisses als bloße Erinnerung an ein Singgefühl.

Der relativ geringe Zusammenhang zwischen Produktions- und Wiedererkennungsleistungen (die Korrelation beträgt nur $r = -.33$, $p = .036$) ist zum Teil durch die in der Wiedererkennungsaufgabe verwendete Klangfarbe (Digitalklavier) zu erklären, die diese Aufgabe erschwert. Anders als z.B. beim Telekom-Jingle gab es hier keine Version, die dem Original identisch ist, denn auch bei der Originalversion musste vom Instrumentalklang abstrahiert und die „reine“ Tonhöheninformation verglichen werden. Dadurch ähnelt die Wiedererkennungsaufgabe dem Test auf absolutes Gehör, mit dem Unterschied, dass mehrere Töne vorliegen, also quasi eine absolute Ton*art*erkennung gefordert wird. Die Ähnlichkeit der beiden Aufgaben wird durch die Tatsache, dass Wiedererkennungs- und Tonbenennungsleistung höher korrelieren als Produktions- und Tonbenennungsleistung, unterstrichen. Die Verwendung des Klaviertimbres könnte auch der Grund dafür sein, dass die Wiedererkennungsleistung der Instrumentalisten wesentlich schlechter war als die der Schüler, während die Produktionsleistung der Instrumentalisten nicht schlechter war als die der Chorsänger.

Außerdem ist der geringe Zusammenhang zwischen Produktions- und Wiedererkennungsleistungen durch gegenläufige Fehlertendenzen zu erklären: Während bei der Produktion (wie bei allen anderen Produktions-Experimenten auch) Fehler nach unten häufiger auftreten als Fehler nach oben, werden in der Wiedererkennung Transpositionen nach unten deutlich besser als „zu tief“ erkannt als Transpositionen nach oben als „zu hoch“. Eine Erklärung hierfür könnte der im Vergleich zum Kammerton a = 440 Hz etwas höhere Stimmton in vielen

deutschen Orchestern sein, allerdings beträgt die Höherstimmung nur einige Hertz und nicht einen Halbton. Ein weiterer Grund hierfür könnte in der Anspannung bei der von vielen Vpn im Vergleich zur Produktion als deutlich schwierigeren Aufgabe empfundenen Wiedererkennungsaufgabe liegen, denn hohe Anspannung führt z.B. bei Chorsängern zu einem Ansteigen der Tonhöhen (persönliche Mitteilung durch verschiedene Dirigenten).

Ein unerwarteter Effekt des Experiments war der Leistungsunterschied zwischen den vier Instrumentengruppen, der in der Produktionsaufgabe zwar nicht signifikant wurde, jedoch lagen auch dort Flötisten und Oboisten ca. einen Viertelton näher an der Originaltonart des Mozartkonzerts und ca. einen Halbton näher an der Originaltonart der Orchesterstelle als Klarinettisten und Fagottisten. In der Wiedererkennungsaufgabe wurden die Leistungsunterschiede signifikant, wobei vor allem die nahezu perfekte Leistung der Flötisten und die zufällige Leistung aller 9 Fagottisten auffallen. Da auch bei den Fagottisten in der Produktionsaufgabe durchaus eine überzufällige Tonarterinnerung vorlag, ist die schlechte Wiedererkennungsleistung nicht allein durch eine ungenauere Tonartrepräsentation zu erklären. Eine mögliche Erklärung könnte in der Prägnanz der Klangfarbe liegen, also darin, wie gut eine Vp ihre eigene Instrumentalstimme zu hören gewohnt ist. Die Flöte ist im Orchesterklang meist sehr gut heraushörbar, das Fagott dagegen geht im Orchester und auch als Solostimme zuweilen unter (vermutlich auch aufgrund der tieferen Lage). Gegen die Interpretation der schlechteren Wiedererkennungsleistung von Fagottisten und Klarinettisten als allgemein geringere Fähigkeit zur Tonarterinnerung spricht die Tatsache, dass bei der Wiedererkennung des Telekom-Jingles diese beiden Instrumentengruppen bessere Leistungen erzielen als Flötisten und Oboisten.

Insgesamt zeigen die Ergebnisse der Wiedererkennungsaufgabe, dass sich auch mit dieser Methode absolutes Tonartgedächtnis abbilden lässt, dass sich aber außer der Fähigkeit, sich an die Tonart eines Stücks zu erinnern, auch noch andere Faktoren auf die Wiedererkennungsleistung auswirken. Einflussfaktoren sind erstens die Klangfarbe, in der die Vergleichsmelodien dargeboten werden, zweitens die Klangfarbe des eigenen Instruments und drittens die Fähigkeit, einzelne Töne zu benennen.

Ein Einwand gegen die Interpretation der Produktionsergebnisse als absolutes Tongedächtnis ist der, dass in dieser Stichprobe von professionellen Musikern eine höhere Rate an Absoluthörern und eine durchschnittlich bessere Fähigkeit, einzelne Töne zu benennen, vorlag als in allen anderen untersuchten Stichproben. Möglicherweise rufen die Instrumentalmusiker lediglich den Namen des Anfangstons aus dem LZG ab und versuchen dann, diesen Ton möglichst genau zu produzieren. In der Tat ließen sich innerhalb dieser Stichprobe höhere Zusammenhänge zwischen Produktions- und Tonbenennungsleistungen nachweisen als in der Gesamt-Analyse (vgl. Kap. 6.11.1). Allerdings widersprechen dieser Interpretation sowohl die Angaben der Vpn, von denen nur ein Viertel angab, sich an eine konkrete Tonart zu erinnern, als auch die Tatsache, dass die Anfangstöne des Mozartkonzerts signifikant genauer produziert als Einzeltöne benannt werden konnten. Dies spricht dafür, dass die beiden bei der Tonartproduktion hauptsächlich genutzten Strategien, die klangliche und spielmotori-

sche Erinnerung, wichtige Informationen zur Bewältigung dieser Aufgabe geliefert haben. Der Abruf von Einzeltönen ist in dieser musikalisch hochtrainierten Stichprobe als eine weitere, aber nicht die einzige, Strategie beim Tonartabruf zu betrachten.

Da die Befragungsdaten zeigen, dass es vorteilhaft zu sein scheint, mehrere Strategien beim Tonartabruf zu verwenden, wurde in einem weiteren Experiment eine Manipulation der Strategien vorgenommen, um den Beitrag der Strategien „inneres Singen“ und „inneres Hören“ zur Genauigkeit des Tonartabrufs zu untersuchen.

9.2 Produktion mit Interferenzaufgaben bei Chorsängern und Laien

9.2.1 Methode

9.2.1.1 Stichprobe

Um u.a. den Einfluss der musikalischen Expertise auf den Tonartabruf mithilfe von Interferenzaufgaben zu untersuchen, wurden zwei Stichproben untersucht.

Die erste Stichprobe umfasste 30 musikalische Laien (15 männlich, 15 weiblich). Sie waren durchschnittlich 35,43 (SD = 6,45) Jahre alt und hatten 2,10 (SD = 3,32) Jahre Instrumentalunterricht genossen. Nur eine Vp verfügte über Erfahrungen mit Gesangs- oder Gehörbildungsunterricht, und nur 2 Vpn haben jemals im Chor gesungen. Die Vpn wurden aus dem Bekanntenkreis der Versuchsleiterin rekrutiert. Die Versuchsteilnahme erfolgte freiwillig und ohne Bezahlung.

Die zweite Stichprobe umfasste 25 semiprofessionelle Chorsänger (15 weiblich, 10 männlich). Sie waren durchschnittlich 32,20 (SD = 9,38) Jahre alt, hatten 11,36 (SD = 5,49) Jahre Instrumentalerfahrung, 17,66 (SD = 10,12) Jahre Chorerfahrung, 5,68 (SD = 4,37) Jahre Gesangsunterricht und 4,16 (SD = 7,06) Jahre Gehörbildungserfahrung. Die Chorsänger wurden aus zwei Chören rekrutiert, die jeweils kurz nach einem Konzert untersucht wurden. Ein Teil der Chorsänger stammte aus dem Kammerchor Ensemberlino Vocale und ein Teil aus dem Ensemble Vocal de Bardou, die hinsichtlich der musikalischen Qualität vergleichbar sind. Die Vpn wurden aus dem Bekanntenkreis der Autorin rekrutiert und nahmen am Experiment freiwillig und ohne Bezahlung teil.

9.2.1.2 Produzierte Melodien

Die Laien produzierten drei ihrer aktuell häufig gehörten Lieblingslieder aus dem Bereich der Unterhaltungsmusik. Sie wurden angewiesen, die Lieder auszuwählen, an die sie sich am besten erinnern können, dabei wurden ihnen keine Lieder vorgegeben (vgl. 6.2.2).

Die Chorsänger produzierten drei Lieblingslieder aus dem Repertoire der gerade gesungenen Konzerte (vgl. 6.4.2). Auch die Chorsänger wurden angewiesen, die Lieder auszuwählen, an die sie sich am besten erinnern können.

9.2.1.3 Versuchsdurchführung

Das experimentelle Vorgehen war für Laien und Chorsänger identisch. Dies wurde durch standardisierte Instruktionen und einen computergesteuerten Versuchsablauf sichergestellt. Zunächst wurden die Vpn angewiesen, ihre drei Lieblingslieder bzw. Lieblingschorlieder zu nennen und in eine Rangfolge (liebstes, zweit- und drittliebstes Lied) zu bringen. Die Stufen individueller Beliebtheit wurden gleichmäßig auf die drei Versuchsbedingungen verteilt.

Als nächstes wurden die Vpn angewiesen, sich ihre Lieblingslieder so genau wie möglich vor ihr inneres Ohr zu rufen. Diese Vorstellungsphase erfolgte unter drei experimentellen Bedingungen. In Bedingung 1 waren die Vpn ungestört. In Bedingung 2 wurde ihre singmotorische Vorstellung (bzw. die Rehearsalkomponente der phonologischen Schleife) durch eine Sekundäraufgabe beansprucht, in Bedingung 3 wurde ihre auditive Vorstellung (bzw. die Speicherkomponente der phonologischen Schleife) durch eine Sekundäraufgabe beansprucht.

Die Sekundäraufgabe für die singmotorische Vorstellung bestand darin, dass die Vpn durch einen Strohhalm ein- und ausatmen sollten. Dadurch sollte verhindert werden, dass die Vpn während der Vorstellungsgenerierung innerlich summen können, um die richtigen Töne zu finden. Bei der Auswahl der Sekundäraufgabe für die singmotorische Vorstellung mussten zwei Voraussetzungen erfüllt sein: Zum einen musste die Aufgabe leise sein, damit nicht gleichzeitig die auditive Vorstellung gestört wird, und zum anderen musste sie das innere Summen oder Singen stören und nicht nur die innere Artikulation. Viele Aufgaben aus Experimenten mit Sprache stören lediglich die innere Artikulation (z.B. lautloses „ta ta ta“ sprechen, Kaugummi kauen, einen Bleistift quer in den Mund nehmen, vgl. Smith, Reisberg & Wilson, 1992), ermöglichen aber ein inneres Summen und können somit die singmotorische Vorstellung für Tonhöhen nicht wirkungsvoll stören. Während die Vpn durch den Strohhalm atmeten, wurden sie durch die Versuchsleiterin genau beobachtet, damit sie nicht trotz des Strohhalms zu summen versuchten.

Die Sekundäraufgabe für die auditive Vorstellung bestand darin, dass die Vpn über einen Kopfhörer beidohrig einen sehr lauten Klang anhören mussten, der ihr inneres Hören quasi übertönen sollte. Bei der Auswahl dieser Sekundäraufgabe mussten wieder zwei Voraussetzungen erfüllt sein: Zum einen musste der Interferenzklang vorgespielt und nicht von den Vpn selbst gesungen werden, um nicht gleichzeitig die singmotorische Vorstellung zu beeinflussen. Zum anderen durfte der Klang kein tonales Zentrum haben, um die Vpn nicht in die Richtung einer bestimmten Tonart zu beeinflussen. Viele Sekundäraufgaben aus Experimenten mit Sprache waren hier außerdem nicht verwendbar, da sie nur sprachliche Informationen darbieten (z.B. gesprochene Wörter) und bislang nicht klar ist, ob Musik und Sprache im gleichen System der phonologischen Schleife verarbeitet werden. Um das innere Hören wirkungsvoll zu stören, musste daher musikalische Information dargeboten werden. Der hier verwendete Klang beinhaltete ein Orgel-Cluster, aus dem kein tonales Zentrum herauszuhören war. Damit die Vpn sich nicht an diesen Klang gewöhnen und ihn dann ignorieren konnten, wurden in kurzen Abständen kurze Pausen (einige 100 ms lang) eingefügt, so dass der Klang einen knatternden Charakter erhielt, der an einen schlechten Radioempfang erinnert.

Der Ablauf des Experiments verlief folgendermaßen: Die Vpn wurden zunächst mit den drei Bedingungen vertraut gemacht und erhielten die Gelegenheit, das Atmen durch den Strohhalm zu üben. Dann wurde ihnen der Versuchsablauf erklärt. Jede Vp wurde in allen drei Bedingungen untersucht, wobei die Reihenfolge der Bedingungen und die Zuordnung der drei Lieder auf die drei Bedingungen zufällig erfolgte und die Vpn folglich in den ersten beiden Bedingungen nicht wussten, welches Lied und welche Bedingung als nächstes kommen würden. Sobald auf dem Bildschirm des Computers der Titel des vorzustellenden Liedes sowie die Bedingung (anhand eines Piktogramms) erschienen, sollten die Vpn sich die erste Phrase dieses Liedes vorstellen. Sobald sie sich die erste Phrase vorgestellt hatten, sollten sie eine beliebige Taste auf der Computer-Tastatur drücken. Die Zeit zwischen Erscheinen von Liedtitel + Bedingung und Tastendruck wurde als Vorstellungszeit registriert. Die Vpn wurden angewiesen, sich nur eine Phrase des Liedes vorzustellen (eine Phrase wurde als Melodieausschnitt definiert, den man in einem Atemzug singen kann) und die Taste erst zu drücken, wenn sie innerlich die ganze Phrase gehört hatten. Sobald die Taste gedrückt worden war, erschien auf dem Bildschirm die Anweisung, die vorgestellte Phrase zu singen. Die Vpn wurden angewiesen, sofort nach Tastendruck mit der Produktion zu beginnen, die Phrase genau in der Lage zu produzieren, die sie sich vorgestellt hatten und keinesfalls nach dem Tastendruck noch über die Phrase nachzudenken. Zur Kontrolle wurde die Reaktionszeit, definiert als die Zeit zwischen Tastendruck und Beginn des Singens, ausgewertet. Die produzierten Liedanfänge wurden gespeichert und hinsichtlich der Anfangstöne analysiert.

Die Sekundäraufgaben wurden nur während der Vorstellungszeit und nicht während der Produktionsaufgabe durchgeführt. Das bedeutet, dass der Klang aus der auditiven Störbedingung mit Erscheinen von Liedtitel + Bedingung begann und mit dem Tastendruck endete. Den Strohhalm hielten die Vpn während aller Bedingungen in der Hand und wurden angewiesen, in der entsprechenden Bedingung sofort den Strohhalm zum Mund zu führen und durch ihn zu atmen zu beginnen, bevor sie sich die Phrase vorzustellen begannen. Zum Singen sollte der Strohhalm dann abgelegt werden.

Die Vpn erhielten die Gelegenheit, diesen Versuchsablauf anhand eines vollständigen Versuchsdurchlaufs zu üben. Außer der Tatsache, dass statt ihrer drei Lieblingslieder die Titel von drei Kinderliedern erschienen, entsprach der Probedurchlauf genau dem Ablauf des eigentlichen Experiments.

Nach der Durchführung des Experiments in allen drei Bedingungen wurde der in Kap. 6.10.1 beschriebene Screening-Test auf absolutes Gehör durchgeführt. Gab eine Vp der Laien-Gruppe an, keine Tonnamen zu kennen, so wurde für diese Vp von der Durchführung des Screening-Tests abgesehen. Zum Schluss des Experiments füllten die Vpn einen Fragebogen zu ihrem musikalischen Hintergrund und ihren Hörgewohnheiten bezüglich der produzierten Lieder aus. Das Experiment wurde in Matlab (Version 6.5) unter Nutzung der Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) programmiert.

9.2.2 Spezifische Hypothesen für die beiden Stichproben

Da es in diesem Experiment um die Frage ging, ob für die Vorstellung von Melodien singmotorische und auditive Gedächtnisinhalte benötigt werden, wurden die Sekundäraufgaben auf die Vorstellungsphase begrenzt. Während der Externalisierung der Vorstellung durch Produktion wurden dagegen keine Sekundäraufgaben eingesetzt. Dieses Vorgehen hat den Vorteil, dass man die auftretenden Effekte eindeutig der Phase des Abrufs aus dem LZG zuordnen kann. Der Nachteil besteht darin, dass man nicht gleichzeitig den Produktionsprozess genauer untersuchen kann. Genauer gesagt, kann man bei vorhandenen Effekten auf die Produktionsgenauigkeit (und schnellen Reaktionszeiten, die darauf hindeuten, dass zwischen Abruf und Produktion keine Korrekturprozesse mehr stattgefunden haben) davon ausgehen, dass die Effekte auf Interferenz während der Vorstellungsphase beruhen. Bei nicht vorhandenen Effekten dagegen ist nicht entscheidbar, ob die Sekundäraufgaben tatsächlich keine spezifische Interferenz erzeugt haben oder ob Effekte durch blitzschnelle Korrekturen während der Produktionsphase, also nach dem Ende der Sekundäraufgaben, nivelliert wurden. Diese Argumentation macht deutlich, dass außer der Produktionsgenauigkeit ein weiteres Maß erforderlich ist, um Gedächtnisprozesse beim Melodieabruf abbilden zu können. Dieses Maß ist die Zeit, die zum Abruf erforderlich ist. In vielen Experimenten zur visuellen Vorstellung hat sich gezeigt, dass man für eine Vorstellungsgenerierung länger braucht, wenn man gleichzeitig eine zweite Vorstellung generieren muss. Analog müsste die Vorstellung einer Melodiephrase länger dauern, wenn die Sekundäraufgabe eine für den Melodieabruf notwendige Gedächtnis-Komponente beansprucht.

Aufgrund der in Kap. 4.4.5 berichteten Befunde zur Intonationsgenauigkeit wurde angenommen, dass Sekundäraufgaben, die die auditive Vorstellung beanspruchen, die Genauigkeit der Produktion beeinträchtigen und die Vorstellungsphase verlängern. Aufgrund der Befundlage wurde außerdem angenommen, dass Sekundäraufgaben, die die singmotorische Vorstellung beanspruchen, die Genauigkeit der Produktion nicht beeinträchtigen und die Vorstellungsphase nicht verlängern. Bei beiden Interferenzarten wurden jedoch aufgrund der unterschiedlichen Lerngeschichte der abzurufenden Lieder Gruppenunterschiede vermutet. Da die Laien ihre Lieblingslieder vorrangig durch Hören von Tonträgern lernen, werden in dieser Stichprobe weniger singmotorische Gedächtnisinhalte vermutet als in der Chorsänger-Stichprobe, die ihre Lieder vorrangig durch das Singen lernte. Folglich wird angenommen,

1. dass bei Laien die Vorstellungsphase bei gleichzeitig dargebotener auditiver Sekundäraufgabe verlängert und die Produktion ungenauer wird im Vergleich zur ungestörten Bedingung. Bei gleichzeitig dargebotener singmotorischer Sekundäraufgabe dagegen wird die Produktion nicht wesentlich ungenauer und die Vorstellungsphase nicht verlängert.

2. dass bei Chorsängern die Vorstellungsphase bei gleichzeitig dargebotener auditiver Sekundäraufgabe verlängert und die Produktion ungenauer wird im Vergleich zur ungestörten Bedingung. Auch bei gleichzeitig dargebotener singmotorischer Sekundäraufgabe wird die Produktion ungenauer und die Vorstellungsphase verlängert.

Bezüglich der Reaktionszeit (also der Zeit zwischen Reaktionstasten-Druck und Produktions-Beginn) wurde kein Effekt der experimentellen Bedingung angenommen, da die Sekundäraufgaben mit dem Tastendruck geendet hatten und sich auf die zur Produktion benötigte Zeit nicht mehr auswirken sollten. Die Reaktionszeit wurde nur zur Kontrolle erfasst, um festzustellen, ob kurze Vorstellungszeiten durch längere Reaktionszeiten „erkauft" worden sind.

9.2.3 Ergebnisse

9.2.3.1 Laien

Zunächst wurde geprüft, ob die Vorstellungs- und Reaktionszeiten sowie die produzierten Abweichungen in allen Bedingungen einer Normalverteilung genügen. Dies konnte durch den Kolmogorov-Smirnov-Test bestätigt werden (Tab. 9.6).

Tab. 9.6: Ergebnisse der Normalverteilungsprüfung aller abhängigen Variablen

	VT 1	**RT 1**	**MV 1**	**MA 1**	**VT 2**	**RT 2**	**MV 2**	**MA 2**	**VT 3**	**RT 3**	**MV 3**	**MA 3**
Z	1,237	0,512	0,987	1,018	1,007	0,699	0,897	1,121	1,198	1,029	1,038	1,138
Sig.	.094	.956	.284	.251	.262	.712	.398	.162	.113	.241	.231	.150

VT = Vorstellungszeit, RT = Reaktionszeit, MV = Mittelpunkte der Fehlerverteilungen, MA = Mittlere Abweichungen ohne Berücksichtigung der Fehlerrichtung. Dargestellt sind die Werte der drei Versuchsbedingungen.

Dann wurde für die Laien geprüft, ob es signifikante Effekte der Produktions-Reihenfolge gab. Tab. 9.7 zeigt die mittleren Vorstellungszeiten, Reaktionszeiten, mittleren Abweichungen und Mittelpunkte der Fehlerverteilungen für die drei gesungenen Lieder.

Tab. 9.7: Mittelwerte der abhängigen Variablen in den 3 Versuchsdurchgängen und Signifikanzprüfung der Unterschiede mit ANOVAs.

	Produktions-Reihenfolge			**Ergebnis der ANOVA**
	Lied 1	Lied 2	Lied 3	
Vorstellungszeit				
M (ms)	10.263	11.183	10.121	F (1,49) = 0,329
SE (ms)	1.115	1.135	1.207	p = .686
Reaktionszeit				
M (ms)	951	1.051	1.042	F (1,38) = 0,335
SE (ms)	71	77	105	p = .641
Mittlere Abweichung				
M (HT)	2,67	3,10	3,27	F (1,50) = 0,784
SE (HT)	0,37	0,31	0,36	p = .446
Mittelpunkt der Fehlerverteilung				
M (HT)	- 1,13	- 0,37	- 0,27	F (1,52) = 0,523
SE (HT)	0,58	0,65	0,70	p = .578

Um die Unterschiede auf Signifikanz zu prüfen, wurden einfaktorielle ANOVAs mit dem Messwiederholungsfaktor *Reihenfolge* durchgeführt. Dabei wurde (wie in allen folgenden ANOVAs mit Messwiederholungen) die Greenhouse-Geisser Korrektur angewandt und die Freiheitsgrade, wenn nötig, angepasst. Die Analysen ergaben keinen signifikanten Haupteffekt der *Reihenfolge*. Somit liegt kein Reihenfolgeeffekt vor.

Dann wurde für die Laien geprüft, ob es einen Einfluss der individuellen Beliebtheit der gesungenen Lieder gab. Tab. 9.8 zeigt die mittleren Vorstellungszeiten, Reaktionszeiten, mittleren Abweichungen und Mittelpunkte der Fehlerverteilungen für die drei gesungenen Lieder, geordnet nach der Beliebtheit (Rang 1 = am beliebtesten). Einfaktorielle ANOVAs ergaben keinen signifikanten Haupteffekt für den Messwiederholungsfaktor *Beliebtheit*. Somit ist kein signifikanter Einfluss dieses Faktors nachweisbar.

Tab. 9.8: Mittelwerte der abhängigen Variablen in Abhängigkeit von der Beliebtheit der produzierten Lieder und Signifikanzprüfung der Unterschiede mit ANOVAs.

	Beliebtheit der produzierten Lieder			**Ergebnis der ANOVA**
	Rang 1	Rang 2	Rang 3	
		Vorstellungszeit		
M (ms)	11.112	10.208	10.247	F (1,56) = 0,257
SE (ms)	1.153	1.190	1.118	p = .770
		Reaktionszeit		
M (ms)	987	1.098	959	F (1,46) = 2,015
SE (ms)	74	96	84	p = .151
		Mittlere Abweichung		
M (HT)	2,77	3,33	2,93	F (1,55) = 0,691
SE (HT)	0,36	0,31	0,36	p = .498
		Mittelpunkt der Fehlerverteilung		
M (HT)	0,43	- 0,93	- 1,27	F (1,49) = 1,981
SE (HT)	0,62	0,67	0,61	p = .155

Da die Analysen der Faktoren *Reihenfolge* und *Beliebtheit* keine signifikanten Effekte dieser Faktoren auswiesen und keine Hypothesen über die Interaktion der Faktoren vorlagen, erschien eine einfaktorielle Analyse der Daten mit dem Faktor *Bedingung* gerechtfertigt. Um einen Einfluss der experimentellen Bedingungen auf die abhängigen Variablen zu überprüfen, wurden die mittleren Vorstellungszeiten, Reaktionszeiten, Abweichungen und Mittelpunkte der Fehlerverteilungen einer einfaktoriellen ANOVA mit dem Messwiederholungsfaktor *Bedingung* zugeführt. Tab. 9.9 zeigt die deskriptiven Daten der drei produzierten Lieder, geordnet nach der experimentellen Bedingung, sowie die Ergebnisse der ANOVAs.

Es ergab sich ein signifikanter Einfluss der experimentellen Bedingung auf die Vorstellungszeit. Paarweise Vergleiche zeigten, dass der Effekt auf einem signifikanten Unterschied zwischen Bedingung 1 und Bedingung 3 beruht (mittlere Differenz = - 4,900, SE = 1,223, p = .001). Die drei anderen abhängigen Variablen wurden durch die experimentellen Bedingungen nicht signifikant beeinflusst. Durch dieses Ergebnis wird Hypothese 1 nicht bestätigt, da

sich in den Vorstellungszeiten ein entgegengesetztes Interferenzmuster ergibt als vorhergesagt. Die Vorstellung einer Melodiephrase wird durch die singmotorische Sekundäraufgabe signifikant, durch die auditive Sekundäraufgabe dagegen nur geringfügig verlängert. Bezüglich der Produktionsgenauigkeit kann Hypothese 1 weder bestätigt noch widerlegt werden, da die Produktionsgenauigkeit selbst in der ungestörten Bedingung sehr gering ist, so dass Interferenzeffekte kaum nachweisbar sind. Abbildung 9.4 stellt die Fehlerverteilungen aller drei Bedingungen gegenüber. Chi²-Tests ergaben, dass in Bedingung 1 und 2 nur zufällige Tonarterinnerung vorliegt, während in Bedingung 3 die Hypothese der Gleichverteilung abgelehnt werden kann (Bedingung 1: $chi^2 = 15{,}12$, $p = .238$, Bedingung 2: $chi^2 = 15{,}12$, $p = .238$, Bedingung 3: $chi^2 = 25{,}55$, $p = .013$).

Tab. 9.9: Mittelwerte der abhängigen Variablen in den 3 Versuchsbedingungen und Signifikanzprüfung der Unterschiede mit ANOVAs.

	Experimentelle Bedingung			**Ergebnis der ANOVA**
	Keine Störung	Störung Motorik	Störung Hören	
Vorstellungszeit				
M (ms)	8.197	13.097	10.273	F (1,57) = 7,424
SE (ms)	897	1.123	1.237	p = .001
Reaktionszeit				
M (ms)	963	987	1.093	F (1,39) = 1,488
SE (ms)	81	73	100	p = .237
Mittlere Abweichung				
M (HT)	3,43	2,80	2,80	F (1,54) = 1,104
SE (HT)	0,36	0,34	0,32	p = .336
Mittelpunkt der Fehlerverteilung				
M (HT)	- 0,50	- 0,27	- 1,00	F (1,57) = 0,324
SE (HT)	0,73	0,62	0,58	p = .722

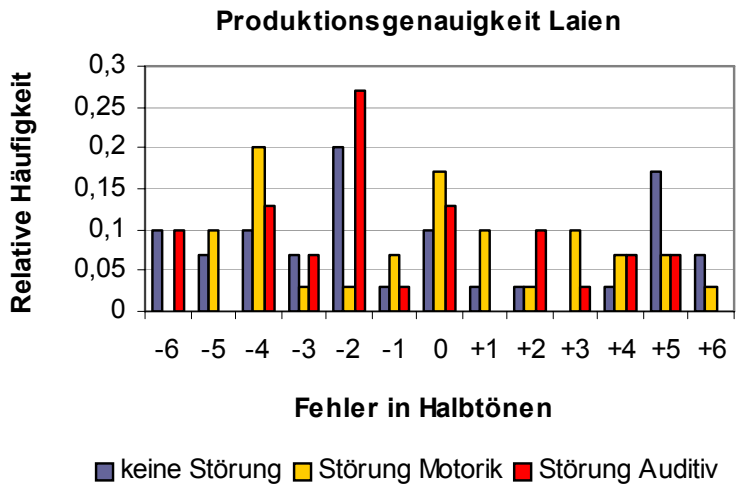

Abb. 9.4: Produktionsgenauigkeit der Laien in den drei experimentellen Bedingungen. Dargestellt sind die relativen Fehlerhäufigkeiten von 30 Vpn. Oktavfehler wurden nicht als Fehler bewertet.

Im Test auf absolutes Gehör wurden durchschnittlich 0,3 (SD = 0,8) Töne richtig benannt. Bei 0,7 (SD = 1,37) Tönen betrug der durchschnittliche Fehler maximal einen Halbton. Damit ist die Tonbenennungsleistung dieser Stichprobe im Mittel als zufällig zu betrachten, denn der Mittelwert liegt unter dem Zufallswert 1. Wie aus Abbildung 9.5 ersichtlich ist, befand sich in der Laien-Stichprobe kein einziger Absoluthörer. Die Tonbenennungsleistungen wurden für diese Stichprobe nicht gesondert mit der Produktionsgenauigkeit korreliert, da es kaum interindividuelle Unterschiede in der Tonbenennungsleistung gibt.

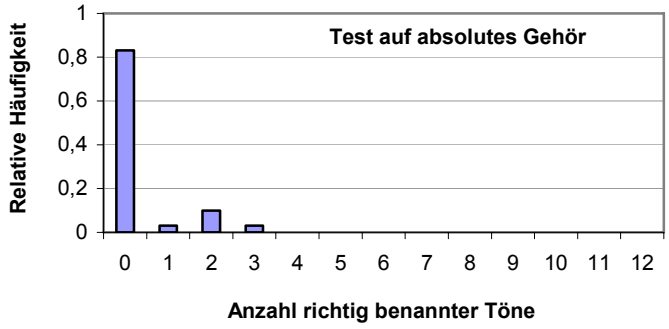

Abb. 9.5: Anzahl korrekt benannter Töne über 30 Vpn.

Um einen Einfluss der musikalischen Ausbildung auf die Produktionsleistung dieser Stichprobe zu untersuchen, wurden die Differenzen mit der Dauer der Instrumentalausbildung korreliert. Es ergab sich ein tendenzieller Zusammenhang zwischen Produktionsleistung und der Dauer des Instrumentalunterrichts (r = - 0,339, p = .067).

9.2.3.2 Chorsänger

Zunächst wurde auch für die Chorsänger geprüft, ob die Vorstellungs- und Reaktionszeiten sowie die produzierten Abweichungen in allen Bedingungen einer Normalverteilung genügen. Dies konnte durch den Kolmogorov-Smirnov-Test bestätigt werden (Tab. 9.10).

Tab. 9.10: Ergebnisse der Normalverteilungsprüfung aller abhängigen Variablen

	VT 1	RT 1	MV 1	MA 1	VT 2	RT 2	MV 2	MA 2	VT 3	RT 3	MV 3	MA 3
Z	0,547	0,504	0,846	0,909	0,703	0,495	0,618	0,811	1,195	0,889	1,050	1,106
Sig.	.926	.962	.471	.380	.706	.967	.839	.526	.115	.409	.220	.173

VT = Vorstellungszeit, RT = Reaktionszeit, MV = Mittelpunkte der Fehlerverteilungen, MA = Mittlere Abweichungen ohne Berücksichtigung der Fehlerrichtung. Dargestellt sind die Werte der drei Versuchsbedingungen.

Um zu prüfen, ob in den beiden Chören vergleichbare Leistungen vorlagen, wurden die Vorstellungszeiten, Reaktionszeiten und mittleren Abweichungen (mit Oktavfehlern) mit t-Tests für unabhängige Stichproben verglichen (vgl. Tab. 9.11). Da keine signifikanten Unterschiede vorliegen, können die Daten aller Chorsänger zusammen analysiert werden.

Tab. 9.11: Vergleich der abhängigen Variablen in den beiden untersuchten Chorstichproben und Signifikanzprüfung der Unterschiede

	Chor	N	Mittelwert	SE	Ergebnis Levene-Test, t-Test
Vorstellungszeit (in ms)	Ensemberlino Vocale	16	9,379	1,036	F = 0,219, p = .644
(alle Bedingungen)	Ensemble Vocal de Bardou	9	10,387	1,431	t (23) = - 0,577 p = .570
Reaktionszeit (in ms)	Ensemberlino Vocale	16	1,185	0,102	F = 0,000, p = .992
(alle Bedingungen)	Ensemble Vocal de Bardou	9	1,446	0,127	t (23) = - 1,592 p = .126
mittlere Abweichung vom Original-Anfangston (in HT)	Ensemberlino Vocale	16	2,467	0,524	F = 0,415, p = .526
(alle Bedingungen)	Ensemble Vocal de Bardou	9	3,111	0,430	t (23) = - 0,850 p = .404

Dargestellt sind die Gruppenmittelwerte, Standardfehler (SE) sowie Ergebnisse des Levene-Tests und des t-Tests für unabhängige Stichproben.

Zunächst wurde auch für die Chorsänger geprüft, ob es signifikante Reihenfolgeeffekte gab. Tab. 9.12 zeigt die mittleren Vorstellungszeiten, Reaktionszeiten, mittleren Abweichungen und Mittelpunkte der Fehlerverteilungen für die drei gesungenen Lieder, geordnet nach der Produktions-Reihenfolge. Einfaktorielle ANOVAs ergaben keinen signifikanten Haupteffekt für den Messwiederholungsfaktor *Reihenfolge*. Somit liegt kein signifikanter Effekt der Produktions-Reihenfolge vor.

Tab. 9.12: Mittelwerte der abhängigen Variablen in den 3 Versuchsdurchgängen und Signifikanzprüfung der Unterschiede mit ANOVAs.

	Produktions-Reihenfolge			Ergebnis der ANOVA
	Lied 1	Lied 2	Lied 3	
Vorstellungszeit				
M (ms)	10.226	9.235	9.763	F (1,40) = 0,340
SE (ms)	1.264	868	1.072	p = .677
Reaktionszeit				
M (ms)	1.223	1.358	1.218	F (1,33) = 0,542
SE (ms)	90	137	95	p = .529
Mittlere Abweichung				
M (HT)	2,80	2,63	2,80	F (1,43) = 0,049
SE (HT)	0,55	0,38	0,45	p = .947
Mittelpunkt der Fehlerverteilung				
M (HT)	- 2,56	-2,29	- 2,56	F (1,42) = 0,088
SE (HT)	0,59	0,47	0,50	p = .903

Als nächstes wurde für die Chorsänger geprüft, ob es einen Einfluss der individuellen Beliebtheit der gesungenen Lieder gab. Tab. 9.13 zeigt die mittleren Vorstellungszeiten, Reak-

tionszeiten, mittleren Abweichungen und Mittelpunkte der Fehlerverteilungen für die drei gesungenen Lieder, geordnet nach der Beliebtheit (Rang 1 = am beliebtesten). Einfaktorielle ANOVAs ergaben keinen signifikanten Haupteffekt für den Messwiederholungsfaktor *Beliebtheit*. Somit ist kein signifikanter Einfluss dieses Faktors nachweisbar.

Tab. 9.13: Mittelwerte der abhängigen Variablen in Abhängigkeit von der Beliebtheit der produzierten Lieder und Signifikanzprüfung der Unterschiede mit ANOVAs.

	Beliebtheit der produzierten Lieder			**Ergebnis der ANOVA**
	Rang 1	Rang 2	Rang 3	
		Vorstellungszeit		
M (ms)	10.233	8.726	10.265	F (1,40) = 1,101
SE (ms)	830	908	1.394	p = .333
		Reaktionszeit		
M (ms)	1.223	1.273	1.297	F (1,36) = 0,373
SE (ms)	99	92	130	p = .641
		Mittlere Abweichung		
M (HT)	2,44	2,71	3,08	F (1,45) = 0,616
SE (HT)	0,42	0,42	0,54	p = .542
		Mittelpunkt der Fehlerverteilung		
M (HT)	- 1,96	- 2,54	- 2,92	F (1,44) = 1,256
SE (HT)	0,52	0,46	0,57	p = .294

Da die Analysen der Faktoren *Reihenfolge* und *Beliebtheit* keine signifikanten Effekte dieser Faktoren auswiesen und keine Hypothesen über eine Interaktion der Faktoren vorlagen, erschien eine einfaktorielle Analyse der Daten mit dem Faktor *Bedingung* gerechtfertigt. Um einen Einfluss der experimentellen Bedingungen auf die abhängigen Variablen zu überprüfen, wurden die mittleren Vorstellungszeiten, Reaktionszeiten, Abweichungen und Mittelpunkte der Fehlerverteilungen einer einfaktoriellen ANOVA mit dem Messwiederholungsfaktor *Bedingung* zugeführt. Tab. 9.14 zeigt die deskriptiven Daten der drei produzierten Lieder, geordnet nach der experimentellen Bedingung, sowie die Ergebnisse der ANOVAs.

Es ergab sich ein signifikanter Einfluss der experimentellen Bedingung auf die Vorstellungszeit. Paarweise Vergleiche zeigten, dass der Effekt auf signifikanten Unterschieden zwischen Bedingung 1 und Bedingung 2 (mittlere Differenz = - 4,124, SE = 0,830, $p < .001$) sowie zwischen Bedingung 2 und Bedingung 3 (mittlere Differenz = 3,508, SE = 1,166, $p = .018$) beruht. Die drei anderen abhängigen Variablen wurden durch die experimentellen Bedingungen nicht signifikant beeinflusst. Durch dieses Ergebnis wird Hypothese 2 nicht bestätigt, da sich in den Vorstellungszeiten ein der Hypothese gegenläufiges Interferenzmuster zeigt: Nur durch die singmotorische Sekundäraufgabe wurden die Vorstellungszeiten verlängert, nicht jedoch wie vermutet durch die auditive Sekundäraufgabe.

Bezüglich der Produktionsgenauigkeit kann Hypothese 2 weder bestätigt noch widerlegt werden, da wie bei den Laien selbst in der ungestörten Bedingung nur eine sehr ungenaue Tonarterinnerung vorliegt (vgl. Ergebnis des t-Tests in Kap. 6.4.5) und somit kein Einfluss

der Sekundäraufgaben nachweisbar ist. Abbildung 9.6 stellt die Fehlerverteilungen der drei Bedingungen gegenüber. Kolmogorov-Smirnov-Tests zeigen, dass alle drei Verteilungen nicht gleichverteilt sind, somit ist in allen drei Bedingungen von zwar ungenauer, aber überzufälliger Tonarterinnerung auszugehen (Bedingung 1: Z = 1,511, p = .021, Bedingung 2: Z = 1,800, p = .003, Bedingung 3: Z = 1,700, p = .006).

Tab. 9.14: Mittelwerte der abhängigen Variablen in den 3 Versuchsbedingungen und Signifikanzprüfung der Unterschiede mit ANOVAs.

	Experimentelle Bedingung			**Ergebnis der ANOVA**
	Keine Störung	Störung Motorik	Störung Hören	
Vorstellungszeit				
M (ms)	8.161	12.286	8.777	F (1,42) = 9,368
SE (ms)	610	1.105	1.232	p = .001
Reaktionszeit				
M (ms)	1.353	1.190	1.255	F (1,33) = 0,597
SE (ms)	89	83	141	p = .505
Mittlere Abweichung				
M (HT)	3,13	2,92	2,20	F (1,41) = 1,836
SE (HT)	0,54	0,38	0,45	p = .176
Mittelpunkt der Fehlerverteilung				
M (HT)	- 3,04	-2,44	- 1,96	F (1,40) = 1,902
SE (HT)	0,56	0,50	0,50	p = .167

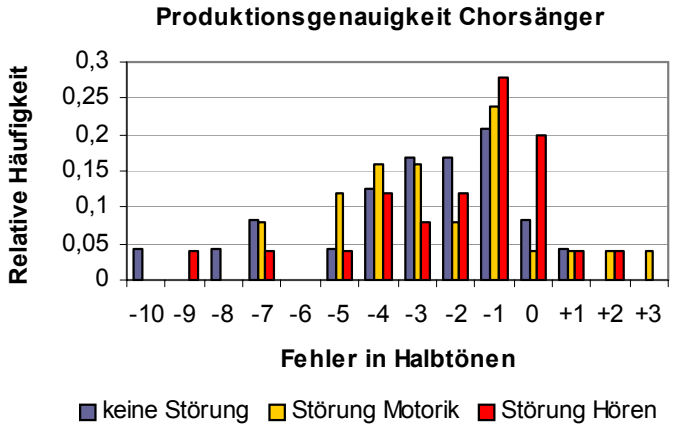

Abb. 9.6: Produktionsgenauigkeit der Chorsänger in den drei experimentellen Bedingungen. Dargestellt sind die relativen Fehlerhäufigkeiten von 25 Vpn. Oktavfehler wurden als Fehler bewertet.

Im Test auf absolutes Gehör wurden durchschnittlich 2,0 (SD = 2,0) Töne richtig benannt. Bei 5,16 (SD = 3,48) Tönen betrug der durchschnittliche Fehler maximal einen Halbton. Damit ist die Tonbenennungsleistung dieser Stichprobe im Mittel als überzufällig zu betrachten (t (24) = 2,5, p = .020). Wie aus Abbildung 9.7 ersichtlich ist, konnten viele Chorsänger 1-2 Töne richtig benennen, während es nur zwei Vpn gelang, mehr als die Hälfte der Töne richtig zu benennen und damit in den Leistungsbereich von Absoluthörern zu kommen.

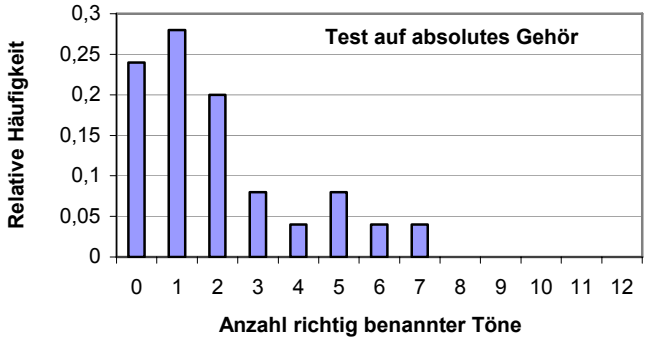

Abb. 9.7: Anzahl korrekt benannter Töne bei 25 Chorsängern.

Um innerhalb dieser Stichprobe einen Einfluss der Fähigkeit, Töne zu benennen, auf die Tonarterinnerung zu überprüfen, wurden die über die drei Bedingungen gemittelten Differenzen aus der Produktionsaufgabe mit der Tonbenennungsleistung korreliert. Es ergab sich kein signifikanter Zusammenhang zwischen Tonbenennungs- und Produktionsleistungen (r = - 0,253, p = .233).

Um außerdem einen Einfluss der musikalischen Ausbildung auf die Produktions- und Tonbenennungsleistung zu untersuchen, wurden weitere Korrelationsanalysen durchgeführt, deren Ergebnisse in Tabelle 9.15 dargestellt sind. Es fanden sich keine Zusammenhänge der Produktionsleistung mit der musikalischen Ausbildung, dagegen ergab sich ein tendenzieller Zusammenhang zwischen Dauer der Instrumentalausbildung und Tonbenennungsleistung.

Tab. 9.15: Korrelationen zwischen den Leistungen in der Produktions- und Tonbenennungsaufgabe und der musikalischen Ausbildung.

		Instrumental-ausbildung in Jahren	**Beginn des Instrumental-unterrichts**	**Gesangs-ausbildung in Jahren**	**Chorerfahrung**
mittlere Produktions-genauigkeit	Pearson Correlation	-,294	,081	-,275	-,235
	Sig. (2-tailed)	,163	,712	,193	,269
	N	25	25	25	25
Absolutes Gehör Tref-ferquote	Pearson Correlation	***,379***	***-,361***	,045	,215
	Sig. (2-tailed)	,061	,083	,829	,302
	N	25	25	25	25

Signifikante Korrelationen auf dem Niveau α =5 % sind fett gedruckt, Tendenzen (α = 10 %) sind kursiv.

Um die Leistung beider Stichproben in der Produktionsaufgabe zu vergleichen, wurden die Vorstellungszeiten, Reaktionszeiten und mittleren Abweichungen (nun bei allen Vpn ohne Berücksichtigung von Oktavfehlern) mit t-Tests für unabhängige Stichproben verglichen. Tabelle 9.16 zeigt die Ergebnisse. Es zeigt sich, dass die Chorsänger etwas kürzere Vorstellungszeiten haben, jedoch signifikant längere Reaktionszeiten. Außerdem sind für die Chorsänger signifikant bessere Produktionsleistungen nachweisbar.

Tab. 9.16: Vergleich der Vorstellungszeiten, Reaktionszeiten und mittleren produzierten Abweichungen bei den Vpn der beiden Stichproben sowie Signifikanzprüfung der Unterschiede

	Stichprobe	N	Mittelwert	SE	Ergebnisse Levene-Test, t-Test
Vorstellungszeit (in ms) (alle Bedingungen)	Laien	30	10,522	0,810	F = 0,078, p = .781
	Chorsänger	25	9,741	0,827	t (53) = 0,670 p = .505
Reaktionszeit (in ms) (alle Bedingungen)	Laien	30	0,991	0,070	F = 0,195, p = .661
	Chorsänger	25	1,283	0,082	t (53) = - 2,721 p = .009
mittlere Abweichung vom Original-Anfangston (in HT) (alle Bedingungen)	Laien	30	3,011	0,192	F = 0,137, p = .713
	Chorsänger	25	2,319	0,221	t (23) = 2,369 p = .022

9.2.4 Diskussion

Mit dem vorliegenden Experiment sollte geprüft werden, inwieweit der Abruf einer vertrauten Melodiephrase durch selektive Störung des inneren Hörens sowie des inneren Singens beeinträchtigt wird. Zu diesem Zweck wurden Laien und Chorsänger gebeten, sich ihre Lieblings-(Chor-)Lieder unter verschiedenen Bedingungen vorzustellen und zu produzieren.

Die Ergebnisse zeigen sowohl bei den Laien als auch bei den Chorsängern eine signifikante Verlängerung der Vorstellungszeit um ca. 5 s beim Vorliegen einer motorischen Sekundäraufgabe. Beim Vorliegen einer auditiven Sekundäraufgabe kam es bei den Chorsängern nicht zu einer verlängerten Vorstellungszeit, bei den Laien nur zu einer nicht signifikanten Verlängerung um ca. 2 s. Die selektive Beanspruchung des inneren Hörens bzw. der klanglichen Vorstellungskomponente hat sich demnach nicht (bei den Chorsängern) bzw. nicht sehr stark (bei den Laien) auf die zur Vorstellung benötigte Zeit ausgewirkt. Dies könnte mehrere Gründe haben. Zum einen könnte es bedeuten, dass keine Interferenz erzeugt wurde, weil die auditive bzw. klangliche Vorstellung für den Melodieabruf aus dem LZG nicht unbedingt benötigt wird. Entweder ist ein Abruf überwiegend auf der Basis singmotorischer Information möglich oder der Abruf funktioniert im Sinne des lexikalischen Ohrs (vgl. Kap. 4.4.5), das durch beide Arten von Interferenz nicht beeinträchtigt wird. Die andere Erklärung wäre, dass die phonologische Speicherkomponente durch den Breitbandton nicht wirkungsvoll genug gestört wurde, um Interferenz zu erzeugen. Gegen beide Erklärungen spricht jedoch die Tatsache, dass sich 66 % der Vpn subjektiv stärker durch den Breitbandton gestört fühlten als durch die motorische Störaufgabe. Eine dritte Erklärung wäre, dass die Sekundäraufgabe zwar gestört hat, sich dadurch die Vorstellungszeit aber nicht verlängert, u.a. weil die Vpn den als unangenehm empfundenen Klang so schnell wie möglich wieder abstellen wollen. Kontrollanalysen zeigen, dass in der auditiven Störbedingung die Reaktionszeit (also die Zeit zwischen Tastendruck und Produktions-Beginn) nicht signifikant verlängert ist. Insofern kann es keinen Tradeoff zwischen Vorstellungszeit und Reaktionszeit in dem Sinne, dass die Vpn

nach Druck der Reaktionstaste und Ende der Störaufgabe noch weiter über die Melodie nachgedacht haben, gegeben haben. Trotzdem wäre eine mögliche Erklärung des erhaltenen Ergebnisses, dass sich die Störung der auditiven Vorstellung eher auf die Qualität der Melodie-Vorstellung auswirkt als auf die zum Abruf benötigte Zeit.

Das Ergebnis, dass es beim Vorliegen einer singmotorischen Sekundäraufgabe zu einer signifikanten Verlängerung der Vorstellungszeit kam, kann mehrere Ursachen haben: Zum einen spricht es für eine notwendige Nutzung singmotorischer Gedächtnisinhalte beim Melodie-Abruf. Eine zweite Erklärung wäre aber die, dass durch die aktive Sekundäraufgabe keine spezifische Interferenz für die singmotorische Vorstellung erzeugt wurde, sondern insgesamt Aufmerksamkeit von der Aufgabenbearbeitung abgezogen wurde. Diese Interpretation lässt sich nur durch weitere Experimente, unter der Nutzung einer zweiten aktiven, jedoch nicht spezifisch auf die Singmotorik ausgerichteten Sekundäraufgabe, also z.B. der Anspannung eines nicht zum Singen benötigten Körperteils, testen. Verwandt mit dem Einwand der zentralen Störung ist die Annahme, dass auch das Starten der Störbedingung, also das Führen des Strohhalms zum Mund und die dadurch veränderte Atmung, zu der langsameren Reaktion in dieser Bedingung geführt haben könnte.

Anhand der Produktionsgenauigkeit in den drei experimentellen Bedingungen ist keine Entscheidung darüber möglich, welche der beiden Gedächtniskomponenten beim Melodieabruf aus dem LZG eine Rolle spielen, da es keine Effekte der Bedingung auf die Produktionsgenauigkeit gab. Bei den Laien ist die Produktionsgenauigkeit auch in der ungestörten Bedingung so schlecht, dass nur von zufälliger Tonartproduktion zu reden ist. Bei Chorsängern dagegen lag zwar eine überzufällige, aber sehr ungenaue Tonartproduktion vor, wie der t-Test gegen zufällige Abweichungsgrößen zeigte. Bei insgesamt sehr ungenauer Tonartproduktion sind Verschlechterungen in den Störbedingungen kaum noch nachweisbar. Tatsächlich kam es bei beiden Gruppen in den Störbedingungen zu leichten Leistungsverbesserungen, die aber nicht signifikant sind und daher nicht weiter interpretierbar.

Unbedingt notwendig erscheint eine weitere Untersuchung des Melodieabrufs mit Interferenz bei einer Stichprobe bzw. mit Melodien, die unter ungestörten Bedingungen so gut produziert werden können wie beispielsweise der Eingangschor des Weihnachtsoratoriums oder die oben genannten Instrumentalkonzerte. Allerdings ergibt sich bei der Suche nach einer unter „Normalbedingungen“ gut produzierenden Stichprobe (bzw. gut produzierten Melodien) das Problem, dass diese Ergebnisse nicht verallgemeinerbar bzw. unter leicht veränderten Bedingungen nicht unbedingt replizierbar sind. In diesem Experiment wurde auf „Levitin-Bedingungen“ (Levitin, 1994) zurückgegriffen, also Laien anhand ihrer Lieblingslieder untersucht, in der Annahme, dass sich dort ein stabiler Tonartgedächtnis-Effekt zeigt. Dies hat sich aber als Irrtum erwiesen. Neben Mängeln in der Standardisierung dieses Experiments könnte dafür auch eine (von vielen Vpn beschriebene) Verunsicherung durch die computergesteuerte Versuchsdurchführung verantwortlich sein, die darauf hinweist, wie störanfällig und kontextabhängig das Tonhöhengedächtnis ist. Diese Störanfälligkeit zeigt sich auch an der Tatsache, dass bei den Laien eine leichte Verschlechterung der Produktion als Reihenfolgeeffekt

auftrat, die zwar mit nur ca. einem Viertelton von Lied 1 nach Lied 3 nicht signifikant wurde, aber dem von Levitin (1994) berichteten Reihenfolgeeffekt entspricht. Ein solcher Reihenfolgeeffekt war bei den Chorsängern nicht nachweisbar, dafür könnte die größere Geübtheit der Chorsänger verantwortlich sein. Dafür zeigte sich bei den Chorsängern ein Effekt der Beliebtheit in gleicher Größenordnung, der dafür spricht, dass bei intentionalem Lernen die emotionale Bewertung gelernter Lieder für das Tonhöhengedächtnis durchaus eine Rolle spielen könnte.

Während bei den Instrumentalmusikern Zusammenhänge zwischen Tonbenennungs- und Produktionsleistungen nachgewiesen wurden, waren solche Zusammenhänge weder bei den Laien noch bei den Chorsängern dieses Experiments zu finden. Dies ist vermutlich darauf zurückzuführen, dass an diesem Experiment keine „echten“ Absoluthörer teilgenommen haben und spricht dafür, dass der Abruf von häufig gehörter oder gesungener Musik nicht davon abhängig ist, ob man die gesungenen Töne auch benennen kann. Dagegen scheint sich bei den Laien vorhandene Instrumentalerfahrung positiv auf die Produktionsleistung auszuwirken. Dies ist bei den Chorsängern wiederum nicht der Fall, vermutlich, weil hier (ähnlich wie bei den übrigen untersuchten Chorsängern) die musikalische Ausbildung wesentlich homogener war als bei den untersuchten Laien. Der Expertiseeffekt bei den Laien beruht möglicherweise auf der Tatsache, dass die erfahreneren Vpn vorgestellte Melodien besser durch Singen externalisieren konnten, was vielen ungeübten Vpn sehr schwergefallen ist.

Die gefundenen Leistungsunterschiede zwischen den beiden untersuchten Stichproben deuten auf etwas unterschiedliche Strategien bei der Aufgabenbewältigung hin. Im Vergleich brauchen die Chorsänger etwas kürzer, um sich eine Melodiephrase vorzustellen, jedoch etwas länger, um nach erfolgreicher Vorstellungsgenerierung die Vorstellung zu singen. Dies könnte ein Hinweis darauf sein, dass bei den Chorsängern nach dem Knopfdruck noch Korrekturprozesse stattfinden und könnte die etwas besseren Produktionsleistungen erklären. Die nicht vorhandene Verlängerung der Vorstellungszeiten der Chorsänger in der auditiven Störbedingung im Vergleich zu der leichten Verlängerung der Laien könnte ein Hinweis dafür sein, dass die Chorsänger singmotorische Gedächtnisinhalte effektiver nutzen können.

9.3 Zusammenfassende Diskussion der Ergebnisse beider Experimente

In den beiden beschriebenen Experimenten sollte der Beitrag motorischer, insbesondere singmotorischer Gedächtnisprozesse für den Abruf von Melodien aus dem LZG untersucht werden. Beide Experimente beruhen auf dem Versuch, die Nutzung singmotorischer Gedächtnisinhalte zu verhindern bzw. einzuschränken. Ein unter solchen Bedingungen erfolgreicher Tonartabruf wird als ein Hinweis darauf angesehen, dass singmotorische Gedächtnisinhalte für den Tonartabruf nicht zwingend notwendig sind.

Im ersten Experiment bestand die Einschränkung der Nutzung singmotorischer Gedächtnisinhalte in der Nutzung einer speziellen Stichprobe. Da Instrumentalmusiker anders als die in anderen Experimenten getesteten Chorsänger und auch die Sänger von inzidentell gelernten Lieblingsliedern ihre im Experiment abgerufenen Stücke nur selten singen, ist davon auszu-

gehen, dass keine singmotorische Übung der Stücke vorliegt. Insofern sind sie beim Abruf auf das Abrufen klanglicher Informationen angewiesen. Die Ergebnisse des ersten Experiments haben gezeigt, dass es Instrumentalmusikern problemlos möglich ist, eine durch häufiges Üben am Instrument gelernte Melodie in annähernd der richtigen Tonart zu singen. Dies spricht für die grundsätzliche Möglichkeit einer Tonarterinnerung ohne die nennenswerte Nutzung singmotorischer Information. Die Befragungsdaten sprechen dafür, dass außer einer klanglichen auch eine Vorstellung von spielmotorischen Informationen hilfreich für den Abruf der Tonart war. Um diesen Aspekt näher zu untersuchen, böte sich eine Untersuchung in Anlehnung an Zatorre & Halpern (1999) an, in der während des Melodieabrufs aus dem LZG die Aktivität in den relevanten motorischen Arealen mithilfe von PET gemessen wird. Eine Stützung der Interpretation, dass spiel- aber nicht singmotorische Informationen wichtig für den Abruf der Instrumentalmelodien sind, wäre durch eine Aktivierung der Areale, in denen z.B. die Finger beider Hände repräsentiert sind und Nichtaktivierung des für Subvokalisationsprozesse als wichtig erachteten supplementären motorischen Kortex gegeben.

Das Experiment mit Instrumentalmusikern hat außerdem gezeigt, dass sich in einer Stichprobe, in der die Tonbenennungsleistung insgesamt überdurchschnittlich ist und stärker variiert als in anderen untersuchten Stichproben, Zusammenhänge zwischen Tonbenennung und Tonartabruf zeigen lassen. Die relativ hohe Korrelation zwischen Wiedererkennungs- und Tonbenennungsleistungen kann zusammen mit der vergleichsweise schlechten Leistung in der Wiedererkennungsaufgabe (verglichen mit der Produktion) aber als Hinweis dafür angesehen werden, dass der Tonartabruf nicht nur auf dem Abruf von Einzeltönen beruht. Ein solcher müsste nämlich in Produktions- und Wiedererkennungsaufgabe gleichermaßen möglich sein. Die besseren Produktionsleistungen sprechen dafür, dass die Nutzung klanglicher und spielmotorischer Information unter den Versuchsbedingungen besser möglich war, in denen keine alternativen Versionen der Melodien, die auch als Interferenz betrachtet werden können (vgl. Kap. 4.3.3), vorhanden waren.

Eine Störung des Melodieabrufs mithilfe von spezifischen Interferenzaufgaben war das Ziel des zweiten beschriebenen Experiments. Einerseits sollte mit einer singmotorischen Interferenzaufgabe geprüft werden, ob ein Melodieabruf auch mit eingeschränkter Nutzung singmotorischer Gedächtnisinhalte möglich ist und ob er dadurch verlangsamt wird. Andererseits sollte mit einer auditiven Interferenzaufgabe geprüft werden, ob ein Melodieabruf, der vor allem auf dem Abruf singmotorischer Gedächtnisinhalte beruht, erfolgreich sein kann und ob er dadurch verlangsamt wird.

Die Ergebnisse des zweiten Experiments zeigten, dass der Melodieabruf durch singmotorische Interferenzaufgaben verlangsamt, aber nicht verschlechtert wird, während sich bei der auditiven Interferenzaufgabe weder eine Verlangsamung noch eine Verschlechterung zeigte. Dies spricht zunächst dafür, dass beim Abruf von Melodien, die entweder durch Singen gelernt oder aber zumindest häufiger mitgesungen werden, die singmotorische Vorstellung nicht ganz unwichtig ist. Allerdings sind einige methodischer Probleme aufgetreten, die weitere Untersuchungen mit der Interferenzmethodik erfordern, bevor die Ergebnisse eindeuti-

ger interpretiert werden können. Der erste Einwand ist, dass die auditive Interferenzaufgabe zu schwach war, um wirkungsvoll zu stören und dass nach dem Ende des Klangs schnelle Korrekturprozesse vor Beginn der Produktion nicht verhindert werden konnten. Ein Lösungsversuch wäre, den Klang auch während der Produktion über den Kopfhörer weiterlaufen zu lassen, so dass die Vpn auch beim Singen überwiegend auf singmotorische Gedächtnisinhalte angewiesen wären. Der zweite Einwand ist, dass die singmotorische Interferenzaufgabe nicht spezifisch, sondern zentral gestört haben könnte. Dieser Einwand wäre durch den Einsatz einer weiteren aktiven Sekundäraufgabe, die den Melodieabruf nicht verlangsamt, zu entkräften. Ein dritter Einwand betrifft die Kompensation der durch die Sekundäraufgaben beanspruchten Ressourcen durch den jeweils ungestörten Gedächtnismechanismus, die durch eine gleichzeitige Störung von auditiver und singmotorischer Vorstellung zu untersuchen wäre. Ein vierter Einwand schließlich betrifft die Auswahl der Stichprobe und Melodien, die produziert werden sollen. Wünschenswert wäre eine Stichprobe, bei der man unter ungestörter Bedingung von relativ guten Produktionsleistungen ausgehen kann, so dass Verschlechterungen überhaupt möglich sind.

Insgesamt lässt sich aus den vorhandenen Ergebnissen schließen, dass singmotorische Gedächtnisinhalte für den Abruf von Instrumentalmelodien keine und für den Abruf von Vokalmelodien keine allein entscheidende Rolle spielen. Die Wichtigkeit einer klanglichen Vorstellung der Melodien sowie bei den Instrumentalmusikern einer spielmotorischen Vorstellung wurde aus Befragungsdaten deutlich. Insgesamt scheint also eine Kombination aus der Erinnerung an den Klang, an ein Sing- und Spielgefühl, sowie vor allem bei hochtrainierten Musikern der Fähigkeit, Töne auf dem Tonhöhenkontinuum zu lokalisieren, vorzuliegen.

10 Zusammenfassung und Diskussion der Hauptergebnisse

In der vorliegenden Arbeit wurde das Phänomen der Tonarterinnerung zunächst aus musikpsychologischer und aus gedächtnispsychologischer Perspektive theoretisch betrachtet. Ein wichtiger Ausgangspunkt war dabei die von Squire & Kandel (1999) vorgenommene Beschreibung des Gedächtnisses. Diese sieht Gedächtnis als Nachwirkung der ursprünglichen Informationsverarbeitung der zu erinnernden Reize an, worauf unter anderem die Tatsache hindeutet, dass Gedächtnisinhalte im Gehirn nicht in einem zentralen Gedächtniszentrum gespeichert werden, sondern in den Arealen, die auch bei der ursprünglichen Informationsverarbeitung aktiv waren. Für die Betrachtung des Melodiegedächtnisses bedeutet diese Beschreibung des Gedächtnisses, dass jene Aspekte von Melodien, die für die Melodiewahrnehmung bedeutsam sind, auch für die Melodieerinnerung bedeutsam sein müssten. Bezogen auf den für diese Arbeit wichtigsten Melodieaspekt, die Tonhöhe, haben viele Untersuchungen von Wahrnehmungs- und Gedächtnisprozessen durch Diana Deutsch (Zusammenfassung: Deutsch, 1999) auf eine starke Kontextabhängigkeit der Tonwahrnehmung hingewiesen. Während einzelne Töne zwar unter bestimmten experimentellen Bedingungen (Stille) für einige Zeit im Gedächtnis gehalten werden können, kommt es unter normalen Hör-Bedingungen zu Wechselwirkungen bei der Wahrnehmung verschiedener Töne. Dies hat für den „normalen“ (Nichtabsolut-) Hörer die Konsequenz, dass er sich beim Hören von Melodien stärker an relationalen Tonhöhenverläufen als an den absoluten Tonhöhen orientiert. Auf Veränderungen der Intervalle oder der Kontur reagieren auch musikalisch untrainierte Hörer sensibel, während Transpositionen weniger sicher bemerkt werden. Wie zahlreiche Untersuchungen sowie die musikalische Praxis zeigen, ist eine Verarbeitung von Tonrelationen sowie eine Reduktion verschiedener gehörter Versionen der gleichen Melodie auf ein melodisches Grundgerüst funktional. Beispielsweise erlaubt eine auf die essentiellen Merkmale der Beziehungen zwischen Tonhöhen und Tondauern reduzierte Melodierepräsentation die Wiedererkennung von Melodien, auch wenn Tonart, Tempo, Instrumentation und Lautstärke im Vergleich zum ersten Hören verändert sind.

Die auf relationale Merkmale reduzierte Melodierepräsentation bedeutet besonders dann eine Entlastung des Gedächtnisses, wenn die Melodien in sehr vielen verschiedenen Versionen gehört werden. Dies ist der Fall bei Melodien, für die keine spezifische Tonart komponiert wurde. Ein klassisches Beispiel für solche Melodien sind Volks- und Kinderlieder, die häufig mündlich überliefert werden. Schriftliche Fixierungen solcher Lieder sind aus Gründen der Lesbarkeit meist in vorzeichenarmen Tonarten gesetzt. Die Nutzung von Noten beim Singen von Volks- oder Kinderliedern bedeutet aber nicht, dass die Lieder in der notierten Tonart gesungen werden, denn sie werden oft entweder ohne Begleitung in einer beliebigen Tonart in angenehmer Singlage angestimmt oder aber in Begleitung einer nicht immer in der Normstimmung gestimmten Gitarre. Neben solchen Melodien, für die keine bestimmte Tonart charakteristisch ist und die folglich in beliebiger Tonart musiziert werden können, gibt es aber auch eine Vielzahl an Melodien, die in immer gleicher Tonart gehört oder gesungen werden. Dies ist der Fall bei Musik, die vom Tonträger gehört wird sowie bei in bestimmter Tonart

komponierter Instrumental- und Vokalmusik, sofern es sich nicht um die Ausnahme der in mehreren Versionen vorliegenden Kunstlieder handelt. Auch im Alltag vieler Menschen gibt es Melodien wie z.B. Werbejingles, Eröffnungsmelodien bestimmter TV- oder Radiosendungen oder die Klänge von Telefonen und PCs, die in immer gleicher Tonart gehört werden. Für solche Melodien wäre eine Tonartrepräsentation weniger aufwändig als für Volkslieder, da es sich nur um eine einzige, statt um viele verschiedene Tonarten handelt. In der vorliegenden Arbeit wird der Hypothese nachgegangen, dass sich für solche in nur einer einzigen Tonart vorliegenden Melodien Repräsentationen auch der absoluten Tonhöhen nachweisen lassen, deren Genauigkeit von einer Reihe von Faktoren abhängt.

Ein zweiter Ausgangspunkt der Arbeit war die Betrachtung der Spezialfähigkeit *absolutes Gehör*, die erforderlich war, weil bisherige Befunde zur Tonarterinnerung häufig im Kontext der Kontinuumsdiskussion des absoluten Gehörs diskutiert wurden. Eine genauere Betrachtung des absoluten Gehörs zeigt, dass diese früh erworbene, eventuell durch genetische Komponenten mit bedingte Fähigkeit auf gelernten Assoziationen zwischen Klängen und Tonnamen zu beruhen scheint, die neben anderen Arten der Kodierung auch eine verbale Kodierung von Tönen erlaubt. Diese ist in experimentellen Bedingungen vorteilhaft, wenn einzelne Töne benannt, produziert oder erinnert werden sollen, was Nichtabsoluthörern nur in bestimmten Grenzen möglich ist. Unter Normalbedingungen jedoch ist absolutes Gehör für die Musikwahrnehmung oder –ausübung nicht notwendig, konnte in Untersuchungen von Miyazaki (1995, 2004) sogar als nachteilig für die Verarbeitung relationaler Informationen nachgewiesen werden. Allerdings sollte es Absoluthörern, die sich an die Namen der Tonarten von gelernten Liedern erinnern können, problemlos möglich sein, diese Lieder in der Originaltonart zu produzieren bzw. wiederzuerkennen.

Leistungskontinua in den Tonidentifikationsleistungen von Absolut- und Nichtabsoluthörern sowie verschiedene Experimente, die latente im Sinne von unbewussten Erinnerungen an Tonarten bei Nichtabsoluthörern zeigten, führten zu der Frage, ob das absolute Gehör kontinuierlich anstatt dichotom ausgeprägt ist. Bisherige Ergebnisse aus der Produktion, Wiedererkennung oder Erinnerung an Einzeltöne sprechen eher für die Nutzung unterschiedlicher Strategien bei Absolut- und Nichtabsoluthörern und legen damit eine Beschreibung absoluten Gehörs als dichotom ausgeprägte Fähigkeit nahe. Offen ist aber bislang, ob die unbewusste Erinnerung an Tonarten als latente Form absoluten Gehörs anzusehen ist. Hier lagen bisher vereinzelte Nachweise der Erinnerung an die Tonarten verschiedener Arten von Melodien vor, aber noch kein Versuch, die Befunde zu einem Erklärungsmodell für die Tonarterinnerung zu integrieren. Eine solche Integration ist einerseits von theoretischem Interesse, da sie z.B. Zusammenhänge zu musikalischer Expertise aufklären und somit aufzeigen könnte, inwieweit es sich um eine musikerspezifische Fähigkeit handelt. Zusammenhänge zwischen Tonarterinnerung und „echtem“ absoluten Gehör sowie eine weitgehende Kontextunabhängigkeit der Tonarterinnerung würden diese eher als latente Form absoluten Gehörs ausweisen. Fehlende Zusammenhänge zwischen Tonarterinnerung und absolutem Gehör und Zusammenhänge mit Faktoren, die sich für „normale“ Gedächtnisprozesse als relevant erwiesen haben, würden die Interpretation der Tonarterinnerung als latentes absolutes Ge-

hör dagegen in Frage stellen. Auf der anderen Seite ist eine integrative Betrachtung der Tonarterinnerung von praktischem Interesse, weil beispielsweise beim Chorsingen explizit bestimmte Tonarten geübt werden und sowohl ein Absinken als auch ein Ansteigen in der Tonart nicht erwünscht ist. Könnte man herausfinden, bei welcher Art des Lernens der Aufbau absoluter Melodierepräsentationen wahrscheinlich ist, so wäre dies für Musikpädagogen praktisch anwendbares Wissen.

Um Korrelate genauer Tonarterinnerung zu finden, wurde in einer Reihe von Experimenten die Genauigkeit der Tonarterinnerung durch Produktionsmethoden untersucht. Dabei wurde sowohl die musikalische Expertise der Melodie-Produzierenden als auch die Art der abzurufenden Melodien variiert. Außerdem kamen mit überwiegend durch Singen und überwiegend durch häufiges Hören gelernten Melodien unterschiedliche Lernformen zum Tragen, und auch die Lernintensität der Melodien wurde variiert.

Bei der Auswahl der experimentellen Methodik stellt die Besonderheit des Lernens von Melodien eine konzeptuelle Herausforderung dar. Melodien werden zwar im Fall von Chor- oder Instrumentalmusik intentional gelernt, dies bezieht sich aber normalerweise nicht auf die Tonarten der Melodien. Diese werden inzidentell, quasi als Nebenprodukt des häufigen Singens oder Spielens der Melodie nur in einer einzigen Tonart, gelernt. Es stellt sich für das Experiment die Frage, ob eine derartig aufgebaute Tonartrepräsentation besser durch explizite oder durch implizite Methoden gemessen werden kann. Ideal wäre ein Vergleich beider Methoden, mit dem man feststellen könnte, ob es bei Nichtabsoluthörern eine stärkere Dissoziation zwischen expliziter und impliziter Tonarterinnerung gibt als bei Absoluthörern. Leistungsunterschiede zwischen beiden Populationen auch bei der Nutzung impliziter Methoden würden die Interpretation des absoluten Gehörs als dichotome Fähigkeit stützen. In der vorliegenden Arbeit wurde die Produktionsmethodik gewählt, die gewissermaßen einen Mittelweg zwischen expliziter und impliziter Erinnerung darstellt. Einerseits wird explizit um die Erinnerung an eine Melodie gebeten. Andererseits wird den Vpn nicht explizit die Produktion einer bestimmten Tonart nahegelegt, sondern vor allem eine *möglichst genaue* Produktion ihrer Vorstellung. Obgleich die Melodieproduktion eine explizite Methode ist, wird die Tonarterinnerung also quasi indirekt getestet. Damit entspricht die Produktionsaufgabe gewissermaßen der Lernsituation von intentional gelernten Melodien.

Die Ergebnisse der Produktions-Experimente zeigen, dass es folgende Einflussfaktoren auf die Qualität der Tonarterinnerung gibt:

1. Es konnte ein Häufigkeitseffekt nachgewiesen werden. Hypothesenkonform werden die Tonarten von häufiger oder länger gehörten Melodien besser produziert als die von seltener oder kürzer gehörten Melodien. Auch bei der Benennung von Tönen durch Absoluthörer und Nichtabsoluthörer konnte ein Effekt der Hör-Häufigkeit nachgewiesen werden.

2. Es konnte ein moderater Expertiseeffekt nachgewiesen werden. Vpn mit längerer musikalischer Ausbildung an einem Musikinstrument konnten Tonarten insgesamt etwas besser produzieren als Vpn mit kürzerer oder ohne musikalische Ausbildung.

3. Es konnte ein moderater Tonbenennungseffekt nachgewiesen werden. Vpn mit besseren Leistungen in der Tonidentifikation konnten Tonarten insgesamt etwas besser produzieren als Vpn mit schlechteren Leistungen in der Tonidentifikation. Es besteht also ein moderater Zusammenhang zwischen Tonarterinnerung und absolutem Gehör.

4. Ein Einfluss singmotorischer Gedächtnisinhalte auf die Tonarterinnerung kann vorsichtig als nicht alleine ausschlaggebend beschrieben werden. Motorische Kontextinformationen scheinen die Tonarterinnerung aber zu erleichtern.

5. Unter den untersuchten Eigenschaften der produzierten Melodien hat sich die Eingängigkeit einer Melodie als Einflussfaktor auf die Genauigkeit der Tonartrepräsentation herausgestellt. Mittelmäßig eingängige Melodien wurden besser produziert als sehr wenig oder extrem eingängige Melodien.

Der gefundene Häufigkeitseffekt ist insofern ein besonders wichtiger Effekt, als er sich sowohl bei Absoluthörern als auch bei Nichtabsoluthörern und sowohl bei der Erinnerung an Einzeltöne als auch bei der Erinnerung an Tonarten als gültig erwiesen hat. Dies verweist darauf, dass es sich bei der Stabilisierung von Gedächtnisinhalten durch Wiederholung um ein grundlegendes und allgemeingültiges Prinzip des Gedächtnisses handelt. Die diesem Gedächtnisprinzip zugrundeliegenden molekularen Prozesse werden nach Squire & Kandel (1999) vor allem während der Speicherungs- bzw. Konsolidierungsphase wirksam. Dabei könnte u.a. der Imprinting-Mechanismus des Wahrnehmungslernens (Goldstone, 1998) eine Rolle spielen, der die Entwicklung von „Detektoren“ im Gehirn als Folge von sehr häufiger Darbietung des gleichen Reizes beinhaltet. Eine weitere Untersuchung des Häufigkeitseffekts beim Tonarterinnern auf der Verhaltensebene könnte von extremerer Variation der Vertrautheit der gleichen Melodie bei verschiedenen Vpn profitieren. Die in dieser Arbeit realisierte Variation spielte sich aufgrund praktischer Einschränkungen nur innerhalb des Bereichs „mittelmäßig vertraut“ ab, da mit Schülern weder häufiger als wöchentlich noch länger als einige Monate die gleiche Melodie geübt werden konnte. Da sich außerdem gezeigt hat, dass das Gruppenlernen von Melodien in Unterrichts- oder Chorproben-Settings mit vielen Problemen – u.a. die schwierige Kontrolle der Anwesenheit – behaftet ist, wäre die Nutzung von Melodien, die im Alltag kontrolliert gelernt werden können, eventuell eine hilfreiche Alternative. Da sich selbst bei der hier realisierten geringen Vertrautheits-Variation ein Effekt der Variation auf die Erinnerungsgenauigkeit gezeigt hat, lassen sich bei einer extremeren Vertrautheits-Variation z.B. über tägliche Lernphasen noch größere Vertrautheits-Effekte erwarten.

Der Effekt musikalischer Expertise auf die Genauigkeit der Tonarterinnerung wurde für die Gesamtgruppe bestätigt. Allerdings ist die nachgewiesene Korrelation nicht sehr hoch. Sie scheint vor allem auf einem Zusammenhang zwischen Tonarterinnerung und musikalischer Ausbildungsdauer innerhalb der musikalisch aktiven Stichprobe (Instrumentalmusiker, Chorsänger, Schüler) zu beruhen, denn während bei den Nichtmusikern bzw. musikalisch nur sehr wenig erfahrenen Vpn sämtliche Ausprägungen der Produktionsgenauigkeit von perfekt bis zufällig auftraten, gab es keine musikalisch hochtrainierte Vp, die Tonarten nur auf dem

Zufallsniveau abrufen konnte. Ein ähnliches Muster ergab sich für den Effekt der Tonbenennungsleistung auf die Genauigkeit der Tonarterinnerung: Während es unter den (wenigen) Vpn mit sehr guten Tonidentifikationsleistungen keine Vp gab, die Tonarten nur auf dem Zufallsniveau produzierte, trat unter den „zufälligen“ Tonbenennern das ganze Leistungsspektrum von perfekter bis zufälliger Tonarterinnerung auf. Die Korrelation zwischen musikalischer Ausbildungsdauer und Tonbenennungsleistung, die etwas höher ist als die jeweiligen Korrelationen dieser beiden Faktoren mit der Tonarterinnerung, könnte eine Konfundierung der Faktoren bewirken, die eine Entscheidung, welcher Faktor für die Genauigkeit der Tonarterinnerung von größerer Bedeutung ist, erschwert.

Da die musikalisch trainierteren Vpn gleichzeitig die waren, die in den hier vorgestellten Experimenten intentional gelernte Melodien produzierten, könnte die Lernart ein weiterer Einflussfaktor in dem Zusammenhang von musikalischer Expertise, Tonbenennung und Tonarterinnerung sein. Denkbar wäre eine Interpretation der Datenlage in dem Sinne, dass sich sowohl musikalische Expertise als auch die Fähigkeit, Töne zu benennen stärker auf die Tonarterinnerung bei intentional gelernten Melodien auswirkt als bei durch häufiges Hören gelernten Melodien. Diese Interpretation wird durch eine zusätzliche Korrelationsanalyse, die Zusammenhänge zwischen musikalischer Expertise, Tonbenennung und Tonarterinnerung nur für die intentional gelernten Melodien fand (Expertise: $r = -0{,}22$, $p = .001$; Tonbenennung: $r = -0{,}22$, $p < .001$), nicht aber für die durch häufiges Hören gelernten Melodien (Expertise: $r = -0{,}05$, $p = .562$; Tonbenennung: $r = -0{,}06$, $p = .510$), bestätigt. Beim intentionalen Lernen könnten bei musikalisch trainierteren Vpn, die meist eine etwas höhere Tonbenennungsfähigkeit aufweisen, veränderte Enkodierungsprozesse, wie sie für Experten in vielen anderen Bereichen (z.B. Schach: Chase & Simon, 1973) nachgewiesen wurden, vorliegen. Möglicherweise können diese Vpn Melodien insgesamt etwas schneller lernen als musikalisch weniger ausgebildete Vpn, so dass mehr Kapazität für die (unbewusste) Aufmerksamkeitsausrichtung auf die Tonarten von Melodien zur Verfügung steht. Außer in der Enkodierungsphase könnten Expertiseeffekte in der Produktionsphase wirksam werden, denn musikalisch erfahrenere Vpn können eine vorgestellte Melodie vermutlich besser externalisieren, d.h. zu ihrer eigenen Zufriedenheit nachsingen. Beim Melodie-Lernen durch häufiges Hören dagegen scheinen andere Einflussfaktoren wichtiger zu sein als die Expertisemerkmale. Beispielsweise wissen die musikalischen Laien oft nicht einmal, was eine Tonart ist, und können folglich auch kaum versuchen, relationale Hinweisreize für den Tonartabruf zur Hilfe zu nehmen. Um zwischen den Effekten musikalischer Expertise, Tonbenennungsfähigkeit und Lernart differenzieren zu können, wären weitere Untersuchungen erforderlich, in denen diese Faktoren vollständig kombiniert werden, also z.B. musikalisch hochtrainierte Vpn auch inzidentell gelernte Melodien produzieren.

Die Effekte verschiedener Abrufmethoden (Produktion und Wiedererkennung) zeigen, dass man kaum von einem allgemeingültigen „Vorteil“ der einen oder anderen Methode sprechen kann. Statt dessen bilden beide Methoden Charakteristika des Lernens unterschiedlich genau ab und sind von weiteren Faktoren abhängig. Bei künstlichen Klangfarben ähnelt die Wiedererkennungsaufgabe der Tonbenennungsaufgabe, daher zeigen sich dort stärkere

Zusammenhänge zwischen den beiden Aufgaben. Bei weniger künstlichen Klangfarben dagegen liegt vermutlich eine stärkere Orientierung am Originalklang vor, weniger die Nutzung der Tonbenennungsfähigkeit. Für eine eindeutigere Interpretation der Einflussfaktoren auf die Wiedererkennung von Tonarten sind weitere Experimente mit Wiedererkennungsmethodik erforderlich, in denen die Lerndauer, Expertise sowie die genutzten Klangfarben gezielt variiert werden.

Ein häufiger Einwand bei der Interpretation von Produktionsergebnissen betrifft den Einfluss singmotorischer Gedächtnisinhalte auf die Genauigkeit der Tonartproduktion. Diesen Einwand auszuräumen und zu zeigen, dass singmotorische Gedächtnisinhalte für die Genauigkeit des Tonartabrufs nicht alleine entscheidend sein können, war das Ziel der beiden Experimente, in denen die Nutzung der Singmotorik durch Interferenz sowie durch Nutzung einer speziellen Stichprobe verhindert werden sollte. Während der Vergleich der Produktionsleistungen von Instrumentalmusikern und Chorsängern zeigt, dass Melodien, die so gut wie nie gesungen worden sind, genauso gut bzw. besser produziert werden können als Melodien, die explizit durch Singen gelernt wurden, erbrachte das Interferenzexperiment weniger eindeutige Ergebnisse. Neben weiteren Untersuchungen mit verfeinerten Interferenzaufgaben wären zur Untersuchung des Motorik-Aspektes bei der Tonarterinnerung auch bildgebende Verfahren (analog Halpern & Zatorre, 1999) hilfreich. Da auch im Interferenzexperiment die Produktionsleistungen in einer singmotorischen Störbedingung nicht zusammenbrachen, kann vorläufig davon ausgegangen werden, dass der Einfluss der Singmotorik nicht alleine ausschlaggebend für das Tonartgedächtnis ist. Ein in weiteren Experimenten nachzuweisender fehlender Effekt von Interferenzaufgaben für die phonologische Schleife auf die Tonarterinnerung würde die Theorie des lexikalischen Ohres von Besner (1987, vgl. Kap. 4.4.5) stützen.

Effekte der verschiedenen für den Abruf genutzten Melodien zeigen, dass eine Generalisierbarkeit erhaltener Ergebnisse auf andere Melodien nur eingeschränkt möglich ist. Auch bei ein und derselben Vp zeigt sich bei der Produktion mehrerer Melodien eine Abhängigkeit der Produktionsgenauigkeit von der Art der Melodien. Welche melodischen Faktoren dabei entscheidend sind, konnte hier nur näherungsweise untersucht werden, wünschenswert wäre eine kontrollierte Studie mit gezielt hinsichtlich bestimmter Faktoren variierten Melodien. Dabei wäre außerdem die persönliche Einschätzung der Melodiefaktoren durch die produzierenden Vpn unerlässlich, vor allem bei Merkmalen, die sich auf die emotionale Beurteilung der Melodien beziehen. Diesbezüglich konnten hier durch die post-hoc Expertenbeurteilung keine Zusammenhänge gezeigt werden, aber das Ergebnis könnte bei persönlicher Einschätzung ganz anders aussehen. Insgesamt deuten die Melodie-Effekte darauf hin, dass eine weitgehende Kontrolle des melodischen Materials, inklusive der Angabe der Vpn, wie oft und in welchen Tonarten sie die jeweilige Melodie als Ohrwurm haben, unerlässlich ist für eine verlässliche Interpretation der Daten.

Zusammenfassend lässt sich festhalten, dass es eine Reihe von Einflussfaktoren auf die Qualität der Tonarterinnerung bei gelernten Melodien gibt, zwischen denen auch Wechsel-

wirkungen vorzuliegen scheinen. Ordnet man die Einflussfaktoren auf der Grundlage theoretischer Überlegungen nach den basalen Gedächtnisprozessen, so lassen sich folgende Einflüsse finden: Während der *Enkodierungsphase*, in der Melodien für die weitere Speicherung vorbereitet werden, können Unterschiede in der musikalischen Expertise zusammen mit Unterschieden in der Fähigkeit Töne zu benennen eine unterschiedliche Aufmerksamkeitsausrichtung bewirken, die eine spätere Tonarterinnerung beeinflussen kann. Allerdings gilt dieser Zusammenhang mehr für intentional als für inzidentell gelernte Melodien. Während der *Speicher- und Konsolidierungsphase* wirkt sich die Häufigkeit oder Intensität des Lernens auf die Qualität der späteren Tonartrepräsentation aus. Auf die Art und Intensität des Lernens wiederum haben vermutlich verschiedene Melodiemerkmale einen Einfluss, zu denen u.a. die Prägnanz und die emotionale Bewertung von Melodien zählen. Außerdem wirkt sich die Lernart auf die Art der motorischen Gedächtnisinhalte aus, die im Zusammenhang mit einer Melodie gespeichert werden. Beim *Abruf* können Produktions- und Wiedererkennungsmethoden unterschiedliche Aspekte des Melodie- und Tonartlernens abbilden, ohne dass sich bisher ein genereller Vorteil für die eine oder andere Methode zeigen lässt.

Die Ergebnisse der vorliegenden Arbeit zeigen, dass die Tonarterinnerung ein komplexes Phänomen ist. Da die Zusammenhänge zwischen Tonarterinnerung und „echtem“ absoluten Gehör nicht sehr hoch sind, da die Tonarterinnerung stark melodie- und somit kontextabhängig ist und da sich eine Reihe von gedächtnispsychologisch begründeten Einflussfaktoren aufzeigen ließ, sprechen die Ergebnisse gegen eine Interpretation des Tonartgedächtnisses als „latentes absolutes Gehör“. Statt einer schwachen Ausprägung einer hochspezialisierten Fähigkeit scheint es sich eher um eine eigene Form des Erinnerns, die auf allgemeingültigen Gedächtnisprinzipien beruht, zu handeln.

Literatur

Abraham, O. (1901). Das absolute Tonbewußtsein. *Sammelbände der Internationalen Musikgesellschaft, 3*, 1-86.

Atkinson, R.C. & Shiffrin, R.M. (1968). Human memory: A proposed system and its control mechanisms. In K.W. Spence (Hg.), *The psychology of learning and motivation* (Bd. 2, S. 89-195). New York: Academic Press.

Attneave, F. (1954). Some information aspects of visual perception. *Psychological Review, 61*, 183-193.

Attneave, F. & Olson, R.K. (1971). Pitch as a medium. *American Journal of Psychology, 84*, 147-166.

Auhagen, W. (1983). *Studien zur Tonartencharakteristik in theoretischen Schriften und Kompositionen vom späten 17. bis zum Beginn des 20. Jahrhunderts* (= Europäische Hochschulschriften XXXVI,6). Frankfurt/Main.

Bachem, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America, 9,* 146-151.

Bachem, A. (1940). The genesis of absolute pitch. *Journal of the Acoustical Society of America, 11*, 434-439.

Bachem, A. (1954). Time factors in relative and absolute pitch determination. *Journal of the Acoustical Society of America, 26*, 751-753.

Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America,27*, 1180-1185.

Baddeley, A.D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A.D. & Hitch, G. (1974). Working memory. In G.H. Bower (Hg.), *The psychology of learning and motivation* (Bd. 8, S. 47-89). New York: Academic Press.

Baharloo, S., Johnston, P.A., Service, S.K., Gitschier, J. & Freimer, N.B. (1998). Absolute pitch: An approach for identification of genetic and nongenetic components. *American Journal of Human Genetics, 62,* 224-231.

Baharloo, S., Service, S.K., Risch, N., Gitschier, J. & Freimer, N.B. (2000). Familial aggregation of absolute pitch. *American Journal of Human Genetics, 67*, 755-758.

Barkowsky, J. (1992). Empirische Studie zum Vergleich von Absolut- und Relativhörern. In K.E. Behne, G. Kleinen & H. de la Motte-Haber (Hg.), *Musikpsychologie: Empirische Forschungen, ästhetische Experimente* (Bd. 9, 65-81). Wilhelmshaven: Noetzel.

Bartlett, J.C. & Dowling, W.J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance, 6*, 501-515.

Bartlett, J.C. & Snelus, P. (1980). Lifespan memory for popular songs. *American Journal of Psychology, 93*, 551-560.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychophysiological Bulletin, 91*, 276-292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J.T. Cacioppo, L.G. Tassinary & G.G. Berntson (Hg.), *Handbook of psychophysiology* (2. Aufl., S. 142-162). New York: Cambridge University Press.

Bergeson, T.R. & Trehub, S.E. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science, 13 (1)*, 72-75.

Bermudez, P. & Zatorre, R.J. (2001). Sexual dimorphism in the corpus callosum: methodological considerations in MRI morphometry. *NeuroImage, 13*, 1121-1130.

Berry, D.C. & Broadbent, D.E. (1985). On the relationship between task performance and associated verbalizable knowledge. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36A(2)*, 209-231.

Berz, W.L. (1995). Working Memory in Music: A Theoretical Model. *Music Perception, 12* (*3*), 353-364.

Besner, D. (1987). Phonology, lexical access in reading, and articulatory suppression: A critical review. *Quarterly Journal of Experimental Psychology, 39A*, 467-478.

Bigand, E. (1990). *Perception et compréhension des phrases musicales*. Ph.D. thesis, Université de Nanterre (Paris).

Bischoff, L.A. (2000). Recent discoveries in the psychophysiology of absolute pitch. In C. Woods, G. Luck, R. Brochard, F. Seddon & J.A. Sloboda (Hg.), *Proceedings of the 6th International Conference on Music Perception and Cognition*. Keele, UK: Dept. of Psychology, Keele University.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler* (4. Aufl.). Berlin, Heidelberg, New York: Springer.

Bortz, J. & Lienert, G.A. (1998). *Kurzgefasste Statistik für die klinische Forschung*. Berlin, Heidelberg, New York: Springer.

Braaten, R.F. Hulse, S.H. & Page, S.C. (1990). Discrimination and classification of rising and nonrising pitch patterns by the European starling (sturnus vulgaris). *Animal Learning and Behavior, 18*, 352-364.

Brady P.T. (1970). Fixed-scale mechanism of absolute pitch. *Journal of the Acoustical Society of America, 48*, 883-887.

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision, 10*, 433-436.

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon.

Brown, W.A., Cammuso, K., Sachs, H., Winklosky, B., Mullane, J., Bernier, R., Svenson, S., Arin, D., Rosen-Sheidley, B. & Folstein, S.E. (2003). Autism-related language, personality, and cognition in people with absolute pitch: Results of a preliminary study. *Journal of Autism and Developmental Disorders, 33*, 163-167.

Buchner, A. & Brandt, M. (2002). Gedächtniskonzeptionen und Wissensrepräsentationen. In J. Müsseler & W. Prinz (Hg.), *Allgemeine Psychologie* (S. 492-543). Heidelberg, Berlin: Spektrum Akademischer Verlag.

Burns, E.M. & Ward, W.D. (1978). Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*, 456-468.

Cahill, L. & McGaugh, J. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness & Cognition: An International Journal, 4(4)*, 410-421.

Cahill, L. & McGaugh, J. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences, 21(7)*, 294-299.

Carroll, J.B. (1975). Speed and accuracy of absolute pitch judgments: Some latter-day results. *Educational Testing Service Research Bulletin* (RB-75-35). Princeton, NJ: Educational Testing Service.

Carroll-Phelan, B. & Hampson, P.J. (1996). Multiple Components of the Perception of Musical Sequences: A Cognitive Neuroscience Analysis and Some Implications for Auditory Imagery. *Music Perception, 13 (4)*, 517-561.

Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B. & Stephenson, K.M. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development,* 61.

Cave, C.B. (1997). Very long-lasting priming in picture naming. *Psychological Science, 8 (4)*, 322-325.

Chase, W.G. & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chin, C.S. (2003). The development of absolute pitch: A theory concerning the roles of music training at an early developmental age and individual cognitive style. *Psychology of Music, 31*, 155-172.

Church, B.A. & Schacter, D.L. (1994). Perceptual specifity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20 (3),* 521-533.

Copp, E.F. (1916). Musical ability. *Journal of Heredity, 7*, 297-305.

Corliss, E.L.R. (1972). Remark on "Fixed-scale mechanism of absolute pitch" (Brady, 1970). *Journal of the Acoustical Society of America, 53,* 1737-1739.

Corso, J.F. (1957). Absolute judgments of musical tonality. *Journal of the Acoustical Society of America, 29 (1)*, 138-144.

Crowder, R.G. (1993). Auditory memory. In S. McAdams & E. Bigand (Hg.), *Thinking in Sound: The Cognitive Psychology of Human Audition* (S. 113-145). Oxford University Press.

Crowder, R.G., Serafine, M.L. & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition, 18,* 469-476.

Cynx, J., Hulse, S.H. & Polyzois, S. (1986). A psychophysical measure of pitch discrimination loss resulting from a frequency range constraint in European Starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology: Animal Behavior Processes, 12*, 394-402.

D'Amato, M.R. (1988). A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune. *Music Perception, 5,* 453-480.

Danko, G. (2005). *Befragung von Hochschullehrern zu Unterrichtsstrategien bei Musikstudierenden*. Unveröffentlichter Arbeitsbericht, HU Berlin.

Deese, V. & Grindley, G.C. (1947). The transposition of visual patterns. *British Journal of Psychology, 37*, 152-163.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in short-term memory. *Science, 168*, 1604-1605.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics, 11*, 411-412.

Deutsch, D. (1974). Generality of interference by tonal stimuli in recognition memory for pitch. *Quarterly Journal of Experimental Psychology, 26*, 229-234.

Deutsch, D. (1978). Delayed pitch comparisons and the principle of proximity. *Perception & Psychophysics, 23*, 227-230.

Deutsch, D. (1986). A musical paradox. *Music Perception, 3*, 275-280.

Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception, 8 (4)*, 335-347.

Deutsch, D. (1992). Paradoxien der Tonhöhenwahrnehmung. *Spektrum der Wissenschaft*, 82-88.

Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Hg.), *The psychology of music* (2. Aufl., S. 349-411). San Diego, CA: Academic Press.

Deutsch, D. & Boulanger, R.C. (1984). Octave equivalence and the processing of tonal sequences. *Music Perception, 3*, 40-51.

Deutsch, D., Henthorn, T. & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception, 21 (3)*, 339-356.

Deutsch, D., Kuyper, W.L. & Fisher, Y. (1987). The tritone paradox: Its presence and form of distribution in a general population. *Music Perception, 5,* 79-92.

Deutsch, D., North, T. & Ray, L. (1990). The tritone paradox: Correlate with the listener's vocal range for speech. *Music Perception, 7*, 371-384.

Donchin, E. (1981). Surprise! … Surprise? *Psychophysiology, 18*, 493-513.

Donchin, E. & Coles, M.G.H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*, 355-425.

Dowling, W.J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*, 342-354.

Dowling, W.J. & Bartlett, J.C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*, 30-49.

Dowling, W.J. & Fujitani, D.S. (1971). Contour, interval and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*, 524-531.

Dowling, W.J. & Harwood, D.L. (1986). *Music cognition*. Orlando, FL: Academic Press.

Dowling, W.J. & Hollombe, A.W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics, 21*, 60-64.

Drobisch, M.W. (1855). Über musikalische Tonbestimmung und Temperatur. *In Abhandlungen der Königlich sächsischen Gesellschaft der Wissenschaften zu Leipzig* (Bd. 4 *(2),* S. 3-121). Leipzig: S. Hirzel.

Fuster, J.M. (1995) *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.

Gigerenzer, G. (1985). Messung und Modellbildung. In H. Bruhn, R. Oerter & H. Rösing (Hg.), *Musikpsychologie* (S. 485-494). München: Urban & Schwarzenberg.

Goldman-Rakic, P.S. (1992). Working memory and the mind. *Scientific American, 267*, 110-117.

Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Granholm, E., Asarnow, R.F., Sarkin, A.J. & Dykes, K.L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33,* 457-461.

Granholm, E. & Steinhauer, S.R. (2004). Introduction: Pupillometric measures of cognitive and emotional processing. *International Journal of Psychophysiology, 52*, 1-6.

Hahn, K. (2002). Absolute memory for melody: A study with choir singers. In C. Stevens, D. Burnham, G. McPherson, E. Schubert & J. Renwick (Hg.), *Proceedings of the 7th international Conference on Music Perception and Cognition* (S. 514-517). Adelaide: Causal Productions.

Halpern, A.R. (1990). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17 (5)*, 572-581.

Halpern, A.R. & Zatorre, R.J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*, 697-704.

Halpern, A.R., Zatorre, R.J., Bouffard, M. & Johnson, J.A. (2002). An fMRI study on timbre perception and imagery. In C. Stevens, D. Burnham, G. McPherson, E. Schubert & J. Renwick (Hg.), *Proceedings of the 7th international Conference on Music Perception and Cognition* (S. 134-137). Adelaide: Causal Productions.

Hamann, S.B. & Squire, L.R. (1996). Level-of-processing effects in word-completion priming: A neuropsychological study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 933-947.

Hamilton, R.H., Pascual-Leone, A., Rodrigues, D. & Schlaug, G. (2000). Increased prevalence of absolute pitch in blind musicians. *Society for Neuroscience Abstracts, 26*, *Abstract No. 739.13*.

Harris, J.D. (1952). The decline of pitch discrimination with time. *Journal of Experimental Psychology, 43*, 96-99.

Heaton, P., Hermelin, B. & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception, 15*, 291-305.

Hespos, S. (1989). *The characteristics of pitch, timbre and loudness in auditory imagery*. Unpublished bachelor's thesis, Reed College, Portland, OR.

Hess, E.H. (1965). Attitude and pupil size. *Scientific American, 4*, 46-54.

Heyde, E.M. (1987). *Was ist absolutes Hören? Eine musikpsychologische Untersuchung*. München: Profil.

Hick, W.E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Hill, B.G. & Lein, M.R. (1987). Function of frequency-shifted songs of black capped chickadees. *The Condor, 89*, 914-915.

Hoeks, B. & Ellenbroek, B.A. (1993). A neural basis for a quantitative pupillary model. *Journal of Psychophysiology, 7(4)*, 315-324.

Hoeks, B., & Levelt, W.J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments & Computers, 25*(*1*), 16-26.

Hüppe, A. (1984). *Prägnanz. Ein gestaltpsychologischer Grundbegriff*. München: Profil.

Hulse, S.H. & Cynx, J. (1985). Relative pitch perception is constrained by absolute pitch in songbirds (*Mimus, Molothrus,* and *Sturnus*). *Journal of Comparative Psychology, 99*, 176-196.

Hulse, S. H., & Cynx, J. (1986). Interval and contour in serial pitch perception by a passerine bird, the European starling (*Sturnus vulgaris*). *Journal of Comparative Psychology, 100*, 215-228.

Hulse, S.H., Cynx, J. & Humpal, J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *Journal of Experimental Psychology: General, 113*, 38-54.

Hulse, S.H., Page, S.C. & Braaten, R.F. (1990). An integrative approach to auditory perception by songbirds. In M. Berkley & W.C. Stebbins (Hg.), *Comparative perception*. New York: Wiley.

Hulse, S.H., Takeuchi, A.H. & Braaten, R.F. (1992). Perceptual invariances in the comparative psychology of music. *Music Perception, 10*, 151-184.

Hurni-Schlegel, L. & Lang, A. (1978). Verteilung, Korrelate und Veränderbarkeit der Tonhöhen-Identifikation (sog. absolutes Musikgehör). *Schweizerische Zeitschrift für Psychologie, 37*, 265-292.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*, 423-432.

Hyönä, J., Tommola, J. & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48A(3*), 598-612.

Idson, W.L. & Massaro, D.W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics, 24*, 551-565.

Jorgenson, O.H. (1991). *Tuning*. East Lansing, MI: Michigan State University Press.

Just, M.J., & Carpenter, P.A. (1993). The intensity of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology. Special Issue: Reading and language processing, 47*(*2*), 310-339.

Kahneman, D. & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics, 2*, 101-105.

Karni, A. (1997). The acquisition of perceptual and motor skills: A memory system in the adult human cortex. *Cognitive Brain Research, 5*, 39-48.

Karni, A. & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Science, 88*, 4966-4970.

Keenan, J., Thangaraj, V., Halpern, A. & Schlaug, G. (2001). Absolute pitch and planum temporale*. Neuroimage 14*, 1402-1408.

Kestler, C. (1999). *Experimentelle Studie zur schnellen Tonhöhenkorrektur bei Sängern und bei Nichtsängern*. Unveröffentlichte Diplomarbeit, Hochschule für Musik und Theater, Hannover.

King, A.J., Schnupp, J.W.H. & Doubell, T.P. (2001). The shape of ears to come: Dynamic coding of auditory space*. Trends in Cognitive Sciences, 5*, 261-270.

Klein, M., Coles, M.G.H. & Donchin, E. (1984). People with absolute pitch process tones without producing a P300. *Science, 223*, 1306-1308.

Klix, F., van der Meer, E. & Preuß, M. (1984). Semantische Relationen: Erkennungsaufwand und psychophysiologische Reaktionstendenzen. In F. Klix (Hg.), *Gedächtnis – Wissen – Wissensnutzung* (S. 156-172). Berlin: Deutscher Verlag der Wissenschaften.

Koester, T. (1945). The time error in pitch and loudness discrimination as a function of time interval and stimulus level. *Archives of Psychology*, 297.

Kries, J.v. (1892). Über das absolute Gehör. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 3*, 257-279.

Krumhansl, C.L. (1991). Memory for musical surface. *Memory & Cognition, 19*, 401-411.

Krumhansl, C.L. & Shepard, R.N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception & Performance, 5*, 579-594.

Langner, G., Albert, M. & Briede, T. (2002). Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (*Chinchilla laniger*). *Hearing Research, 168,* 110-130.

Large, E.W., Palmer, C. & Pollack, J.B. (1995). Reduced memory representations for music. *Cognitive Science, 19*, 53-96.

Lenhoff, H.M., Perales, O. & Hickok, G. (2001). Absolute pitch in Williams syndrome. *Music Perception, 18*, 491-503.

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Levitin, D.J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56 (4)*, 414-423.

Levitin, D.J. (1999). Absolute pitch: Self-reference and human memory. *International Journal of Computing and Anticipatory Systems, 4*, 255-266.

Libby, W.L. Jr., Lacey, B.C. & Lacey, J.I. (1973). Pupillary and cardiac activity during visual attention. *Psychophysiology, 10*, 270-294.

Loftus, E.F., Miller, D.G. & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19-31.

Logothetis, N.K., Pauls, J. & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*, 552-563.

Marvin, E.W. & Brinkman, A.R. (2000). The effect of key color and timbre on absolute pitch recognition in musical contexts. *Music Perception, 18*, 111-137.

Meyer, M. (1899). Is the memory of absolute pitch capable of development by training? *Psychological Review, 6*, 514-516.

Miller, G.A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception & Psychophysics, 44*, 501-512.

Miyazaki, K. (1989). Absolute pitch identification: Effects of timbre and pitch region. *Music Perception, 7*, 1-14.

Miyazaki, K. (1990). The speed of musical pitch identification by absolute-pitch possessors. *Music Perception, 8*, 177-188.

Miyazaki, K. (1995). Perception of relative pitch with different references: Some absolute-pitch listeners can't tell musical interval names. *Perception & Psychophysics, 57*, 962-970.

Miyazaki, K. (2004). Recognition of transposed melodies by absolute-pitch possessors. *Japanese Psychological Research, 46*, 270-282.

Miyazaki, K. & Rakowski, A. (2002). Recognition of notated melodies by possessors and non-possessors of absolute pitch. *Perception & Psychophysics, 64*, 1337-1345.

Moody, C.B. & Stebbins, W.C. (1989). Salience of frequency modulation in primate communication. In R.J. Dooling & S.H. Hulse (Hg.), *The comparative psychology of audition: Perceiving complex sounds* (S. 353-376). Hillsdale, NJ: Erlbaum Associates.

Moody, C.B. & Stebbins, W.C. (1990). Auditory perception of communication signals by Japanese monkeys. In W.C. Stebbins & M.A. Berkley (Hg.), *Comparative perception* (Bd. 2, S. 311-343). New York: Wiley.

Mottron, L., Peretz, I. & Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: Beyond central coherence? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*, 1057-1065.

Mudd, S., Conway, C.G. & Schindler, D.E. (1990). The eye as music critic: Pupil response and verbal preferences. *Studia Psychologica, 32*, 23-30.

Müller, G.E. & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, Ergänzungsband 1*, 1-300.

Murry, T. (1990). Pitch matching accuracy in singers and non-singers. *Journal of Voice, 4*, 317-321.

Musen, G. & Squire, L.R. (1991). Normal acquisition of novel verbal information in amnesia. *Journal of Experimental Psychology: Learning, Memory & Cognition, 17(6)*, 1095-1104.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Erlbaum.

Nauck-Börner, C. (1984). Gedächtnisrepräsentation von Musik: Analoge oder aussagenartige Kodierung? In K.E. Behne, G. Kleinen & H. de la Motte-Haber (Hg.), *Musikpsychologie: Empirische Forschungen, ästhetische Experimente* (Bd. 1, 93-101). Wilhelmshaven: Noetzel.

Nauck-Börner, C. (1988). Strukturen des musikalischen Gedächtnisses. Anmerkungen zu formalen Modellen der Repräsentation. In K.E. Behne, G. Kleinen & H. de la Motte-Haber (Hg.), *Musikpsychologie: Empirische Forschungen, ästhetische Experimente* (Bd. 5, 55-66). Wilhelmshaven: Noetzel.

Nettl, B. (1956). *Music in primitive culture*. Cambridge, MA: Harvard University Press.

Nuthmann, A. (2002). *Die aufgabenbezogene Pupillenreaktion als Indikator für kognitiven Verarbeitungsaufwand. Untersuchungen zur Kodierung der zeitlichen Vorzugsrichtung innerhalb der Struktur eines Ereignisbegriffes im semantischen Gedächtnis*. Unveröffentlichte Diplomarbeit, Humboldt Universität zu Berlin.

Nuthmann, A. & van der Meer, E. (2005). Time's arrow and task-evoked pupillary response. *Psychophysiology, 42*, 306-317.

Oakes, W.F. (1955). An experimental study of pitch naming and pitch discrimination reactions. *Journal of Genetic Psychology, 86*, 237-259.

Olsen, R.K. & Hanson, V. (1977). Interference effects in tone memory. *Memory & Cognition, 5*, 32-40.

Page, S.C., Hulse, S.H. & Cynx, J. (1989). Relative pitch perception in the European starling (*Sturnus vulgaris*): Further evidence for an elusive phenomenon. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 137-146.

Pantev, C., Engelien, A., Candia, V. & Elbert, T. (2003). Representational cortex in musicians. In I. Peretz & R.J. Zatorre (Hg.), *The cognitive neuroscience of music* (366-381). Oxford University Press.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811-814.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12*, 169-174.

Paulesu, E., Frith, C.D. & Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of short-term memory. *Nature, 362*, 342-345.

Peavler, W.S. (1974). Pupil size, information overload, and performance differences. *Psychophysiology, 11*, 559-566.

Pechmann, T. & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition, 20*, 314-320.

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Petran, L.A. (1932). An experimental study of pitch recognition. *Psychological Monographs, 42*, 1-120.

Piaget, J. (1950). *The psychology of intelligence*. London: Routledge & Kegan Paul.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America, 24*, 745-749.

Prinz, W. (1966). Quantitative Versuche über die Prägnanz von Punktmustern. *Psychologische Forschung, 29*, 297-359.

Rakowski, A. (1972). Direct comparison of absolute and relative pitch. In F.A. Bilsen (Hg.), *Symposium on hearing theory* (S. 105-108). Eindhoven, Niederlande: Instituut voor Perceptie Underzoek.

Rakowski, A. (1978). Investigations of absolute pitch. In E.P. Asmus (Hg.), *Proceedings of the Research Symposium on the Psychology and Acoustics of Music* (S. 45-57). Lawrence: University of Kansas, Division of Continuing Education.

Rakowski, A. (1994). Investigating short term auditory memory with the method of vocal pitch control. In *Proceedings of the Stockholm Music Acoustics Conference*, 53-57.

Rakowski, A. & Morawska-Büngeler, M. (1987). In search for the criteria of absolute pitch. *Archives of Acoustics, 12*, 75-87.

Rao, S.M., Binder, R.J., Bandettini, P.A., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Lisk, L.M., Morris, G.L., Mueller, W.M., Estkowski, L.D., Wong, E.C., Haughton, V.M. & Hyde, J.S. (1993). Functional magnetic resonance imaging of complex human movements. *Neurology, 43*, 2311-2318.

Ratcliffe, L. & Weisman, R.G. (1985). Frequency shift in the fee bee song of the black-capped chickadee. *The Condor, 87*, 555-556.

Rausch, E. (1966). Das Eigenschaftsproblem in der Gestalttheorie der Wahrnehmung. In W. Metzger (Hg.), *Handbuch der Psychologie*, I.1. Göttingen: Hogrefe.

Rauschecker, J.P. (2001). Cortical Plasticity and Music. In R. Zatorre & I. Peretz (Hg.), *The biological foundations of music,* Annals of the New York Academy of Science (Bd. 930, S. 330-336).

Reber, P.J. & Squire, L.R. (1998). Encapsulation of implicit and explicit memory in sequence learning. *Journal of Cognitive Neuroscience, 10(2)*, 248-263.

Révész, G. (1913). *Zur Grundlegung der Tonpsychologie*. Leipzig: Veit.

Révész, G. (1946). *Einführung in die Musikpsychologie*. Bern: A. Francke.

Riker, B.L. (1946). The ability to judge pitch. *Journal of Experimental Psychology, 36*, 331-346.

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S.A. & Neville, H.J. (1999). Improved auditory spatial tuning in blind humans. *Nature, 400*, 162-166.

Rössger, P. (1997). *Die Entwicklung der Pupillometrie zu einer Methode der Messung mentaler Beanspruchung in der Arbeitswissenschaft*. Sinzheim: Pro Universitate Verlag.

Roland, P.E., Larsen, B., Lassen, N.A. & Skinhoj, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology, 43*, 118-136.

Ross, D.A., Olson, I.R., Marks, L.E. & Gore, J.C. (2004). A nonmusical paradigm for identifying absolute pitch possessors. *Journal of the Acoustical Society of America, 116(3)*, 1793-1799.

Saffran, J.R. (2003). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science, 6*, 35-47.

Saffran, J.R. & Griepentrog, G.J. (2001). Absolute pitch in infant auditory learning : Evidence for developmental reorganization. *Developmental Psychology, 37*, 74-85.

Salamé, P. & Baddeley, A.D. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology, 41A*, 107-122.

Salasoo, A., Shiffrin, R.M. & Feustel, T.C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114*, 50-77.

Sameroff, A.J. & Haith, M.M. (1996). *The five to seven year shift: The age of reason and responsibility*. Chicago: University of Chicago Press.

Schacter, D.L. & Tulving, E. (1994). What are the memory systems of 1994? In D.L. Schachter & E. Tulving (Hg.), *Memory systems 1994* (S. 1-38). Cambridge, MA: MIT Press.

Schellenberg, E.G. & Trehub, S.E. (2003). Good pitch memory is widespread. *Psychological Science, 14 (3)*, 262-266.

Schlaug, G., Jäncke, L., Huang, Y. & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267*, 699-701.

Scoville, W.B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry, 20*, 11-21.

Semal, C. & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America, 89*, 2404-2410.

Semal, C. & Demany, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America, 94*, 1315-1322.

Serafine, M.L., Davidson, J., Crowder, R.G. & Repp, B.H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language, 25*, 123-135.

Sergeant, D. (1969). Experimental investigation of absolute pitch. *Journal of Research in Music Education, 17*, 135-143.

Sergeant, D.C. & Roche, S. (1973). Perceptual shifts in the auditory information processing of young children. *Psychology of Music, 1*, 39-48.

Shepard, R.N. (1964). Circularity in jugdments of relative pitch. *Journal of the Acoustical Society of America, 36 (12)*, 2346-2353.

Siegel, J.A. (1972). The nature of absolute pitch. In E. Gordon (Hg.), *Experimental research in the psychology of music: VIII. Studies in the psychology of music* (S. 65-89). Iowa City: Iowa University Press.

Siegel, J.A. (1974). Sensory and verbal coding strategies in subjects with absolute pitch. *Journal of Experimental Psychology, 103*, 37-44.

Siegler, R.S. (1996). Unidimensional Thinking, multidimensional thinking, and characteristic tendencies of thought. In A.J. Sameroff & M.M. Haith (Hg.), *The five to seven year shift: The age of reason and responsibility* (S. 63-84). Chicago: University of Chicago Press.

Simpson, J. & Huron, D. (1994). Absolute pitch as a learned phenomenon: Evidence consistent with the Hick-Hyman Law. *Music Perception, 12*, 267-270.

Sloboda, J.A. & Parker, D.H.H. (1985). Immediate recall of melodies. In I. Cross, P. Howell & R. West (Hg.), *Musical Structure and Cognition* (S. 143-167). London: Academic Press.

Smith, S.M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 460-471.

Smith, J.D., Reisberg, D. & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In D. Reisberg (Hg.), *Auditory Imagery* (S. 95-119). Hillsdale, NJ: Lawrence Erlbaum.

Spitzer, M. (1996). *Geist im Netz*. Spektrum Akademischer Verlag, Heidelberg.

Spitzer, M. (2002). *Musik im Kopf*. Stuttgart: Schattauer.

Squire, L.R. (1986). Mechanisms of memory. *Science, 232*, 1612-1619.

Squire, L.R. & Kandel, E.R. (1999). *Memory. From mind to molecules*. New York: Scientific American Library.

Squire, L.R. & Frambach, M. (1990). Cognitive skill learning in amnesia. *Psychobiology, 18(1)*, 109-117.

Stadler, M., Stegagno, L. & Trombini, G. (1979). Quantitative Analyse der Rauschschen Prägnanzaspekte. *Gestalt Theory, 1*, 39-51.

Steinhauer, S.R., Boller, F., Zubin, J. & Pearlman, S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology, 20*, 472.

Stumpf, C. (1883). *Tonpsychologie* (Bd. 1). Leipzig: S. Hirzel.

Takeuchi, A.H. & Hulse, S.H. (1991). Absolute pitch judgments of black and white-key pitches. *Music Perception, 9*, 27-46.

Takeuchi, A.H. & Hulse, S.H. (1993). Absolute pitch. *Psychological Bulletin, 113*, 345-361.

Terhardt, E. & Seewann, M. (1983). Aural key identification and its relationship to absolute pitch. *Music Perception, 1*, 63-83.

Terhardt, E. & Ward, W.D. (1982). Recognition of musical key: Exploratory study. *Journal of the Acoustical Society of America, 72 (1)*, 26-33.

Tervaniemi, M., Alho, K., Paavilainen, P., Sams, M. & Näätänen, R. (1993). Absolute pitch and event-related brain potentials. *Music Perception, 10*, 305-316.

Trehub, S.E. (2003). Absolute and relative pitch processing in tone learning tasks (Commentary). *Developmental Science, 6*, 44-45.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Hg.), *Organization of memory* (S. 382-403). New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon.

Ungerleider, L.G. (1995). Functional brain imaging studies of cortical mechanisms for memory. *Science, 270*, 769-775.

van der Meer, E. (im Druck). Langzeitgedächtnis. In P. Frensch & J. Funke (Hg.), *Handbuch Allgemeine Psychologie*. Göttingen: Hogrefe.

van der Meer, E., Friedrich, M., Nuthmann, A., Stelzel, C., & Kuchinke, L. (2003). Picture-word-matching: Flexibility in conceptual memory and pupillary responses. *Psychophysiology, 40*, 904-913.

van Krevelen, A. (1951). The ability to make absolute judgments of pitch. *Journal of Experimental Psychology, 42*, 207-215.

Vazan, P. (2000). Memory for music: An overview. *Systematische Musikwissenschaft, 7*, 7-31.

Verney, S.P., Granholm, E., & Dionisio, D.P. (2001). Pupillary responses and processing resources on the visual backward masking task. *Psychophysiology, 38*(*1*)*,* 76-83.

Vitouch, O. (2005). Absolutes Gehör. In T.H. Stoffer & R. Oerter (Hg.), *Allgemeine Musikpsychologie* (Enzyklopädie der Psychologie, Vol. D/VII/1, S. 717-766). Göttingen: Hogrefe.

Vitouch, O. & Gaugusch, A. (2000). Absolute recognition of musical keys in non-absolute-pitch-posessors. In C. Woods, G. Luck, R. Brochard, F. Seddon & J.A. Sloboda (Hrsg.), *Proceedings of the 6th Int. Conference on Music Perception and Cognition* [CD-ROM]. Keele, UK: Dept. of Psychology, Keele University.

von Ehrenfels, C. (1890). Über Gestaltqualitäten. *Vierteljahresschrift für Wissenschaftliche Philosophie, 14*, 249-292.

Wallace, J.D. (1985). *An investigation of extrinsic laryngeal muscle responses to auditory stimulation*. Denton, TX: Unpublished doctoral dissertation, University of North Texas.

Ward, W.D. (1990, Mai). *Relative versus absolute pitch and the key of auralized melodies*. Vortrag beim Karajan Symposium, Wien (berichtet in Ward, 1999).

Ward, W.D. (1999). Absolute pitch. In D. Deutsch (Hg.), *The psychology of music* (2. Aufl., S. 265-298). San Diego, CA: Academic Press.

Ward, W.D. & Burns, E.M. (1978). Singing without auditory feedback. *Journal of Research in Singing & Applied Vocal Pedagogy, 1*, 24-44.

Weinberger, N.M. (1993). Learning-induced changes of auditory receptive fields. *Current Opinion in Neurobiology, 3*, 570-577.

Weinberger, N.M. (1999). Music and the auditory system. In D. Deutsch (Hg.), *The psychology of music* (2. Aufl., S. 47-87). San Diego, CA: Academic Press.

Weisman, R.G., Ratcliffe, L., Johnsrude, I.S. & Hurly, T.A. (1990). Absolute and relative pitch production in the song of the black-capped chicadee (*Parus atricapillus*). *Condor, 92*, 118-124.

Welch, G.F. (1988). Observations on the incidence of absolute pitch ability in the early blind. *Psychology of Music, 16*, 77-80.

White, D.J., Dale, P.S. & Carlsen, J.C. (1990). Discrimination and Categorization of pitch direction by young children. *Psychomusicology, 9*, 39-58.

Wickelgren, W.A. (1966). Consolidation and retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology, 72*, 250-259.

Wickelgren, W.A. (1969). Associative strength theory of recognition memory for pitch. *Journal of Mathematical Psychology, 6*, 13-61.

Wilson, S.J., O'Keefe, G. & Reutens, D.C. (2002). Using absolute pitch to examine brain plasticity and cognitive skill development. In C. Stevens, D. Burnham, G. McPherson, E.

Schubert, J. Renwick (Hg.), *Proceedings of the 7th International Conference on Music Perception and Cognition*. Adelaide: Causal Productions.

Winkelhaus, E. (2004). *Zur kognitionspsychologischen Begründung einer systematischen Melodielehre* (= Systemische Musikwissenschaft Bd. 7). Frankfurt am Main: Peter Lang Verlag.

Wise, R.J., Chollet, F., Hadar, U., Friston, K., Hoffner, E. & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain, 114*, 1803-1817.

Witelson, S. & Pallie, W. (1973). Left hemisphere specialization for language in the newborn: Neuroanatomical evidence of asymmetry. *Brain, 96*, 641-646.

Wyke, B.D. (1974). Laryngeal reflex mechanisms in phonation. *Proceedings of the XVIth International Congress on Logopedics and Phoniatrics* (S. 528-537). Basel: Karger.

Wynn, V.T. (1972). Absolute pitch revisited. *British Journal of Psychology, 83*, 129-131.

Wynn, V.T. (1973). Accuracy and consistency of absolute pitch. *Perception, 22*, 113-121.

Zajonc, R.B. & Hazel, M. (1982). Affective and cognitive factors in preferences. *Journal of Consumer Research, 9*, 123-131.

Zatorre, R.J. (1989). Intact absolute pitch ability after left temporal lobectomy. *Cortex, 25*, 567-580.

Zatorre, R.J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience, 6*, 692-695.

Zatorre, R.J. & Beckett, C.A. (1989). Multiple coding strategies in the retention of musical tones by possessors of absolute pitch. *Memory & Cognition, 17*, 582-589.

Zatorre, R.J. & Halpern, A.R. (1979). Identification, discrimination and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics, 26*, 384-395.

Zatorre, R.J., Halpern, A.R., Perry, D.W., Meyer, E. & Evans, A.C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience, 8*, 29-46.

Zatorre, R.J., Perry, D.W., Beckett, C.A., Westbury, C.F. & Evans, A.C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences, 95*, 3172-3177.