

Zur Tonhöhenwahrnehmung von oktav-komplexen Tönen

Psychophysik, psychoakustische Theorie und
computationale Modellierung

Dissertation

zur Erlangung des
Doktorgrades der Philosophie (Dr. phil.)

vorgelegt

der Philosophischen Fakultät I
der Martin-Luther-Universität Halle-Wittenberg,
Fachbereich Sozialwissenschaften und historische Kulturwissenschaften

von

Sebastian Krüger
geb. am 4.5.1975 in Gardelegen

Erstgutachter: Prof. Dr. J. Lukas
Zweitgutachter: PD Dr. habil S. Blankenberger
Tag der mündlichen Verteidigung: 15.12.2011

*Wichtige Dinge nur halb zu tun, ist
nahezu wertlos; denn meistens ist es die
andere Hälfte, die zählt.*

Emil Oesch

Danksagung

Ich danke Prof. Dr. Josef Lukas, der mir seit Beginn des Studiums der Psychologie ein besonderer Lehrer und Mentor war und stets ein offenes Ohr hatte. Dank gilt auch meinem Freund und Kollegen Dr. Michael Hanke für die Unterstützung bei der Entwicklung der verwendeten Software, für die zahlreichen motivierenden Kaffees und die nötige Ablenkung in den richtigen Augenblicken. PD Dr. Sven Blankenberger, Dr. Sandra Fritsche und Dr. Sven Nicklas danke ich für die Unterstützung in inhaltlichen wie in methodischen Fragen. In diesem Zuge sei auch Prof. Dr. Dieter Heyer erwähnt, ohne Ihn wäre die 'andere Hälfte' wohl nie geschrieben worden. Dank gilt weiterhin den zahlreichen Versuchspersonen für die Teilnahme an den mitunter „spannenden“ Experimenten und für die Beurteilung von über 100000 Tonpaaren und der Open-Source-Community für die Arbeit an und mit freier Software (R, Qt, Psignifit, Dev-Cpp, MinGW, Praat, AMS, L^AT_EX, ...), ohne die diese Arbeit nicht möglich gewesen wäre. Spezieller Dank gilt meiner Frau, meinen Eltern, meiner Familie und meinen Freunden für die Gewissheit, dass immer jemand da ist, wenn man ihn braucht.

Sebastian Krüger
Kalbe (Milde), März 2011

Inhaltsverzeichnis

1	Zusammenfassung	1
2	Einleitung	3
3	Tonhöhe: Einordnung und Abgrenzung	4
3.1	Tonhöhe	4
3.1.1	Definition von Tonhöhe	5
3.1.2	Existenzbedingungen	6
3.1.3	Eindeutigkeit der Tonhöhe	7
3.1.4	Messung von Tonhöhe	7
3.1.5	Einflussgrößen	10
4	Theorie der Tonhöhenwahrnehmung	13
4.1	Überblick	13
4.1.1	Zur Vorhersage von Tonhöhe: Ort und Zeit	15
4.2	Ortstheorien	16
4.2.1	Auf dem Weg zum Algorithmus	16
4.2.2	Das Modell der virtuellen Tonhöhe	19
4.2.3	Zum Konzept der virtuellen Tonhöhe	21
4.3	Zeittheorien	23
4.3.1	Autokorrelation	24
4.3.2	Das Unitary Pitch Modell	26
4.3.3	Das Auditory Image Modell	30
4.4	Integration von Orts- und Zeittheorien	32
4.5	Theorien zur binauralen Tonhöhenwahrnehmung	35
5	Zur Psychophysik oktav-komplexer Töne	40
5.1	Tonhöhe oktav-komplexer Töne	40
5.1.1	Tonhöhe und Chroma	40

5.1.2	Reizbeschreibung	42
5.1.3	Tonhöhenwahrnehmung bei oktav-komplexen Tönen	44
5.1.4	Das Tritonus Paradox	46
5.1.5	Einflussfaktoren	52
5.2	Psychometrische Modellierung	57
5.2.1	Das Modell des subjektiven Tonklassenkreises	57
5.2.2	Das P4-Modell	60
5.3	Modellierung der Tonhöhe oktav-komplexer Töne	63
6	Experimentelle Untersuchungen	70
6.1	Vorexperiment I:	
	Hörschwellenaudiometrie	73
6.1.1	Zur adaptiven Messmethodik	73
6.1.2	Methode	74
6.1.3	Ergebnisse	78
6.1.4	Diskussion	83
6.2	Vorexperiment II:	
	Tonhöhendiskrimination	86
6.2.1	Methode	86
6.2.2	Ergebnisse	88
6.2.3	Diskussion	89
6.3	Zur Stabilität der Urteile bei der Tritonusaufgabe	92
6.3.1	Fragestellung	92
6.3.2	Methode	93
6.3.3	Ergebnisse	94
6.3.4	Diskussion	99
6.4	Vollständiger Paarvergleich zum Einfluss der Teiltonanzahl . . .	101
6.4.1	Fragestellung	101
6.4.2	Methode	102
6.4.3	Ergebnisse	105
6.4.4	Diskussion	114
6.5	Vollständiger Paarvergleich zum Einfluss der Hüllenposition und -form	117
6.5.1	Fragestellung	117
6.5.2	Methode	118

6.5.3	Ergebnisse	120
6.5.4	Diskussion	125
6.6	Dichotische Darbietung von oktav-komplexen Tönen	127
6.6.1	Fragestellung	127
6.6.2	Methode	132
6.6.3	Ergebnisse	135
6.6.4	Diskussion	138
6.7	Dichotische Darbietung II: Erweiterung um das Tritonusintervall	141
6.7.1	Fragestellung	141
6.7.2	Methode	141
6.7.3	Ergebnisse	142
6.7.4	Diskussion	147
6.8	Zusammenfassung und Diskussion der experimentellen Befunde	149
6.9	Ausblick	156
 Literaturverzeichnis		 158
 Anhang		 169

1 Zusammenfassung

In der vorliegenden Dissertation werden spezifische Fragestellungen bei der Wahrnehmung von oktav-komplexen Tönen, sog. Shepard-Tönen, untersucht. Den Kern der Arbeit bilden sieben experimentelle Studien zur Tonhöhenwahrnehmung bei Shepard-Tönen.

Ausgehend von einer allgemeinen Definition, Einordnung und Abgrenzung der Tonhöhe als Empfindungsgröße in Kapitel 3 wird in Kapitel 4 auf die bestehenden Theorien zur Tonhöhenwahrnehmung eingegangen. Dabei werden die Eigenarten und Besonderheiten von Orts- und Zeittheorien anhand jeweils eines einschlägigen Vertreters beider Kategorien näher beschrieben.

In Kapitel 5 werden oktav-komplexe Reize definiert und deren Paradoxien, wie z.B. zirkuläre Urteile oder das Tritonus-Paradox, beschrieben und anhand einschlägiger und alternativer Theorien erklärt. Auch wird an dieser Stelle auf die computationale Modellierung der Antworten aus den Experimenten eingegangen. Dazu werden zwei Modelle vorgeschlagen, gegenübergestellt und angewendet.

Beginnend mit einer kurzen Einführung in die zirkuläre Statistik, sind in Kapitel 6 die einzelnen Experimente dargestellt. Die beiden Vorexperimente dienen zur Überprüfung der Hörfähigkeit der Probanden und zur Erhebung der Fähigkeit zur Diskrimination von Tonhöhe. Hierbei wird neben den inhaltlichen Fragestellung auch auf die besondere adaptive Methodik und die Effizienz der verwendeten Algorithmen zur Schwellenschätzung eingegangen.

Das Experiment in Abschnitt 6.3 ist eine Replik der vielfach zitierten Tritonusaufgabe. Dabei war die Kontrolle der experimentellen Bedingungen, speziell der isolierten Darbietung von Tritonusintervallen und der vollständig randomisierten Reizdarbietung, von wesentlicher Bedeutung für die Interpretation der Ergebnisse. Es konnte gezeigt werden, dass sich der vermutete Zusammenhang zwischen Sprachsozialisation bzw. -gemeinschaft und der Beurteilung von Tonhöhen oktav-komplexer Töne, bekannt als Tritonusparadoxon, nicht finden lässt.

In den Abschnitten 6.4 und 6.5 werden zwei Paarvergleichsexperimente beschrieben, welche den Einfluss verschiedener Parameter (Anzahl der Teiltöne, Position und Form der Hüllkurve) auf die Beurteilung der Tonhöhe untersuchen. Als Erklärung werden zwei Strategien (Chroma- vs. Fundamentalurteil) zur Beurteilung der Tonhöhe von oktav-komplexen Tönen gegenübergestellt. Es kann gezeigt werden, dass beide Strategien fließend ineinander übergehen. Im Nachgang werden individuelle Parameterschätzungen zur Auswertung der experimentellen Befunde verwendet.

Zwei Experimente zur binauralen Tonhöhenwahrnehmung werden in den Abschnitten 6.6 und 6.7 beschrieben. Dabei werden oktav-komplexe Töne in drei Bedingungen so aufgeteilt, dass die Beantwortung der Frage nach dem Ort der Tonhöhenbeurteilung (zentral vs. peripher) und nach der Art des Mechanismus (Fusion vs. Suppression) beantwortet werden kann. Die Ergebnisse weisen auf einen zentralen Mechanismus hin, welcher die lateralen Teilreize fusioniert.

2 Einleitung

Sinnestäuschungen, Paradoxien und Trugwahrnehmungen wohnt seit jeher eine Faszination inne, da sie einen scheinbar unauflösbaren oder tatsächlichen Widerspruch zu dem Erwarteten eröffnen. Die in dieser Arbeit untersuchten oktav-komplexen Töne bilden die Grundlage für zahlreiche akustische Phänomene. Darunter ist das wohl bekannteste das von Shepard (1964) demonstrierte Phänomen der unendlich aufsteigenden Tonhöhe. Paradox dabei ist, dass wiederholt die gleichen Töne abgespielt werden. Hier entsteht der Widerspruch zwischen der Auffassung von einer begrenzten, linearen Tonhöhe von tief nach hoch und der nicht endenden Tonhöhenänderung im Perzept.

Ein weiteres Phänomen, welches durch oktav-komplexe Töne hervorgerufen werden kann, ist das Tritonusparadoxon (Deutsch, 1974). Dabei wird die Tonhöhenänderung (auf- bzw. absteigend) eines Tritonusintervalls entgegen der Erwartung nicht eher zufällig, sondern individuell verschieden und stabil beurteilt. So kann z.B. die Situation entstehen, dass ein derartiges Tonpaar von zwei Hörern gegensätzlich beurteilt wird, wobei beide ihrer Entscheidung sicher sind.

Eine weiteres akustisches Paradoxon wurde im Rahmen dieser Arbeit entwickelt. Man stelle sich die folgende Situation vor: Über einen Kopfhörer wird ein Tonpaar präsentiert. Erst auf dem einen, dann auf dem anderen Ohr. Auf die Frage, wie sich die Tonhöhe vom ersten zum zweiten Ton ändert, wird in beiden Fällen sicher mit z.B. 'aufsteigend' geantwortet. Werden nun beide 'aufsteigenden' Paare gemeinsam präsentiert, kehrt sich das Urteil um. Das Tonpaar wird nun 'absteigend' wahrgenommen.

Sinnestäuschungen oder Illusionen sind keine Fehler des Wahrnehmungssystems, sie lassen sich nicht berichtigen. Sie sind 'normale' Reaktionen eines Systems auf ungewöhnliche, meist künstliche Stimuli und bieten gerade dadurch das Potential, bestehende Theorien zu hinterfragen.

3 Tonhöhe: Einordnung und Abgrenzung

In der folgenden Einführung soll in aller Knappheit die theoretische Basis für ein Verständnis der später vorgestellten experimentellen Arbeiten gelegt werden. Um die Frage zu beantworten, was Töne und Klänge mit Psychologie zu tun haben, soll nun die eigentlich interessierende psychischen Größe vorgestellt werden, die Tonhöhe.

3.1 Tonhöhe

Das Hören ist die älteste Form der Wahrnehmung unserer Umwelt und seither durchzogen von Phänomenen und Paradoxien im Dunstkreis von Sprache und Musik. In dem wissenschaftlichen Themenraum, welcher diese Phänomene umreißt, der Psychoakustik, steht die Beschreibung und Modellierung des Zusammenhangs zwischen Schallereignissen und der daraus resultierenden Empfindung, den Hörereignissen, im Mittelpunkt. Ist also das Ziel die Beschreibung von Schallereignissen wie z.B. Musik, Sprache, Töne, Klänge oder Geräusche, so beschränkt sich das psychologische Interesse auf empfindbare Schalle im Gegensatz zu nicht empfindbaren Schallen wie Infra- oder Ultraschall. Diese Abgrenzung mag trivial erscheinen, demonstriert sie doch die Unterschiedlichkeit von physikalisch zu beschreibenden Schallereignissen und psychologisch zu beschreibenden Hörereignissen.

Versucht man Hörempfindungen zu beschreiben und damit greifbar und klassifizierbar zu machen, so gehört die Lautheit mit der Unterscheidung zwischen laut und leise zu den Dimensionen, welche Kinder bereits mit dem ersten Lebensjahr beherrschen und somit wohl zu den prägnantesten Wahrnehmungsmerkmalen (Wilkening & Krist, 2002). Eine nicht weniger nachvollziehbare Unterteilung liefert die Tonhöhe mit den Attributen hoch oder tief. Auch die Tatsache, dass es uns keine Schwierigkeiten bereitet, einen Klang oder ein Ge-

räusch als angenehm oder unangenehm zu beschreiben, mag individuellen oder gar kulturellen Einflüssen unterliegen, gibt jedoch Hinweise auf eine weitere Dimension zwischen Wohlklang und Dissonanz. Bedient man sich des Vokabulars der Musik, so lassen sich noch andere Größen wie z.B. Rauheit oder Brassigkeit finden um Klänge zu beschreiben (Hall, 1991).

Terhardt (1998) unterteilt die Hörempfindungen in zwei große Gruppen, die *prothetischen* und die *metathetischen* Empfindungen. Zu den prothetischen Aspekten des Hörens zählen alle Hörempfindungen, die mit „Wie viel, Wie stark“ erfragt werden, wie z.B. die Lautheit, Volumen, Schärfe oder Klanghaftigkeit. Zu den durch „Was bzw. Wo“ erfragten metathetischen Aspekten des Hörens gehören neben der Tonhöhe auch alle klanglichen Wahrnehmungsattribute wie z.B. die Klangfarbe oder Sonanz.

Abhebend von der Vielzahl der bereits genannten Dimensionen der Hörempfindungen, soll nun näher auf die Tonhöhe eingegangen werden. Was ist Tonhöhe und wie lässt sie sich definieren?

3.1.1 Definition von Tonhöhe

Das deutsche Institut für Normung (DIN) definiert Tonhöhe unter DIN 1320 (Grundbegriffe der Akustik) als die Zuordnung eines Tones oder Tongemisches zu einer Skala zwischen tiefen und hohen Tönen. Ähnlich dieser Definition beschreibt auch der amerikanische Standard (ANSI) Tonhöhe, allerdings wird hier explizit auf eine musikalische Skala verwiesen: „Pitch may be defined as that attribute of auditory perception in terms of which sounds may be ordered on a musical scale extending from low to high“ (Moore, 2003, S. 195). Während die deutsche und die amerikanische Definition Tonhöhe als eine rein psychologische Größe definieren, welche auf einer Skala von tief nach hoch eingeordnet wird, assoziiert das französische Institut für Normung (AFNOR) Tonhöhe mit der Frequenz (de Cheveigné, 2004). Dieser psychophysischen Definition zufolge wird ein Ton in Abhängigkeit von einer größeren oder kleineren Frequenz als höher oder tiefer empfunden.

Eine zugleich treffende und unverfängliche Einordnung liefert Terhardt (1998): „Wenn man die Tonhöhe als diejenige Hörempfindung bezeichnet, welche den Aspekt der Periodizität und der Periodendauer der Schallsignale repräsentiert, so trifft man durchaus den Kern der Sache.“ (S. 307). Gleichwohl fügt er an, dass sich die Beziehungen zwischen Tonhöhe und Periodizität weitaus komplizierter erweisen, als dies auf den ersten Blick erscheint.

3.1.2 Existenzbedingungen

Vorausgesetzt ein Schallereignis enthält Frequenzen im wahrnehmbaren Bereich zwischen ca. 20 und 20000 Hz, so muss dieser Reiz, um eine auditive Empfindung auszulösen, die absolute Wahrnehmungsschwelle überschreiten. Mit anderen Worten, ein Reiz muss laut genug sein, um wahrgenommen zu werden. Dabei besteht zwischen der bloßen Wahrnehmung (Detektion) eines Tones und dem Auslösen einer Tonhöhenempfindung ein Unterschied (Terhardt, 1998). Die Tonschwelle, welche etwa 1-3 dB über der absoluten Hörschwelle liegt, beschreibt die untere Grenze der Tonhöhe. Befindet sich der Schallpegel eines Tones zwischen Hörschwelle und Tonschwelle kann er zwar detektiert werden, seine Tonhöhe bleibt jedoch undefiniert.

Der Vollständigkeit halber sei hier auch erwähnt, dass der entgegengesetzte Fall, nämlich ein zu lauter Reiz, ebenfalls mit einer indifferenten Tonhöhe wahrgenommen wird, da bei zu hoher Lautstärke die Schmerzschwelle erreicht wird.

Weiterhin sei darauf hingewiesen, dass das alleinige Vorhandensein von hinreichender Lautstärke und Frequenz kein Garant für die Empfindbarkeit eines Schallereignisses ist. So kann z.B. ein Ton durch einen zweiten Ton überlagert werden. Hierbei spricht man von Maskierung. Maskierte Töne sind nicht wahrnehmbar und erzeugen demnach keine Tonhöhe (Moore, 2003). Dieses Phänomen tritt bei der Interaktion mehrere Schalle innerhalb eines bestimmten Frequenzbereiches, der sog. kritischen Bandbreite auf. Zwicker (1954) konnte zeigen, dass zwar die ursprüngliche Tonhöhe des maskierten Reizes nicht mehr wahrnehmbar ist, ein Residualschall als Produkt der Maskierung jedoch detektierbar bleibt.

Bei den Überlegungen zur Existenz von Tonhöhe stellt sich auch die Frage nach der minimalen Tondauer, die notwendig ist, um die Wahrnehmung von Tonhöhe auszulösen. Erstaunlicherweise reicht die Darbietung einer einzigen Periode aus, um einen 1kHz Ton von einem 2kHz Ton zu unterscheiden (Hartmann & Zhang, 2003). Allerdings kann bei dieser Diskriminationsaufgabe noch nicht von Tonhöhe gesprochen werden. Doughty und Garner (1948) unterscheiden abhängig von der Dauer zwei Schwellen, „click-pitch-threshold“ und „tone-pitch-threshold“. Dabei stellt die erste Schwelle die Dauer eines Tones dar, die notwendig ist, um einen Frequenzbereich identifiziert zu können (z.B. bei einem 1kHz Ton 5 ms). Ein Reiz wird hier als ein „Klick“ wahrgenommen. Ab der zweiten Schwelle erreicht der Ton die Qualität eines längeren statischen Tones mit einer ausgeprägten Tonhöhe (z.B. bei einem 1000 Hz Ton

11 ms). Generell kann behauptet werden, je kürzer ein Ton wird, je weniger ausgeprägt wird die Tonhöhe wahrgenommen.

3.1.3 Eindeutigkeit der Tonhöhe

Ein weiterer wesentlicher Aspekt der Tonhöhenwahrnehmung bezieht sich die Ausgeprägtheit oder Eindeutigkeit der Tonhöhe eines Reizes (pitch strength). Sind die o.g. Bedingungen erfüllt und eine Tonhöhe ist wahrnehmbar, so kann diese in ihrer Qualität stärker oder schwächer ausgeprägt sein. Untersuchungen hierzu wurden u. a. von Fastl und Stoll (1979) durchgeführt, welche zu dem Ergebnis kamen, dass die Eindeutigkeit der Tonhöhe eines Reizes unabhängig ist von dessen eigentlicher Tonhöhe, d. h. wie hoch oder tief ein Reiz wahrgenommen wird. Ausgenommen hiervon sind sehr tiefe und sehr hohe Töne. Sie konnten weiter zeigen, dass die Tonhöhe bei Reizen mit einem Linienspektrum (zusammengesetzt aus einzelnen abgrenzbaren Frequenzen) deutlicher ausgeprägt ist als bei Reizen mit einem kontinuierlichen Spektrum (nicht einzelne Frequenzen, sondern Frequenzbereiche sind enthalten). So ist leicht nachzuvollziehen, dass ein harmonischer Klang z.B. einer Gitarrensaite, welcher ein Linienspektrum aufweist, eine ausgeprägtere Tonhöhe aufweist als z.B. das kontinuierliche Frequenzspektrum eines rauschenden Fernsehers. Der höchste Grad an Tonhöhenausprägung wird nach Aussage der Autoren bei Darbietung eines einzelnen Sinustones wahrgenommen, der geringste bei breitbandigem Rauschen. Neuere Arbeiten weisen darauf hin, dass die wahrgenommene Eindeutigkeit der Tonhöhe eines Reizes stark von zeitlichen Parametern wie z.B. der Wellenform beeinflusst wird (Shofner & Selas, 2002; Patterson, Handel, Yost & Datta, 1997).

Zusammenfassend kann konstatiert werden, dass die Ausgeprägtheit der Tonhöhe eine weitere unabhängige Empfindungsgröße darstellt, bei der kein direkter Zusammenhang mit einem physikalischen Parameter postuliert werden kann, wie es—wenn auch stark vereinfacht—z.B. bei Tonhöhe und Frequenz oder Lautheit und Amplitude der Fall ist.

3.1.4 Messung von Tonhöhe

Wie lässt sich Tonhöhe messen? Wie jede Empfindung stellt auch Tonhöhe eine nicht direkt beobachtbare Größe dar. Um einer Empfindung einen Messwert zuzuordnen, bietet die Psychophysik ein breites Repertoire an Untersuchungsmethoden an. Die Art der Frage stellt dabei unterschiedliche Anforderungen

an den Probanden. Nach Hall (1991) lassen sich generell drei Arten von Tonhöhenurteilen unterscheiden:

- Bei einem *absoluten Tonhöhenurteil* ist ein Proband angehalten, die Tonhöhe eines isolierten Reizes zu benennen, was i.d.R. durch die Zuordnung der entsprechenden Note oder einer Frequenz geschieht. Die Fähigkeit des absoluten Gehörs besitzen nur wenige Probanden.
- Bei *Intervallurteile* wird der Abstand zwischen zwei gegebenen Tönen erfragt. Die Antwort enthält die Distanz (in z.B. Terz, Quinte, Quarte, ...) und die Richtung der Tonhöhenänderung (auf- oder absteigend). Probanden, die reliabel Intervallurteile abgeben können, sind aller Regel nach trainierte Musiker.
- Die einfachste Form des Tonhöhenurteils ist das *Vergleichsurteil*, bei dem nur die Richtung der Tonhöhenänderung erfragt wird. In Abhängigkeit von der Distanz und der Art der dargebotenen Töne, kann diese Aufgabe von den meisten Probanden erfüllt werden.

Aus nachvollziehbaren Gründen liefern die einfachsten Aufgaben mit den geringsten Anforderungen an den Probanden aller Regel nach die reliabelsten Messergebnisse. So wurden erste Untersuchungen zur Messung der Tonhöhe mit einzelnen Sinustönen als Referenzgröße durchgeführt. Diese bieten, neben dem Vorteil der hohen Eindeutigkeit der Tonhöhe auch eine klar begrenzte Anzahl an variierbaren Parametern.

Die Mel-Skala

Die von Stevens, Volkman und Newman (1937) entwickelte *Mel-Skala* beschreibt den Zusammenhang zwischen der Frequenz eines Sinustones und der wahrgenommenen Tonhöhe. Bei der Erstellung dieser Skala verwendeten sie die Methode der direkten Größenschätzung. Bei dieser Methode werden den dargebotenen Sinustönen Werte zugeordnet, die die Ausprägung der Tonhöhe im Vergleich zur Tonhöhe des Referenzsinustons beschreiben, so entspricht 1000 Hz einem Skalenwert von 1000 Mel. Stevens konnte zeigen, dass sich der Zusammenhang zwischen Frequenz und Tonhöhe in einem Bereich bis 500 Hz gut durch eine lineare Funktion beschreiben lässt. Frequenz und Tonhöhe sind hier äquivalent, so dass eine Verdoppelung der Frequenz zu einer Verdoppelung der wahrgenommenen Tonhöhe führt. Für Töne ab etwa 500 Hz stehen Frequenz und Tonhöhe in einem nichtlinearen Zusammenhang. So muss z.B.

die Frequenz eines Tones mit 1500 Hz auf 10000 Hz erhöht werden, um zu einer wahrgenommenen Verdoppelung der Tonhöhe zu führen.

Da diese Skala ausschließlich für die Tonhöhe bei Darbietung reiner Sinustönen konzipiert wurde, ist auch ihre Gültigkeit nur für diese Reize aufrecht zu erhalten. Die Tonhöhe komplexer Schalle, welche im Allgemeinen mehrdeutig ist, lässt sich mit dieser eindimensionalen Skala nicht vorhersagen. Der Zusammenhang zwischen Reiz und Tonhöhe scheint weitaus komplizierter zu sein, als dass er durch eine einzige Funktion beschrieben werden könnte (Terhardt, 1998).

Die Bark-Skala und die ERB-Skala

Eine weitere Skala, welche in gleicher Weise die Beziehung von Tonhöhe und Frequenz beschreibt, ist die nach Heinrich Barkhausen benannte *Bark-Skala*. Mel- und Bark-Skala lassen sich ineinander überführen. So entspricht 1 Bark = 100 Mel. Weiterhin beschreibt die Bark-Skala den Ort der Erregung der Nervenzellen im Innenohr. Ein Ton setzt im Innenohr die Basilarmembran in Schwingung, welche wiederum die an ihr anliegenden Nervenzellen (Haarzellen) erregt. Die Tonhöhe eines Tones hängt davon ab, an welcher Stelle die Basilarmembran ihr Schwingungsmaximum erreicht. Ein Bark entspricht etwa 1,3 mm der Basilarmembran, was der Breite eines auditiven Filters entspricht.

Die Auffassung, dass die Basilarmembran einer Folge von einzelnen Filtern entspricht, findet sich auch in der von Moore und Glasberg (1983) vorgeschlagenen ERB-Skalierung. Die Größe eines Filters δ_{ERB} entspricht dabei der eines Rechteckfilters (Equivalent Rectangular Bandwidth) und ist eine Funktion der Frequenz f .

$$\delta_{ERB} = 6.23f^2 + 93.39f + 28.52 \quad (3.1)$$

Die Bandbreite der Filterung nimmt mit steigender Frequenz zu, d.h. je höher die Frequenz, desto eher überschneiden sich benachbarte Filter. Hier lässt sich auch das Konzept der „Auflösung“ verdeutlichen, je eindeutiger eine Frequenz einem Filter zuzuordnen ist, je weniger Interaktionen mit benachbarten Filter bestehen, je genauer kann diese Frequenz in nachgeschalteten Prozessen aus dem Gesamtsignal extrahiert bzw. aufgelöst werden.

Sowohl die Bark- als auch die ERB-Skala basieren auf dem Konzept der kritischen Bandbreite. In Maskierungsexperimenten konnten Fletcher und Munson (1937) zeigen, dass die Entdeckungsleistung für ein Sinuston nur begrenzt abhängig ist von der Bandbreite eines Maskierungsrauschens. Steigt die Bandbreite des Rauschens, so muss die Intensität des Sinustones gesteigert werden,

um diesen noch zu detektieren. Ab einer bestimmten kritischen Bandbreite bewirkt eine weitere Erweiterung des Rauschens keine zusätzliche Maskierung mehr, da hier die Filtergrenze erreicht ist und bereits benachbarte Filter angesprochen werden. Dieses Experiment belegt die bereits von Helmholtz (1863) vorgetragene Annahme, das Gehör bestehe aus einer Reihe einzelner Filter.

3.1.5 Einflussgrößen

Generell kann postuliert werden, dass die Tonhöhe im Allgemeinen von der Frequenz bzw. der Periode eines Tones abhängt. Jedoch führen auch andere Parameter eines Tones zur Veränderung der Tonhöhe. Dieser als *Tonhöhenabweichung* bezeichnete Effekt tritt auf, wenn die Tonhöhe eines Tones mit fester Frequenz sich ändert, sobald andere Parameter des Tones geändert werden. Auch das Hinzufügen weiterer Schallreize kann zur Tonhöhenabweichung führen.

Der Einfluss des Schallpegels auf die Tonhöhe. Untersuchungen zur Tonhöhe von einfachen Sinustönen bei Variation des Schallpegels konnten zeigen, dass der Schallpegel eines Reizes die wahrgenommene Tonhöhe in Abhängigkeit von dessen Frequenz beeinflusst (Terhardt, 1974; Stevens, 1935). Tiefe Töne mit einer Grundfrequenz von z.B. 200 Hz, werden durch Anheben des Schallpegels als noch tiefer empfunden. Hingegen kommt es bei Tönen mit Frequenzen im oberen Bereich von z.B. 8000 Hz zu einer Anhebung der Tonhöhe, wenn der Schallpegel des Reizes steigt, der Ton wird dann als höher empfunden. Dieses Phänomen tritt um so stärker auf, je extremer die Frequenzen gewählt werden.

Der Einfluss der Tondauer auf die Tonhöhe. Auch die Dauer eines Tones kann sich auf die Tonhöhe auswirken, wenngleich dieser Einfluss relativ gering ist und erst bei einer Tondauer von weniger als 200 ms auftritt (Terhardt, 1998). Hierbei kann jedoch keine eindeutige Richtung (höher oder tiefer) postuliert werden, sondern eine generelle Erhöhung der Varianz der wahrgenommenen Tonhöhe (Doughty & Garner, 1948). In Abgleichsexperimenten von Liang und Chistovich (1961) präsentierten diese Töne mit einer Dauer von 3 ms bis 600 ms und konnten zeigen, dass die Schwankung im Abgleichsurteil auch von der Frequenz des Testtones abhängt. Bei Tönen mit einer Dauer kleiner als ca. 150 ms erhöht sich demnach mit steigender Frequenz auch die Varianz im Tonhöhenurteil.

Der Einfluss von Zusatzschall auf die Tonhöhe. Die Tonhöhe eines Reizes kann durch die Darbietung eines zusätzlichen Reizes beeinflusst werden. Dies tritt in der Regel jedoch nur auf, wenn der erste Reiz durch den zweiten teilmaskiert wird (vgl. 3.1.4).

Dies gilt auch für kontralateral dargebotene Reize, mit dem Unterschied, dass hierbei keine Teilmaskierung zum Tragen kommt, was auch erklärt, dass dieser Effekt durchweg gering ausfällt.

Interaurale Tonhöhendifferenz. Die interaurale Tonhöhendifferenz bezieht sich auf die physiologischen Unterschiede zwischen den beiden Ohren (interaural). Bei Darbietung eines z.B. Sinustones auf dem einen Ohr wird eine andere Tonhöhe empfunden als bei Darbietung des gleichen Reizes auf dem anderen Ohr. Dieser Effekt lässt sich z.B. durch die abwechselnde Präsentation eines Sinustones mal auf dem einen und mal auf dem anderen Ohr demonstrieren (Terhardt, 1998). Auch wenn die Unterschiede nur von geringem Ausmaß sind, ist die interaurale Tonhöhendifferenz ein Beleg dafür, dass die (spektrale) Tonhöhe in beiden Ohren autonom gebildet wird.

Die erstaunliche Tatsache, dass wir trotz dieser messbaren Unterschiede bei binauraler Darbietung eine einheitliche Tonhöhe wahrnehmen, lässt sich damit erklären, dass beide wenig verschiedenen Tonhöhen zentral zu einer einzigen mittleren Tonhöhe zusammengefasst werden (binaurale Tonhöhenfusion).

Auf die bei der Bildung von Tonhöhe beteiligten Mechanismen wird in Abschnitt 6.6 näher eingegangen.

Tonhöhenurteil und Referenzierung: Das absolute Gehör

Die Fähigkeit, einen Ton unabhängig von vorgegebenen Referenztönen zu erkennen, wird als das absolute Gehör (perfect pitch) bezeichnet (vgl. 3.1.4). Da dieses Thema im Rahmen der weiter hinten vorgestellten Theorie zur Wahrnehmung von oktav-komplexen Tönen relevant sein wird, sei im Folgende näher darauf eingegangen.

Schätzungen zufolge verfügt etwa 1% der Bevölkerung über diese, auch als Tonhöhengedächtnis bezeichnete Fähigkeit (Moore, 2003). Das absolute Gehör scheint primär angeboren zu sein. Versuche, diese Fähigkeit zu trainieren, wie z.B. das Wiedererkennen eines 1000 Hz Tones, wiesen zwar Erfolge auf, die trainierten Probanden ohne absolutes Gehör erreichten jedoch im Vergleich zu genuinen Hörern nicht deren Leistung (Cuddy, 1968).

Neuere Untersuchungen deuten jedoch darauf hin, dass interne Referenzen

bestehen müssen, welche bei weniger künstlichen Aufgabenstellungen exakte Tonhöhenurteile erlauben. Evidenz für ein vorbewusstes Tonhöhengedächtnis liefert Levitin (1994). Von 46 Studenten ohne einschlägige musikalische Vorerfahrung, waren 26 in der Lage, ein selbst gewähltes Lied mit weniger als einem Halbton Abweichung zum Original zu singen. Physiologische Gegebenheiten wie die Tonlage der eigenen Ruhestimme oder der höchste bzw. tiefste Ton, welchen eine Person in der Lage ist, zu erzeugen, können hier ebenfalls Referenzfunktionen übernehmen. So sind Versuchspersonen in der Lage kleinste Tonhöhenverschiebungen in der elektronisch aufgezeichneten eigenen Stimme wahrzunehmen (Jones & Munhall, 2000). Weitere Einflüsse werden alltäglichen Hörgewohnheiten, wie z.B. bei Rezeption von populärer Musik aus Funk und Fernsehen, zugeschrieben. Bei der Untersuchung der Tonhöhe in 2400 aufgezeichneten Sätzen 15 holländischer Sprecher fand Braun (2001) einen Frequenzbias in Richtung der geläufigen musikalischen Skalierung (definiert durch Kammerton A=440 Hz). Weiterhin war der Bias beschränkt auf die dominanten Ganztöne A,B,C,D,E und F. Diese Befunde deuten darauf hin, dass ein absolutes Gehör eine normale, wenn auch vorbewusste, Fähigkeit ist, und Ergebnis eines Lernprozesses ist.

4 Theorie der Tonhöhenwahrnehmung

Im Folgenden Abschnitt soll auf die Theorien zur Tonhöhenwahrnehmung näher eingegangen werden. Beginnend bei historischen Ansätzen des neunzehnten und frühen zwanzigsten Jahrhunderts von z.B. Ohm, Seebeck oder von Helmholtz (zusammengefasst in Turner, 1977) über die Beiträge zur Gestaltpsychologie von z.B. Stumpf (1883) bis hin zu frühen Arbeiten von Fletcher oder Schouten (1940) unterliegt dieses Themenfeld ständigem Wachstum. Ein kleiner Ausschnitt der Vielzahl an Theorien und Modellen zur Tonhöhenwahrnehmung ist, ohne Anspruch auf Vollständigkeit, in Tabelle 4.1 dargestellt.

4.1 Überblick

Im Cepstrum-Verfahren von Noll (1967) wird die tonhöhenäquivalente Frequenz eines komplexen Tones durch Fourier-Transformation des logarithmisch dargestellten Betragsspektrums gewonnen. Das Frequenz-Histogramm-Verfahren von Schroeder (1968) berechnet die tonhöhenäquivalente Frequenz als größten gemeinsamen Teiler der Teiltonfrequenzen. Das HIPEX-System von Miller (1970) beruht auf dem Histogramm-Verfahren von Schroeder (1968). Die Theorie der virtuellen Tonhöhe von Terhardt (1972b) leitet die virtuelle Tonhöhe aus der extrahierten spektralen Tonhöhe ab. Eine harmonische Schablone kommt in der subharmonischen Koinzidenzdetektion zum tragen. Auf diese einschlägige Ortstheorie wird später genauer eingegangen.

Die Optimum Prozessor Theorie Goldstein (1973) ist eine Ableitung einer Tonhöhentheorie aus der statistischen Schätztheorie. Die Frequenz in einem Signal wird kanalweise in normalverteilte Variablen transformiert. Das Anwenden einer harmonischen Schablone geschieht hier mittels Maximum Likelihood Schätzung. Die Pattern Transformation Theorie von Wightman (1973) stellt eine ans Gehör angepasste Version des Cepstrum-Verfahrens dar.

Nr.	Theorie	Referenz
1	Cepstrum-Verfahren	Noll (1967)
2	Frequenz-Histogramm-Verfahren	Schroeder (1968)
3	HIPEX-System	Miller (1970)
4	Virtual-Pitch-Theorie	Terhardt (1972b)
5	Optimum Prozessor Theorie	Goldstein (1973)
6	Pattern Transformation Theorie	Wightman (1973)
7	Harmonic Pitch Detector	Seneff (1978)
8	Verfahren von Piszgall	Piszcalski und Galler (1979)
9	Pitch for Ripple Noise	Yost und Hill (1979)
10	Harmonic Sieve	Duifhuis, Willems und Sluyter (1982)
11	DWS pitch meter	Scheffers (1983)
12	Central-Spektrum-Theorie	Raatgever und Bilsen (1986)
13	Subharmonic Summation	Hermes (1988)
14	SPINET Model	Cohen, Grossberg und Wyse (1995)
15	Unitary Pitch Model	Meddis und Hewitt (1991b) Meddis und O'Mard (1997)
16	Auditory Image Model	Patterson, Allerhand und Giguere (1995)
17	Neural Cancellation Model	de Cheveigné (1998a)
18	Multi-Pitch and Periodicity Model	Karjalainen und Tolonen (1999)
19	Multiple Fundamental Frequency Estimation	Klapuri (2006)

Tabelle 4.1: Eine Auswahl vom Modellen zur Tonhöhenwahrnehmung seit 1960.

Im Harmonic Pitch Detector von Seneff (1978) wird die Grundfrequenz eines Teiltonspektrums aus den Frequenzabständen der Teiltöne eines bestimmten Frequenzbereichs abgeleitet. Das Verfahren von Piszcalski und Galler (1979) wird ebenfalls die tonhöhenäquivalente Frequenz aus den Verhältnissen der Teiltöne erschlossen. In ihrem Modell beschreiben Yost und Hill (1979) die Vorhersage der Tonhöhe für u.a. sich wiederholendes Rauschen (rippled noise). Das Harmonic Sieve von Duifhuis et al. (1982) stellte eine Implementation der Goldstein'schen Theorie für technische Anwendungen dar. Es wurde u.a. entwickelt, um die Tonhöhe von Sprachsignalen zu ermitteln. Mit dem DWS pitch meter knüpft Scheffers (1983) an die Arbeiten von Duifhuis et al. (1982) an. Das Modell ist in der Lage, teilmaskierte komplexe Töne zu identifizieren. Die Central-Spektrum-Theorie von Raatgever und Bilsen (1986) ist neben der Tonhöhentheorie auch noch eine Theorie zum binauralen Hören

und ermöglicht die Vorhersage dichotischer Tonhöhe.

Die Theorie der Subharmonic Summation von Hermes (1988) kann als Abwandlung und Erweiterung der Theorie der virtuellen Tonhöhe angesehen werden. Das SPINET Model von Cohen et al. (1995) ist eine Kombination aus der Theorie der virtuellen Tonhöhe und der Optimum Prozessor Theorie. Die Architektur des Modells beruht auf einem neuronalen Netzwerk. Das Unitary Pitch Modell von Meddis und Hewitt (1991b) bzw. Meddis und O'Mard (1997) ist ein zeitliches Modell, das keine explizite Extraktion der Spektraltonhöhe durchführt, sondern eine summarische Autokorrelation von quasi-neuronalen Signalen. Das Auditory Image Model von Patterson et al. (1995) setzt auf das Unitary Pitch Modell auf und bildet in der letzten Stufe statt der Autokorrelation ein „auditives Bild“ des gegebenen Signals. Das Neural Cancellation Model de Cheveigné (1998a) ist ein auf der Zeit Domain operierendes Modell. Das Multi-Pitch and Periodicity Model von Karjalainen und Tolonen (1999) ist eine Weiterentwicklung des Unitary Pitch Modells. Es ist, so die Autoren, effizienter und praktischer als das Ausgangsmodell. Das Modell kann neben der Vorhersage von Tonhöhe auch mehrere Schallquellen unterscheiden und isolieren. Klapuri (2006) stellt ein iteratives Verfahren vor, welches aus einer Wellenform eines Signals die Hüllkurve der ersten Harmonischen schätzt und diese Hülle vom Ursprungssignal subtrahiert. Diese Prozedur wird sodann wiederholt, um die übrigen harmonischen Residuen zu schätzen.

Die hier präsentierte Auswahl einschlägiger psychoakustischer Modelle zur Tonhöhenwahrnehmung kann problemlos erweitert werden und sollte dem Anspruch des Überblicks auf diesem Gebiet genügen. Neben der aufgeführten Originalliteratur kann für einen vertiefenden Blick z.B. die Arbeit von Wightman und Green (1974) für eine historische Zusammenfassung oder de Cheveigné (2004) für eine Darstellung der methodischen Grundlagen empfohlen werden.

4.1.1 Zur Vorhersage von Tonhöhe: Ort und Zeit

Allein die Vielzahl der Theorien macht deutlich, dass die Vorhersage der Tonhöhe kein triviales Unterfangen zu sein scheint, in dem unterschiedlichste Wissenschaften involviert sind. So wurden die oben skizzierten Theorien von z.B. Physikern, Ingenieuren, Physiologen, Neurowissenschaftlern oder Psychologen entworfen und weiterentwickelt, ohne zu einer allgemeingültigen Theorie zu gelangen.

Versucht man die Liste der Theorien zu ordnen, so lassen sich zwei konkurrierende Ansätze unterscheiden, *Ortstheorien* und *Zeittheorien*. Beide haben

den Anspruch, für möglichst viele Reize die Tonhöhe vorherzusagen und das möglichst genau. Im Kern unterscheiden sich beide Theorierichtungen bereits in der Betrachtung eines Schallsignals. Ortstheorien benötigen die einzelnen Frequenzen eines Reizes zur Vorhersage der Tonhöhe. Im Fokus steht daher das Frequenzspektrum. Da im Innenohr, genauer entlang der Basilarmembran, eine Zerlegung des eingehenden Signals in Frequenzkomponenten geschieht und hierbei der Ort auf der Membran mit der wahrgenommenen Tonhöhe korrespondiert, werden diese Theorien Ortstheorien genannt.

Die Zeittheorien arbeiten mit dem zeitlichen Verlauf des eingehenden Signals, der sog. Wellenform oder Feinstruktur, um die Tonhöhe vorherzusagen.

Beide Ansätze konkurrieren um die Erklärung einzelner auditiver Phänomene, wobei die Ortstheorien nicht alle durch die Zeittheorien erklärten Phänomene vorhersagen kann und vice versa. Neuere Theorien, sog. hybrid Modelle, integrieren Mechanismen aus beiden Richtungen.

4.2 Ortstheorien

Wie bereits dargestellt wurde, korrespondiert die Tonhöhe eines Reizes mit dessen Periode. Ein Mechanismus zur Vorhersage der Tonhöhe sollte folglich in der Lage sein, die Periode eines Reizes oder dessen Inverse, die fundamentale Frequenz F_0 , zu schätzen. Dass dies allein zwar notwendig jedoch nicht hinreichend ist, soll für beide Theorierichtungen anhand von Beispielen illustriert werden.

4.2.1 Auf dem Weg zum Algorithmus

Gegeben sei ein Frequenzspektrum eines beliebigen Schallsignals, gesucht ist eine Schätzung der wahrgenommenen Tonhöhe dieses Signals (siehe Abb. 4.1). Dass die Tonhöhe mit der Frequenz assoziiert ist, wurde bereits in den eingehend aufgeführten Definitionen beschrieben und führt leicht zur Schätzung der Tonhöhe in Abbildung A. Bei mehr als einer Frequenz könnte man zu der Annahme gelangen, dass der Teilton mit der größten Amplitude—vereinfacht gesagt der lauteste—die wahrgenommene Tonhöhe bestimmt. Mag diese Annahme für Sinustöne und bestimmte komplexe Töne zutreffend sein, wird durch eine Blick auf Abbildung B leicht klar, dass sie nicht haltbar ist. Trotzdem der Amplitude eine wesentliche Rolle bei der Bestimmung der Tonhöhe zukommt, entspricht die wahrgenommene Tonhöhe nicht der in Abbildung B dargestellten zweiten Harmonischen, sondern korrespondiert mit der kleinsten Frequenz

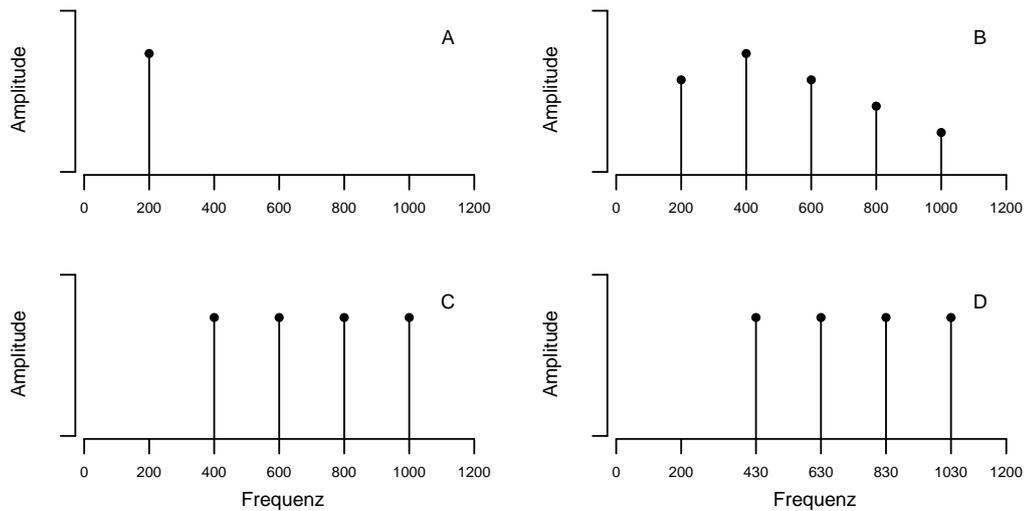


Abbildung 4.1. Das Frequenzspektrum eines Sinustons (A), eines harmonischen Tons (B), eines harmonischen Tons mit fehlendem Grundton (missing fundamental) und eines harmonisch-verschobenen Tons (D).

bei 200 Hz.

Wenn die Tonhöhe mit der Periode eines Signals assoziiert ist und das Inverse der Periode die fundamentale Frequenz F_0 ist, könnte eine einfache Regel zur Schätzung der Tonhöhe wie folgt lauten: Die Tonhöhe entspricht der kleinsten im Spektrum vorkommenden Frequenz. Im Falle eines einfachen Sinustones (A) stimmt diese Behauptung mit empirischen Befunden überein. Auch für harmonisch-komplexe Töne (B), deren Teiltonfrequenzen der Regel $F_i = i \cdot F_0$ folgen, trifft diese Behauptung zu. Da die meisten Spektren alltäglicher Klänge wie Sprache oder Musik den Regeln harmonischer Töne folgen wäre, würde man sich nur auf diese Reize beziehen, das Problem somit gelöst.

Was jedoch, wenn die fundamentale Frequenz F_0 entfernt wird? Durch das von Schouten (1940) durchgeführte Experiment konnte dieser nachweisen, dass, wenn die erste Komponente eines harmonischen Tones eliminiert wird, sich zwar die Klangfarbe ändert, die Tonhöhe jedoch nicht (C). Dieser als Phänomen des fehlenden Grundtons (missing fundamental) bezeichnete Effekt macht deutlich, dass Tonhöhe nicht zwingend auf das Vorhandensein von spektralen Komponenten zurückzuführen ist und somit keine einfache Funktion des Spektrums ist. Die vorgeschlagene Regel ist somit ebenfalls nicht haltbar. Dieses Phänomen lässt sich auch noch stabil nachweisen, wenn mehrere Komponenten entfernt werden.

Eine alternative Regel, um auch das Phänomen des fehlenden Grundtons

zu modellieren, könnte so formuliert werden: Der Abstand zwischen benachbarten Teiltönen entspricht der Tonhöhe. Da die Obertöne eines harmonischen Tones stets den gleichen Abstand aufweisen, würde die Differenz der Frequenzen zweier benachbarter Teiltöne $F_{n+1} - F_n$ zu einer Schätzung der Tonhöhe führen. Das Problem des fehlenden Grundtons wäre modellierbar.

Damit aber nicht genug. Das in Abbildung 4.1-D dargestellte Frequenzspektrum ist im Unterschied zu dem in C dargestellten Spektrum um 30 Hz verschoben (shifted harmonic). Der Abstandsregel folgend, sollte dieser Ton eine Tonhöhe von 200 Hz hervorrufen. Es konnte jedoch gezeigt werden, dass, in Abhängigkeit von der Weite der Verschiebung, diese Reize eine signifikant höhere Tonhöhe aufweisen (Schouten, 1940). Weiterhin führt die Verschiebung der Teiltöne neben einem unnatürlichen Klang auch zu Mehrdeutigkeiten in der Tonhöhenwahrnehmung (vgl. Abschnitt 3.1.3).

Die dargestellte Prozedur, immer artifizellere Reize zu konstruieren, um bestehende Theorien zu testen, macht deutlich, welchen hohen Ansprüchen aktuelle Theorien gerecht werden müssen. Um diesen Vorgang abzukürzen, sei im Folgenden ein einschlägiger Algorithmus auf Basis der Terhardschen Ortstheorie vorgestellt.

Subharmonische Koinzidenzdetektion

Der Algorithmus, welcher allen angeführten Schwierigkeiten gerecht wird und deshalb in einschlägigen Ortstheorien (z.B. Terhardt, 1979; Cohen et al., 1995) zum Einsatz kommt, basiert auf einem *Histogramm subharmonischer Tonhöhen* und kann wie folgt skizziert werden. Dividiert man jede in einem Spektrum vorhandene Frequenz durch kleine, positive, ganzzahlige Teiler (z.B. 1,2,3,...,10), so erhält man die sog. Subharmonischen der jeweiligen Frequenzen. Trägt man diese in einem Histogramm ab, so entspricht die Tonhöhe der am häufigsten vorkommenden Subharmonischen, d.h. dem höchsten Balken des Histogramms. Alle Subharmonischen entsprechen potentiell wahrnehmbaren Tonhöhen. Kommen mehrere Subharmonische gleich häufig vor, so entspricht die primäre Tonhöhe der Subharmonischen mit dem größeren Klassenmittelwert (de Cheveigné, 2004).

Die in Abb. 4.2 dargestellten Histogramme zeigen das Ergebnis der Tonhöhenschätzung für die Spektren C und D aus Abbildung 4.1.

Die Höhe der Balken des Histogramms entsprechen den Häufigkeiten der subharmonischen Tonhöhen je Intervall. Bei einer Intervallbreite von 4 Hz und einem maximalen Teiler von 10 resultieren Tonhöhenschätzungen entsprechend

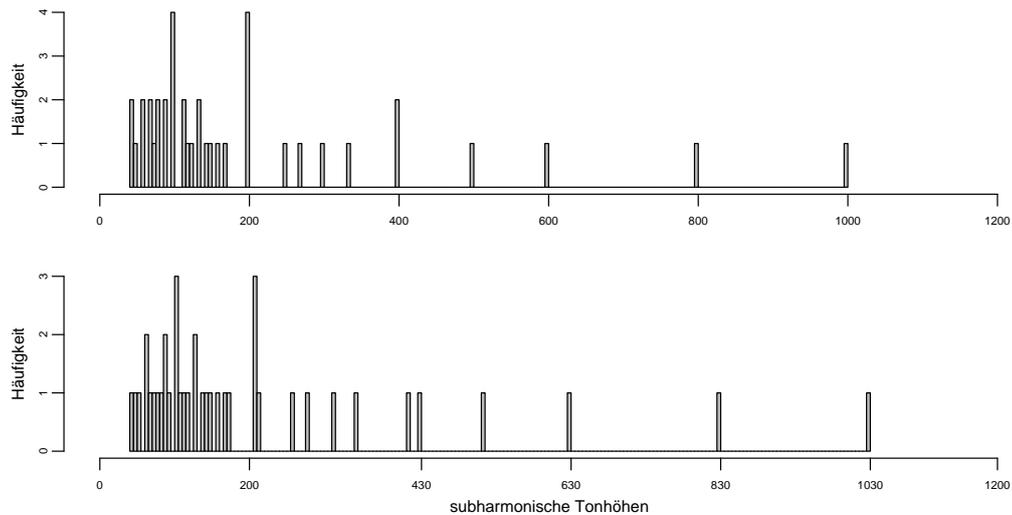


Abbildung 4.2. Histogramm subharmonischer Tonhöhen, A: für das Reizspektrum des harmonischen Tones mit fehlendem Grundton aus Abb. 4.1 C, B: für das Reizspektrums des verschobenen harmonischen Ton aus Abb. 4.1 D.

den Frequenzen von 200 pu (C, missing fundamental) und 206 pu (D, shifted harmonic). Die u.a. von Terhardt (1971) gewählte Einheit pu steht für pitch units und verdeutlicht, dass es sich hierbei um Schätzungen der Tonhöhe handelt. Der hier vorgestellte Algorithmus geht zurück auf Arbeiten von Walliser (1969) und Schroeder (1968) und wurde von Terhardt (1972a, b) weiterentwickelt. Er stellt einen wesentlichen Bestandteil des „Modells der virtuellen Tonhöhe“ dar. Die verschiedenen verwendeten Parameter, wie z.B. die Anzahl der ganzzahligen Teiler oder die Breite der Histogrammintervalle sind prinzipiell frei wählbar, orientieren sich aber an empirisch überprüften Modellanpassungen.

4.2.2 Das Modell der virtuellen Tonhöhe

Die von Terhardt (1972) entwickelte Theorie der virtuellen Tonhöhe unterscheidet zwei Arten von Tonhöhe, die spektrale und die virtuelle. In die Kategorie der spektralen Tonhöhe fallen alle die Tonhöhenempfindungen, bei denen die wahrgenommene Tonhöhe direkt auf eine im Spektrum des Signals vorhandene Frequenz zurückführbar ist. Das einfachste Beispiel hierfür ist die durch einen Sinuston hervorgerufene Tonhöhe. Diese ist allein auf die Frequenz des Sinustones zurückzuführen. Bei komplexeren Schallereignissen kann es mehrere Spektraltonhöhen geben, welche mit hinreichender Aufmerksamkeit meist bewusst

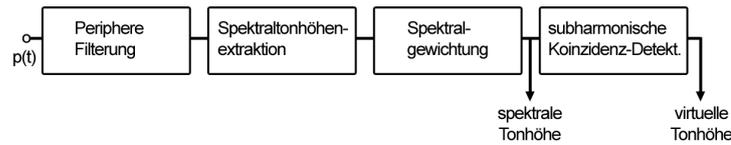


Abbildung 4.3. Schematische Darstellung der Theorie der virtuellen Tonhöhe. Ein Schallsignal $p(t)$ ruft im Allgemeinen mehrere Spektraltonhöhen und mehrere virtuelle Tonhöhen hervor.

wahrgenommen werden können (Terhardt, 1998). Sie bilden die Grundlage für die Entstehung weiterer Tonhöhen, den sogenannten virtuellen Tonhöhen. Ein Beispiel ist der als Phänomen des fehlenden Grundtons bekannte Effekt, der auftritt, wenn die fundamentale Frequenz eines harmonisch komplexen Klanges entfernt wird. Die Tonhöhe des resultierenden Klanges entspricht der Tonhöhe vor der Entfernung der Fundamentalen. Da diese Tonhöhe keine korrespondierende Frequenz hat, wird sie als virtuell bezeichnet.

Terhardt geht in seiner Theorie davon aus, dass die Signalverarbeitung bei der Wahrnehmung von Tonhöhe ein hierarchischer Prozess ist, d.h. erst werden die spektralen Tonhöhen ermittelt, um daraus die virtuellen Tonhöhen abzuleiten. Ausgehend von dieser Theorie stellt Terhardt (1979) ein computationales Modell vor, welches vier nacheinander ablaufende Prozesse beinhaltet (siehe Abb. 4.3). Die einzelnen Prozesse können wie folgt beschrieben werden.

Der erste Verarbeitungsschritt, die *Periphere Filterung* (Peripheral Ear Transduction, PET), besteht aus zwei Elementen: 1. dem Filter zur Nachbildung der Gehörgangsresonanz (Ear Canal Resonances, ECR) und 2. dem Filter zur Nachbildung der cochleären Übertragung (Cochlear Transmission Functions, CTF). Der erste Filter verändert das eintreffende Signal in der Weise, wie es auch im menschlichen Ohr der Fall ist. Dabei wird durch einen linearen Filter die im Gehörgang entstehende Resonanz modelliert. Die zweite Einheit zur Nachbildung der cochleären Übertragungsfunktion stellt ein System von parallel verschalteten Filtern dar, wobei die einzelnen Kanäle in vielerlei Hinsicht den Fasern des Hörnervs, welche über die inneren Haarzellen erregt werden, entsprechen.

In der zweiten Stufe der Signalverarbeitung erfolgt die *Extraktion der Spektraltonhöhen*, welche aus den Ausgangssignalen der Kanalfilter des PET-Systems gewonnen werden. Das Ergebnis dieser Stufe ähnelt einem Fourier-Spektrum, dem Resultat der Fourieranalyse. Die Funktionsweise dieser Stufe kann jedoch auf anderen Mechanismen basieren, beispielsweise einer Gleichrichtung der Ka-

nalsignale mit anschließender Tiefpassfilterung.

In der dritten Stufe erfolgt eine *Spektralgewichtung* der in der zweiten Stufe extrahierten Spektraltonhöhen. Diese Gewichte bestimmen in welchem Ausmaß die einzelnen Spektraltonhöhen bei der Bildung der virtuellen Tonhöhen einfließen. Sie ist abhängig vom Pegelüberschuss und von der Frequenzlage der Spektraltonhöhe.

Die so gewichteten Spektraltonhöhen bilden die Grundlage für die Berechnung der virtuellen Tonhöhe durch *subharmonische Koinzidenzdetektion*. Dabei werden aus jeder extrahierten Spektraltonhöhe 12 potentielle virtuelle Tonhöhen berechnet (Terhardt, 1998). Werden aus einem Signal nach Filterung, Extraktion und Gewichtung N Spektraltonhöhen abgeleitet, so werden in der vierten Stufe insgesamt $12N$ potentielle virtuelle Tonhöhen berechnet. Jede potentielle virtuelle Tonhöhe kann mehr oder minder genau mit einer oder mehreren anderen übereinstimmen. Dabei hat eine virtuelle Tonhöhe ein umso größeres Gewicht,

- je größer das Gewicht derjenigen Spektraltonhöhe ist, von der sie über die Schablone abstammt;
- je kleiner die subharmonische Ordnungszahl (der Teiler) ist;
- je größer die Anzahl der damit Übereinstimmenden, von anderen Teiltönen abstammende Tonhöhe ist;
- je genauer die Übereinstimmung ist.

Da die spektralen und virtuellen Tonhöhen in der Hörwahrnehmung konkurrieren und einzelne mit gezielter Aufmerksamkeit bewusst wahrgenommen werden können, stellt die Bildung einer Rangreihe eine *Zusammenfassung aller Tonhöhen* dar. Dabei stellt die Tonhöhe mit dem höchsten Gewicht *die* Tonhöhe dar, die spontan ohne subjektive Analyse wahrgenommen wird.

4.2.3 Zum Konzept der virtuellen Tonhöhe

In seiner Theorie unterscheidet Terhardt (1972b) zwischen der *spektralen* Tonhöhe und der *virtuellen* Tonhöhe. Bei der spektralen Tonhöhenwahrnehmung bezieht sich die empfundene Tonhöhe auf tatsächlich im Spektrum vorhandene Frequenzen, während die virtuellen Tonhöhen mit Frequenzen korrespondieren, welche gerade nicht im Signal vorhanden sind.

Zur Verdeutlichung dieser Unterscheidung bzw. als Existenzbeweis der virtuellen Tonhöhe sei ein Experiment von Smoorenburg (1970) angeführt. Dieser

präsentierte seinen Probanden Paare komplexer Töne mit zwei Harmonischen, jedoch stets ohne die fundamentale Frequenz. Die Aufgabe der Hörer war, zu entscheiden, ob die Tonhöhe vom ersten zum zweiten Ton auf- oder absteigt. Die Töne waren so konzipiert, dass die Tonhöhe z.B. steigt, wenn nur die vorhandenen spektralen Komponenten beachtet werden, hingegen fällt, wenn der Hörer gemäß der fehlenden Grundfrequenz antwortet. Im damaligen Experiment antwortete die Hälfte der Hörer gemäß den Teiltonfrequenzen (spektrale Tonhöhenwahrnehmung) und die andere Hälfte gemäß der nicht vorhandenen Grundfrequenz (virtuelle Tonhöhenwahrnehmung). In Folgestudien konnte Hartmann (1989, zitiert in Houtsma & Fleuren, 1991) zeigen, dass 90% der Hörer die spektrale Wahrnehmung bevorzugen.

Aus dieser Debatte um die Präferenzen für die spektrale oder virtuelle Wahrnehmung von Reizen stammt auch die Unterscheidung von Hörertypen in analytische Hörer und synthetische Hörer. Ein analytischer Hörer ist demnach in der Lage, die spektralen Komponenten eines Reizes zu extrahieren und bewertet die Tonhöhe entsprechend den spektralen Wahrnehmung. Demgegenüber steht der synthetische Hörer, welcher nach holistischen Aspekten den Ton als Einheit wahrnimmt und die Tonhöhe gemäß der virtuellen Wahrnehmung bewertet.

Bei der Differenzierung individueller Präferenzen scheint es jedoch auch physikalische Faktoren zu geben. So fanden Rousseau et al. (1996) heraus, dass die bevorzugte Art der Wahrnehmung, virtuell oder spektral, von der Anzahl der Teiltöne abhängt. So werden Zwei-Ton-Reize eher spektral und Vier-Ton-Reize eher virtuell wahrgenommen. Auch scheint die Lateralität einen entscheidenden Einfluss zu haben. Bei monauraler Darbietung werden Reizen, welche auf dem rechten Ohr und somit der linken Hemisphäre dargeboten werden, schneller und im Sinne der virtuellen Wahrnehmung bewertet.

Zusammenfassend kann gesagt werden, dass, folgt man Terhardt, die virtuelle Tonhöhe durch einen Abgleich der spektralen Tonhöhe mit einem erlernten Muster, dem Muster der (Sub-)Harmonischen, gebildet wird. Sie wird in einem physiologisch nachgeschaltetem Prozess aus den spektralen Tonhöhen ermittelt. Dieser Mechanismus des Abgleichens von Mustern (patter-matching models) fußt auf der These, dass unser Wahrnehmungssystem unvollständige Muster, und so auch den fehlenden Grundton, rekonstruiert. Dies ist deshalb so entscheidend, da davon ausgegangen wird, dass die fundamentale Frequenz eines Signals *das* Korrelat zur Tonhöhe darstellt. Anders ausgedrückt, kennt man den Grundton eines Klanges, Schalls, Reizes, so ist der wesentlichste Schritt auf dem Weg zur Tonhöhenschätzung getan.

Die Annahmen, welche den „pattern-matching“ Modellen zugrunde liegen, stehen in enger Beziehung mit den Annahmen der Ortstheorie bzw. Methoden, welche das Spektrum eines Signals als Grundlage haben (de Cheveigné, 2004). Beide Begriffe werden oft synonym verwendet, wenngleich sie unterschiedliche Mechanismen des Schätzvorgangs abbilden. So werden z.B. bei den (Orts-)Modellen in Tabelle 4.1 unterschiedliche Methoden zur Schätzung der fundamentalen Frequenz verwendet (subharmonic summation, harmonic sieve, autocorrelation, cepstrum). Alle operieren auf dem gegebenen Frequenzspektrum (bzw. je nach Methode Power- oder Log-Spektrum), welches die harmonische Struktur eines Signals beinhaltet.

4.3 Zeittheorien

Zeittheorien gehen davon aus, dass der zeitliche Verlauf eines Signals, spezieller die Zeit zwischen bestimmten wiederkehrenden Mustern, z.B. Schwingungsmaxima oder Nulldurchgänge, in der Wellenform dieses Signals die Grundlage der auditiven Wahrnehmung darstellt. Wobei angenommen wird, dass sich allein aus dieser Information die verschiedenen auditiven Qualitäten, wie z.B. auch die Tonhöhe, bestimmen lassen.

Die entlang der Basilarmembran gefilterten Teiltöne lassen sich in aufgelöste (resolved) und nicht aufgelöste (unresolved) unterscheiden. Trifft nur eine Frequenz auf einen Filter, so spricht man von aufgelösten Teiltönen bzw. Harmonischen. Mit zunehmender Frequenz treffen mehr harmonische Teiltöne in den Bereich eines Filters. Das Ergebnis dieser Filterung ist eine komplexe Wellenform, zusammengesetzt aus den interagierenden Frequenzen dieses Bereiches. Diese Wellenform wiederholt sich, aufgrund der Regelmäßigkeit der Frequenz in harmonischen Klängen, mit der Frequenz des Grundtones (Hartmann, 1998).

Als einer der ersten Vertreter der Zeittheorien geht Schouten (1940) davon aus, dass allein die nicht aufgelösten Harmonischen hinreichend sind, um die Periode und somit die Fundamentale Frequenz ergo eine Schätzung der Tonhöhe, zu ermitteln.

Weitere physiologische Hinweise, dass zeitliche Informationen verarbeitet werden, liefern Beobachtungen über die Feuerrate der Zellen im auditorischen Nerv (Moore, 2003). Diese korrespondiert z.B. mit der Frequenz eines Tones im dargebotenen Signal. So ist die Wahrscheinlichkeit, dass eine bestimmte Zelle ein Aktionspotential auslöst, am höchsten, wenn im dargebotenen Signal ein Schwingungsmaximum erreicht wird. Die Zelle feuert „im Takt“. Da eine

Nervenzelle aufgrund der Refraktärzeit nur etwa alle 2 ms in der Lage ist zu feuern, ist die mögliche Übertragungsfrequenz einer Zelle auf ca. maximal 500 Hz beschränkt. Diese Limitierung wird durch den Zusammenschluss mehrerer Zellen aufgehoben. Bei dem sog. phase-locking übernimmt eine Gruppe von Nervenzellen die Übertragung einer bestimmten Frequenz, so dass nicht jede Zelle immer feuert, aber immer in der gleichen Phasenlage zu den übrigen Zellen der gleichen Gruppe.

Bevor näher auf konkrete Zeit-Modelle eingegangen wird, soll—wie auch bei den Ortstheorien—eine einfache Methode zur Schätzung der Tonhöhe anhand zeitlicher Charakteristiken an Beispielen vorgestellt werden.

4.3.1 Autokorrelation

Die Autokorrelation gilt als eine Standardmethode bei der Analyse von Zeitreihen und dient dem Auffinden von Selbstähnlichkeit (Hartmann, 1998). Bei der Autokorrelationsfunktion (AKF bzw. ACF) wird ein Signal mit einem Teil seiner selbst korreliert. Die ACF eines diskreten Signals x ist definiert durch die Funktion

$$r_t(\tau) = \sum_{j=t+1}^{t+w} x_j x_{j+\tau} \quad (4.1)$$

wobei $r_t(\tau)$ die ACF, τ den Versatz (Lag) zum Zeitpunkt t bei gegebener Fensterbreite w beschreibt. Das Lag τ gibt den Versatz des w -langen Teils des Signals an, der mit dem ebenfalls w -langen Anfangsteil des Signals korreliert wird. Bei einem Signal der Länge N erhält man auf diese Weise maximal $N - w$ Autokorrelationen und somit ein Maß für die Selbstähnlichkeit in Abhängigkeit vom Lag τ .

Eigenschaften der Autokorrelationsfunktion Die Autokorrelationsfunktion hat ein globales Maximum bei $\tau = 0$. Lassen sich weitere globale Maxima finden und handelt es sich um ein stationäres Signal (d.h. mit konstanten Statistiken wie z.B. Phase und Spektrum), so ist das Signal *periodisch* und es existiert ein Lag T_0 , die *Periode*. Alle weiteren globalen Maxima lassen sich durch das ganzzahlige Vielfache der Periode nT_0 beschreiben, wobei $r_t(nT_0) = r_t(0)$ (Boersma, 1993). Da die *fundamentale Frequenz* F_0 eines periodischen Signals dem Reziproken der Periode entspricht, gilt: $F_0 = 1/T_0$.

Eine Gegenüberstellung von Wellenform und Autokorrelationsfunktion ist in Abbildung 4.4 für zwei Stimuli dargestellt. Im Falle des Kammertons A (440 Hz) findet sich die erste maximale positive Korrelation (abgesehen von $r_t(0)$)

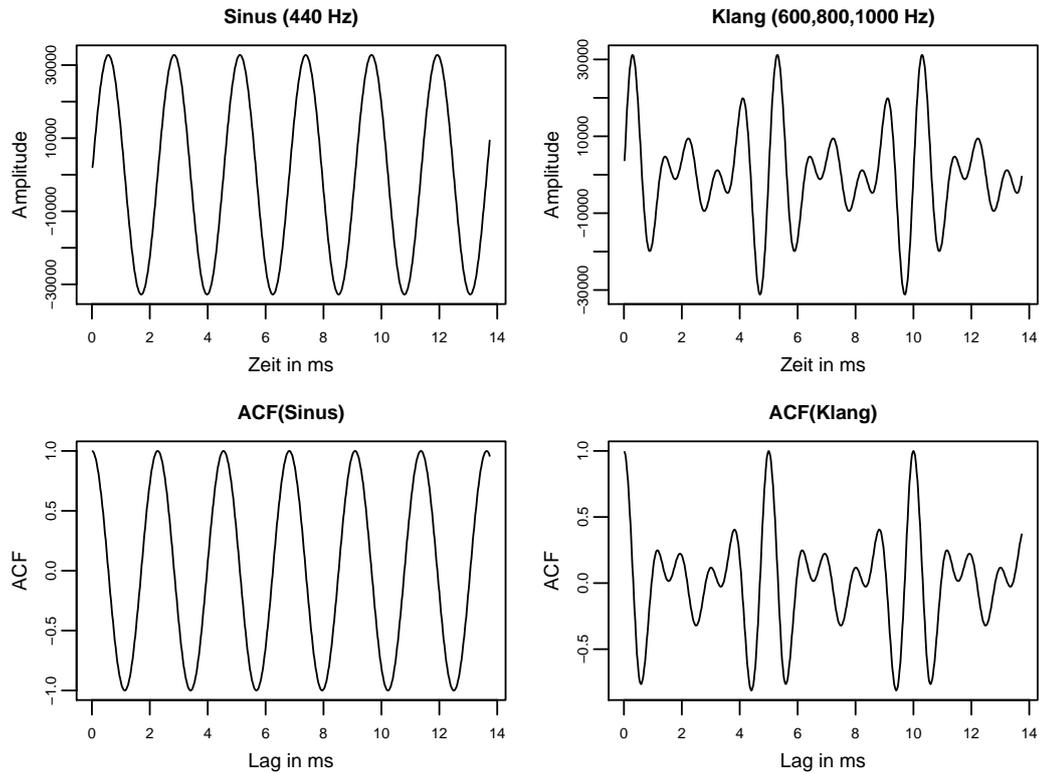


Abbildung 4.4. Die Wellenform und die Autokorrelationsfunktion für einen Sinuston mit einer Frequenz von 440 Hz (links) und einen komplexen Ton zusammengesetzt aus den Frequenzen 600 Hz, 800 Hz und 1000 Hz (rechts).

bei 2,27 ms, was wiederum einer Frequenz von $1/0.00227s = 440$ Hz entspricht. Im Falle des komplexen Signals (Abb. 4.4 rechts) liegt das erste relevante Maximum bei 5,0 ms, was einer Frequenz von $1/0.005s = 200$ Hz entspricht. Die Autokorrelation ist somit in der Lage, diese fundamentale Frequenz, welche nicht im Signal enthalten ist, zu ermitteln.

Die Autokorrelationsfunktion findet Verwendung in vielen einschlägigen Zeitmodellen¹. Als Inputsignal können unterschiedliche Reize verwendet werden. So kann das reine Schallsignal (z.B. cancellation model, de Cheveigné, 1998a) oder die zeitlichen Muster (spikes) im auditiven Nerv (z.B. Unitary Pitch Modell, Meddis & Hewitt, 1991b) als Eingangssignal Verwendung finden.

Wie an den dargestellten Beispielen demonstriert wurde, ist die Anwendung der Autokorrelation auf ein gegebenes Signal, um die Periodizität bzw. die Grundfrequenz und damit eine erste Schätzung der Tonhöhe zu erhalten,

¹Auch in Ortsmodellen findet der Algorithmus der Autokorrelation gelegentlich Verwendung. Allerdings wird er dort auf das Spektrum angewandt, z.B. zum Auffinden von harmonischen, wiederkehrenden Mustern (de Cheveigné, 2004).

auch ohne den sehr umfangreichen Überbau vieler computationalen Modelle hinreichend.

Eine Erweiterung der Autokorrelation wird von de Cheveigné und Kawahara (2002) vorgeschlagen und in dem von ihnen vorgestellten YIN-Algorithmus realisiert. Beginnend mit der ACF werden weitere Funktionen appliziert, welche sukzessiv zu einer besseren Vorhersage führen. Dies mit dem Vorteil der dreifach verringerten Fehlerrate bei der Bestimmung der Tonhöhe von Musik- und Sprachsignalen. Einen Überblick über die Verwendung zeitlicher Codes liefert Cariani (2001). Einen gestrafften Überblick über bestehende Zeitmodelle zur Extraktion von z.B. Lokalisation, Segregation, Tonhöhe und Timbre findet sich in de Cheveigné (1998b). Einwände gegen die Anwendung der Autokorrelation bei der Modellierung der Tonhöhe sind z.B. bei von Kaernbach und Demany (1998) formuliert worden.

Bevor die Eigenschaften von Orts- und Zeitmodellen gegenübergestellt werden, soll im Folgenden die Funktionsweise zeitlicher Modelle an zwei ausgewählten Beispielen verdeutlicht werden.

4.3.2 Das Unitary Pitch Modell

Bei dem von Meddis und Hewitt (1991b) und Meddis und O'Mard (1997) beschriebenen Tonhöhenmodell handelt es sich um ein zeitliches Modell, in welchem die Charakteristik der Wellenform als Grundlage zur Vorhersage von Tonhöhe dient. Das Modell ist kein ausschließlich von Meddis und Hewitt (1991b) entworfenes Modell, es vereinigt aktuelle in der Psychoakustik diskutierte Theorien und Modelle. Es orientiert sich in 8 Stufen stark an der auditiven Physiologie.

1. Außenohr (Periphere Filterung, Bandpassfilterung)
2. Mittelohr (Periphere Filterung, Bandpassfilterung)
3. Cochlea/Basilarmembran (Gammatone-Filterbank)
4. Haarzellenmodell (mechanisch-neuronale Transformation)
5. Auditiver Nerv (Inhibitions- und Refraktärzeitmodellierung)
6. Interspikeintervall (Autokorrelation)
7. Aggregation (Summe der Autokorrelationen über die Kanäle)
8. Extraktion von Tonhöhe und Klangfarbe

1 und 2: Außen- und Mittelohr

Im äußeren Ohr trifft ein Schallsignal nach dem Passieren des Gehörgangs auf das Trommelfell, wobei aufgrund der Gehörgangsresonanz die Frequenzen von etwa 2kHz bis 5kHz verstärkt werden (Shaw, 1974). Der im Modell verwendete Bandpassfilter trägt diesem Effekt Rechnung. Das mittlere Ohr überträgt die Energie des Schallsignals vom Trommelfell über die Knochen Hammer, Amboss und Steigbügel zum ovalen Fenster der Cochlea. Auch auf diesem Signalweg kommt es zur Verstärkung bestimmter Frequenzbereiche, was ebenfalls durch einen Bandpassfilter mit einem maximalen Durchlass bei 1kHz modelliert wird. Im Modell werden die Funktionen der beiden Bandpassfilter kombiniert und bilden die Stufe der peripheren Filterung.

3: Cochlea/Basilarmembran

Die Hörschnecke oder Cochlea ist ein mit Flüssigkeit gefülltes Organ und bildet zusammen mit dem Gleichgewichtsorgan das Innenohr. Die Basilarmembran ist eine von zwei Membranen, welche der Länge nach die Cochlear durchzieht. Sie reicht vom ovalen Fenster (Base) bis zur Spitze der Cochlear (Apex). Ein Schallsignal, welches als Schwingung des ovalen Fensters die Cochlea erreicht, verursacht Druckunterschiede entlang der Basilarmembran, was zur Bewegung der Basilarmembran führt. Die Form und Geschwindigkeit der Bewegung ist dabei abhängig vom eintreffenden Schallsignal. Im Falle eines einfachen Sinustons ähnelt diese Bewegung einer Welle, die entlang der Basilarmembran wandert. Der Ort der maximalen Auslenkung der Basilarmembran korrespondiert mit der Frequenz des Schallsignals. Tiefe Töne verursachen maximale Auslenkungen im Bereich der Apex, während hohe Töne zu Auslenkungen im Bereich des ovalen Fensters (Base) führen. Auf diese Weise ist es, analog zur Fourier-Analyse, möglich die Komponenten eines Schallsignals zu extrahieren.

Im Modell wird die Funktion der Basilarmembran durch eine Gammaton-Filterbank, bestehend aus 128 digitalen Bandpassfiltern im Bereich von 80Hz bis 8kHz, realisiert. Die Zentren der überlappenden Filter sind gleichabständig auf der logarithmischen Frequenzachse angeordnet. Der Abstand zwischen den einzelnen Filtern entspricht etwa 0.25 equivalent rechteckiger Bandbreite² (equivalent rectangular bandwidth, ERB, siehe Abschnitt 3.1.4), wobei ein ERB etwa einer Länge von 0.89 mm auf der Basilarmembran entspricht (Moo-

²Die Form eines Gammaton-Filters kann variieren und lässt sich nicht immer durch eine einzige Zahl beschreiben. Die ER-Bandbreite ist ein alternatives Maß zur Beschreibung der Bandbreite eines Filters.

re & Glasberg, 1987). Das Ergebnis dieser Stufe des Modells ist ein auf 128 einzelne Frequenzkanäle aufgeteiltes Signal.

Die Anzahl der verwendeten Gammatonfilter variiert jedoch und scheint von der Art des zu lösenden Problems abzuhängen. In einer Simulation von Meddis und O'Mard (1997) werden z.B. statt 128 nur 60 Gammatonfilter verwendet. Weitere Variationen finden sich im *auditory image model* von Patterson (2000).

4: Die Haarzelle

In der vierten Stufe wird das bislang mechanische Signal (nach Stufe 3 nun 128 Signale) in ein neuronales Signal gewandelt. Die physiologische Grundlage für diese Transduktion sind die zwischen den Membranen der Cochlea angeordneten Haarzellen. Die Bewegung der Basilarmembran führt zu Scherkräften zwischen der Tektorialmembran und den Haaren der Haarzellen (Stereocilia). Das Spannen der Stereocilia wiederum veranlasst die Freisetzung von Neurotransmittern, was zu einem Auslösen von Aktionspotentialen im auditiven Nerv führt (Moore, 2003). Kurz, eine Haarzelle feuert, wenn ein Signal Frequenzen enthält, dass die Basiliarmembran in der Nähe der Haarzelle bewegt.

Im Modell wird diese Umwandlung durch das Haarzellenmodell von Meddis (1986) erreicht. Für jeden der 128 Kanäle aus Stufe 3 wird im Haarzellenmodell die Wahrscheinlichkeit für das Auftreten von Spikes, die Wahrscheinlichkeit, dass diese Haarzelle in diesem Moment feuert, berechnet. Im Modell wird angenommen, dass die Wahrscheinlichkeit linear von der Menge der Transmittersubstanz im synaptischen Spalt zwischen Haarzelle und auditivem Nerv abhängt. Die Menge der Transmitter hängt wiederum von der Permeabilität der Haarzellenmembran ab, welche von der Amplitude des eintreffenden Signals moduliert wird.

Das Modell hat eine Vielzahl von Parametern, wie z.B. Sättigungsrate, spontane Feuerrate oder Dynamikbereich. Rein rechnerisch umfasst das Haarzellenmodell die Permeabilitätsfunktion und drei Differentialgleichungen. Die Ausgabe entspricht bei einem Signal mit hoher Amplitude etwa dem gleichgerichteten³, gefilterten Inputsignal.

Das Haarzellenmodell ist in der Lage, wichtige Eigenschaften des auditiven Nerves zu simulieren. So können z.B. Adaptation, on-set Verhalten und phase-locking modelliert werden. Eine ausführliche Beschreibung aller Parameter und

³Gleichgerichtet meint hier half-wave-rectified (h-w-r). Ein Signal wird h-w-r, indem der negative Teil der Amplitude auf Null gesetzt wird.

Eigenschaften des Modells findet sich in Meddis (1988, 1986).

5: Refraktäreffekte

Die Wahrscheinlichkeit, dass ein Neuron feuert, hängt von der vorhergehenden Aktivität dieses Neurons ab. Die Refraktärphase, in der der Potentialunterschied zwischen dem Inneren der Zelle und ihrer Umgebung hergestellt wird, um erneut feuern zu können wird durch die fünfte Stufe des Modells simuliert. Durch eine Funktion der vergangenen Aktivierung wird die Wahrscheinlichkeit zu feuern neu berechnet. Diese Transformation beeinflusst vor allem pulsartige Stimuli mit längeren Phasen von Ruhe. Bei stationären Reizen mittlerer Lautstärke hat die Refraktärphase nur wenig Auswirkung auf die Wahrscheinlichkeit zu feuern (Meddis & Hewitt, 1991a).

6: Verteilung der Zeitintervalle

In der sechsten Modellstufe wird in einem Histogramm die Verteilung der Zeitintervalle zwischen den Spikes aus Stufe 5 berechnet. Zu einem Zeitpunkt t werden dazu alle Spike-Intervalle der letzten z.B. $t - 2.5$ ms von einem der 128 Kanäle zu einem Histogramm vereint. Dieses stellt nun eine Grundlage für die Schätzung der Periode dieses einen Kanals dar. Die Limitierung der Integrationszeit auf 2.5 ms geht auf Licklider (1959) zurück, welcher dafür keine Erklärung liefert.

Rechnerisch entspricht die Bildung von Interspike-Intervall-Histogrammen einer permanent laufenden Autokorrelationsfunktion, welche im Modell für jeden Kanal berechnet wird.

7: Summation über die Kanäle

In dieser Stufe werden die Autokorrelogramme aller 128 Kanäle ungewichtet gemittelt, es entsteht ein neues summiertes Autokorrelogramm (SACF).

8: Schätzung der Tonhöhe

Jeder Punkt in der SACF aus Stufe 7 entspricht einem Zeitintervall. Der höchste Punkt entspricht der Periode, dessen korrespondierende Frequenz als Schätzer für die Tonhöhe fungiert. Für den Fall, dass in der SACF mehrere qualitativ gleich hohe Spitzen (peaks) vorhanden sind, so Meddis und Hewitt (1991b), ist der Peak mit der kleinsten Periode (der höchste Frequenz) zu wählen.

Eigenschaften des Modells

In zahlreichen Untersuchungen wurde die Vorhersage des Unitary Pitch Modells bei Verwendung unterschiedlichster Stimuli getestet und mit empirischen Befunden verglichen. Meddis und Hewitt (1991b) berichten, dass das Modell z.B. in der Lage ist zahlreiche psychoakustische Phänomene (missing fundamental, pitch-shift, am-modulation, deleyed noise, ...) zu modellieren.

Der Prozess bei der Modellierung des Problems des fehlenden Grundtons sei auch hier kurz skizziert. Ein Reiz bestehend aus der 3., 4. und 5. Komponente eines harmonischen komplexen Tones (600 Hz, 800 Hz, 1000 Hz) mit einer fehlenden Fundamentalen von 200Hz wird trotz Abwesenheit der unteren Frequenzkomponenten mit einer Tonhöhe entsprechend 200 Hz wahrgenommen. Mittels der Gammaton-Filterbank wird der Stimulus in unterschiedliche Frequenzbereiche (Kanäle) aufgeteilt. Auf diese Weise finden sich neben Kanälen mit Mischsignalen auch die drei Komponenten des Reizes in einzelne Kanäle wieder. Die ACF angewandt auf jeden Kanal führt zu Signalspitzen bei 1.66, 3.33, 5.00, 6.67, 8.33,..., ms für die 600 Hz Komponente, 1.25, 2.5, 3.75, 5.0, 6.67, 8.33,..., ms für die 800 Hz Komponente und 1,2,3,4,5,6,..., ms für die 1 kHz Komponente. In der SACF, welche die einzelnen Kanäle aggregiert, zeigt sich eine deutliche Spitze bei 5 ms, der gemeinsamen Periode der einzelnen Komponenten. Die korrespondierende Frequenz und somit die geschätzte Tonhöhe ist folglich $1/0.005 \text{ s} = 200 \text{ Hz}$.

4.3.3 Das Auditory Image Modell

Das *Auditory Image* Modell von Patterson (1987) unterscheidet sich im Aufbau nicht wesentlich vom Modell von Meddis und Hewitt (1991b). Beide Modelle verwenden die gleiche periphere Filterung und die Gammaton-Filterbank von Patterson (1987). Zu Transduktion der Signale, also die Umwandlung von mechanisch in neuronal codierte Informationen, wird auch hier das Haazellenmodell von Meddis (1986) verwendet. In beiden Modellen wird die Tonhöheninformation aus zeitlich repräsentierten neuronalen Signalen extrahiert. Statt der Wahrscheinlichkeit, dass ein Neuron feuert arbeitet das auditory image model mit diskreten Signalspitzen (spikes).

Ein wesentlicher Unterschied findet sich in dem Mechanismus zur Extraktion der Tonhöhe. Während im Modell von Meddis und Hewitt (1991b) die Tonhöhe durch den höchsten Punkt in der summierten Autokorrelationsfunktion (SACF) definiert ist, beschreibt das Auditory Image Modell das Perzept auf der Basis des sog. auditiven Bildes.

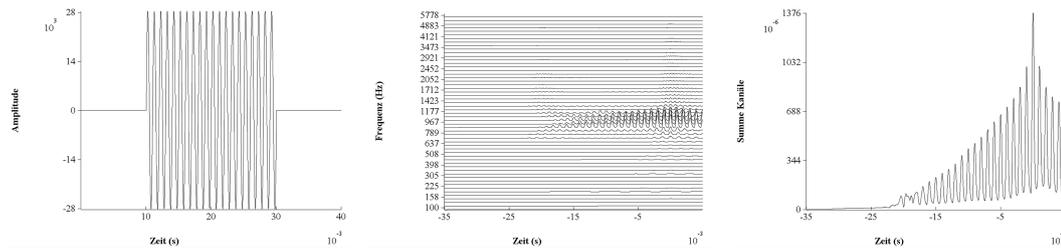


Abbildung 4.5. Simulation mit dem Auditory Image Modell bei Verwendung eines 1kHz Sinustones mit 60 dB Lautstärke. Abgetragen sind der Stimulus (links), die Ausgabe der STI-Stufe (mittig) und die Summation aller Kanäle zu einem auditiven Bild (rechts).

Vom neuronalen Aktivitätsmuster zum auditiven Bild

Nach der Verarbeitung des eintreffenden Signals durch das Haarzellenmodell und der Refraktäreinheit des auditiven Nervs, liegt der Stimulus in einer bestimmten Anzahl von Frequenzkanälen vor. Die finale Stufe des Auditory Image Modell ist die Transformation der neuronalen Aktivitätsmuster des auditiven Pfades in sogenannte stabilisierte auditive Bilder (stabilised auditory images, SAI). Ein SAI ist eine zusammenfassende Darstellung der neuronalen Signale eines oder mehrerer Kanäle des auditiven Pfades für ein bestimmtes Zeitintervall. Anders als die von Meddis und Hewitt (1991b) verwendete ACF, die mit dem zeitlichen Verlauf des Inputsignals korrespondiert, ist ein SAI eines Kanals eine Art Standbild des aktuellen Reizes in diesem Frequenzbereich.

In Abbildung 4.5 sind die Ausgaben von drei Stufen des Auditory Image Modells bei Eingabe eines 1 kHz Sinustones (60dB) mit einer Dauer von 20 ms dargestellt.

Die von Patterson (2000) beschriebene Methode zur Erstellung eines SAI ist die wiederholt abgetastete zeitliche Integration (strobed temporal integration, STI). Dabei wird angenommen, dass das auditive System (1.) über einen Mechanismus zur Verzögerung der Muster im auditiven Pfad verfügt (bank of delay lines), sowie (2.) über eine Reihe von Zeitintervall-Histogrammen, in denen die Intervalle zwischen den einzelnen Signalspitzen aggregiert werden.

Der *strobe* Mechanismus, der die Integration der Zeitintervalle steuert, lässt sich als adaptive Schwelle beschreiben, welche auf die Feuerspitzen eines Kanals reagiert. Nachdem eine Spitze registriert wurde, sinkt die Schwelle etwa um 5%/ms. Wird nach weiteren 5 ms keine größere Signalspitze registriert, so wird die letzte Signalspitze als Impuls (strobe pulse) deklariert und die STI wird initiiert.

Zur STI wird für jeden aktuell im Puffer vorhandenen Impuls das Zeitintervall zu den benachbarten Impulsen berechnet. Im Histogramm wird nun der korrespondierende Wert um einen von der Höhe des Impulses abhängigen Betrag erhöht. Die im Puffer gespeicherten Impulse unterliegen ebenfalls einem Verfall von 2.5%/ms, was dazu führt, dass ein auditives Bild bei Abwesenheit eines Schallsignals langsam zerfällt. Das Histogramm ist auf Zeitintervalle kleiner als 40 ms beschränkt. Zeitintervalle zwischen Impulsen, die länger sind als 40 ms und somit einer Frequenz kleiner als 25 Hz entsprechen, werden nicht mehr beachtet.

Autokorrelation versus strobed integration

Als ein Vorteil der STI gegenüber der ACF nennen Patterson und Irino (1998) die Sensitivität für Symmetrie neuronaler Aktivierungsmuster. Patterson argumentiert, dass die durch Sprache und natürliche Töne hervorgerufenen neuronalen Muster hochgradig asymmetrisch sind und diese Asymmetrie (i.S. der gleichen positiven und negativen Auslenkung der Wellenform) ein wesentlicher Bestandteil bei der Wahrnehmung von Klangfarbe (timbre) ist. Bei Anwendung der ACF, welche stets symmetrische Muster produziert, wird die Asymmetrie zerstört. Die Modelle auf der Grundlage der ACF können folglich nicht die feinen Unterschiede in der Klangfarbe natürlicher Klänge vorhersagen.

4.4 Integration von Orts- und Zeittheorien

Betrachtet man die zahlreichen, sowohl örtlichen als auch zeitlichen, Modellansätze, so stellt sich die Frage nach der Plausibilität beider Theorierichtungen. Beide Theorienrichtungen sind in der Lage, die Tonhöhen unterschiedlicher Stimuli vorherzusagen. Sieht man von einigen Limitierungen ab, so scheint sowohl die örtliche als auch die zeitliche Information für die Modellierung einschlägiger auditiver Phänomene hinreichen zu sein.

Bei der Klärung der Frage, welche Theorierichtung nun die „wahre“ sei, kommt dem Grad der Auflösung der Signale (siehe 3.1.4) eine Schlüsselrolle zu. Grund dafür ist, dass zeitliche Theorien mit den Mustern der nicht aufgelösten Teiltönen (unresolved) oder Residuen arbeiten, während Ortstheorien einzelne, aufgelöste Teiltöne (resolved) benötigen, um die Periode zu ermitteln. Vereinfacht ausgedrückt: Reize mit aufgelösten Harmonischen enthalten Ortsinformation und Reize mit nicht aufgelösten Harmonischen enthalten Zeitinformation.

Carlyon und Shackleton (1994) führten dazu ein viel zitiertes Experiment durch, in dem die Versuchspersonen u.a. aus einem Paar sequentiell dargebotener harmonisch-komplexer Töne den höheren Ton eines Paares wählen sollten. Die Stimuli waren so gefiltert, dass sie entweder auflösbare (resolved) oder nicht auflösbare (unresolved) Harmonische enthielten. Variiert wurde die Zusammensetzung der Tonpaare (resolved-resolved, resolved-unresolved, unresolved-unresolved). Die Diskriminationsleistung für den Vergleich von aufgelösten Harmonischen (resolved-resolved) lag über der von Paaren mit nicht aufgelösten Harmonischen (unresolved-unresolved). Dieser Befund ist nicht verwunderlich, da bekannt ist, dass Reize mit aufgelösten Teiltönen eine prominente Tonhöhe haben. Die geringste Diskriminationsleistung jedoch findet sich im Vergleich von aufgelösten mit nicht aufgelösten Harmonischen (resolved-unresolved). Dieses Ergebnis nehmen die Autoren als Evidenz für zwei unterschiedliche Tonhöhenmechanismen, da hier ein Vergleich über zwei separate Systeme angenommen wird.

In einem Lernexperiment bei dem Versuchspersonen die F_0 -Diskrimination von Tonpaaren trainierten, konnten Grimault, Micheyl, Carlyon und Collet (2002) zeigen, dass das Training mit aufgelösten Reizen sich auf die Diskriminationsleistung von aufgelösten Reizen anderer Frequenzbereiche auswirkt, nicht jedoch auf Paare mit nicht aufgelösten Harmonischen. Gleiches gilt für den umgekehrten Fall. Da kein Lerntransfer von Reizen mit zeitlichen Charakteristika (unresolved) auf Reize mit Ortseigenschaften (resolved) erfolgt, postulieren die Autoren zwei separate Tonhöhenmechanismen.

Houtsma und Smurzynski (1990) konnten nachweisen, dass ein Mechanismus zu Extraktion von Tonhöhe existiert, welcher auf der Basis rein zeitlicher Information operiert und bei Frequenzen des zehnfachen der Fundamentalen wirksam wird. Zu ähnlichen Ergebnissen kamen Kaernbach und Bering (2001) bei Experimenten mit hochpassgefilterten Klicks.

Ein Kritik an den Ortstheorien stellt auch das 'Problem der Phase' dar. Bei komplexen Reizen kann die Variation der Phase der Teiltonfrequenzen einen Einfluss auf die Tonhöhe ausüben (Bilsen, 1973). Wobei z.B. eine über die Teiltöne randomisierte Phase nicht zu eine Tonhöhenverschiebung (pitch shift), sondern zu einer weniger eindeutigen Tonhöhenwahrnehmung führt. Auch spielt die Phase eine entscheidende Rolle bei den in Abschnitt 4.5 dargestellten Phänomenen zur binauralen Tonhöhenwahrnehmung. In den vorgestellten Ortstheorien wird die Phase nicht erfasst und kann somit auch nicht als moderierender Faktor fungieren.

McKay, McDermott und Carlyon (2000) stellen die Frage nach der Unabhängigkeit von Orts- und Zeitinformationen in auditiven Signalen („Place and time cues in pitch: Are they truly independent?“). In ihrer Studie mit vier Trägern von Cochleaimplantaten wurde die relativ geringe zeitliche Auflösung dieser Geräte verwendet, um zu überprüfen, ob die Darbietung konsistenter Zeit- und Ortsinformation effektiver ist, als die Darbietung nur einer Information. Die Ergebnisse bei der Messung verschiedener Differenzlimen führen zu der Schlussfolgerung, dass beide Mechanismen unabhängig voneinander arbeiten. Es besteht kein Vorteil bei der Präsentation von konsistentem Reizmaterial (Orts- und Zeitinformation sind vorhanden) gegenüber der Präsentation von Reizen mit nur Ortsinformation oder nur Zeitinformation.

Micheyl und Oxenham (2005) führen erneute Berechnungen zu den von Carlyon und Shackleton (1994) präsentierten Daten durch und schlagen eine alternative Erklärung vor. Speziell der Vergleich der Ergebnisse von Experimenten, in denen Reize auf unterschiedliche Weise dargeboten werden (sequentiell vs. simultan), wird kritisch beleuchtet. So könnten diese Ergebnisse nicht nur mit unterschiedlichen Tonhöhenmechanismen, sondern auch mit wesentlich basaleren, unterschiedlichen Wahrnehmungsmechanismen erklärt werden.

Zusammenfassend kann behauptet werden, dass Zeit- und Ortsinformation von unserem auditiven System gleichermaßen genutzt und zentral zu einem konsistenten Perzept integriert werden. Analogien hierzu finden sich in anderen Wahrnehmungsqualitäten, z.B. die zahlreichen mono- und binokularen Tiefenkriterien zur Wahrnehmung von Entfernung. Redundanz, wie das parallele Extrahieren und Verwerten von örtlicher und zeitlicher Information, ist nicht ausschließlich Mehraufwand, sondern ein vermutlich evolutionärer Vorteil. Neuere sog. spektro-temporale Modelle versuchen beide Theorierichtungen zu integrieren (Hall, Haggard & Fernandes, 1984, Moore, 2003).

Letztlich scheint die Frage um die Mechanismen bei der Tonhöhenwahrnehmung, welche sich im Kern um die Gültigkeit von Orts- und Zeitmodellen dreht, jedoch nicht eindeutig geklärt zu sein. Um mit den Worten von Plack (2005) zu diesem Thema zu schließen: „The debate rages on.“

4.5 Theorien zur binauralen Tonhöhenwahrnehmung

Bei der binauralen Wahrnehmung wird das Zusammenspiel der Informationen aus beiden Ohren betrachtet. Im Gegensatz dazu steht die monaurale Wahrnehmung, welche eine Situation bezeichnet, bei der nur einem Ohr ein Reiz dargeboten wird. Ist ein binaural präsentierter Stimulus für beide Ohren identisch, so wird dieser als diotisch bezeichnet. Hingegen spricht man von einem dichotischen Reiz, wenn beiden Ohren unterschiedliche Informationen dargeboten werden. Davon abzugrenzen sind die Begriffe mono und stereo, welche sich auf eine Eigenschaft des distalen Reizes, nämlich die Anzahl der potentiell verschiedenen Klangquellen, beziehen. So kann ein mono dargebotener Stimulus diotisch sein, wenn er über Kopfhörer präsentiert wird und dichotisch, wenn er z.B. über Lautsprecher in einem Raum wiedergegeben wird.

Ein zentrales Thema beim binauralen Hören stellt daher auch die räumliche Wahrnehmung mit z.B. Fragen zur Lokalisation von sog. Hörobjekten in Bezug auf Richtung und Distanz dar. Auch für die Wahrnehmung von Tonhöhe lassen sich interessante Fragen ableiten. Wie z.B. ändert sich die Wahrnehmung der Tonhöhe, bei Darbietung dichotischer Reize?

Sind die Stimuli auf beiden Ohren konsistent, so wird für gewöhnlich eine Tonhöhe wahrgenommen. Werden zwei Sinustöne mit leicht variierender Frequenz auf jeweils einem Ohr präsentiert, so werden diese zu einer Tonhöhe fusioniert. Die maximale Abweichung für die eben noch eine Verschmelzung zweier Tonhöhen stattfindet, liegt zwischen einem Halb- und einem Ganzton (van den Brink, Sintnicolaas & van Stam, 1976). Wird die Differenz größer, so werden binaurale Schwebungen (binaural beats) wahrgenommen, welche—abhängig von der Differenz—als langsam zwischen den Ohren hin- und herbewegend wahrgenommen werden (Moore, 2003). Interessant dabei ist, dass binaurale Schwebungen auch auftreten, wenn ein Sinuston unterhalb der absoluten Wahrnehmungsschwelle liegt (Groen, 1964).

Dass jedes Ohr unterschiedliche Reizinformationen erhält, ist nicht ungewöhnlich, sondern entspricht unserem alltäglichen Hörerleben. Aufgrund von Laufzeit- und Frequenzunterschieden, direktem und indirektem Schall aufgrund von räumlichen Gegebenheiten kommt es zwangsläufig zu Unterschieden zwischen dem rechten und dem linken Signal. Ein zentrales Thema bei der binauralen Tonhöhenwahrnehmung ist der Mechanismus der Integration der Informationen aus beiden Ohren. Bezüglich des Ortes der Entstehung von

Tonhöhe wurden bislang zwei Theorien formuliert.

Bei einem *peripheren* Mechanismus wird davon ausgegangen, dass die Tonhöhe in der auditiven Peripherie, d.h. vor der Kreuzung der Hörnerven, genauer ausgedrückt vor Erreichen der oberen Olivenkerne, getrennt für beide Ohren extrahiert wird. Dieser Ansatz, welcher bis auf die von Helmholtz postulierte Resonanztheorie (Tonhöhe entsteht im Innenohr aufgrund von Resonanzen, Helmholtz, 1863) zurückzuführen ist, scheint heute nicht mehr ernsthaft verfolgt zu werden, wenngleich bestimmte Befunde dafür sprechen.

Ein *zentraler* Mechanismus hingegen geht von einer Fusion kontralateraler Signale aus, auf dessen Grundlagen dann die Tonhöhe bestimmt wird. In einem Experiment von Houtsma und Goldstein (1971) präsentierten diese Paare von harmonisch-komplexen Tönen. Jeder Ton eines Paares bestand aus zwei Obertönen, so dass der Grundton, welcher mit der Tonhöhe korrespondiert, „virtuell“ erzeugt werden musste (missing fundamental). Aufgabe der Versuchsteilnehmer war es, acht musikalische Intervalle zu identifizieren. In einer zweiten dichotischen Bedingung des Experimentes wurden die beiden Teiltöne eines Tones jeweils einem Ohr dargeboten. Aufgrund der Übereinstimmung der Ergebnisse der dichotischen und der dichotischen Bedingung schlussfolgerten die Autoren, dass ein zentraler Mechanismus bei der Bildung der Tonhöhe beteiligt sein muss, da die unilaterale Reizinformation (ein Sinuston je Ohr) nicht hinreicht, um die virtuelle Tonhöhe des fehlenden Grundtones zu erzeugen. Virtuelle Tonhöhe entsteht demnach erst durch die Fusion kontralateraler Schallinformation.

Ein anderer wesentlicher Beleg für einen zentralen Mechanismus findet sich bei dem Phänomen der *binauralen Tonhöhe*, welche durch Reize erzeugt wird, die monaural keine bzw. keine eindeutige Tonhöhenempfindung provozieren. Der älteste und zugleich stärkste binaurale Tonhöheneffekt ist die Huggins-Tonhöhe (Huggins pitch).

Bei dem von Cramer und Huggins (1958) entwickelten Stimulus wird auf einem Ohr weißes Rauschen dargeboten und auf dem anderen Ohr das selbe Rauschen mit dem Unterschied, dass ein bestimmter Frequenzbereich in der Phase versetzt wurde. Wie bei den Julesz'schen Random-Dot Mustern in der visuellen Wahrnehmung wird auch hier bei unilateraler Darbietung nur zufällige Information (Rauschen) wahrgenommen (Julesz, 1971). Wird der Huggins-Reiz beiden Ohren dargeboten, kann zusätzlich zum Rauschen ein feiner Ton gehört werden, dessen Tonhöhe einem Sinuston mit der Frequenz des Zentrums des phasenversetzten Bereiches entspricht.

Eine Erweiterung zur Huggins-Tonhöhe berichtet Bilsen (1977). Durch ge-

zielte Phasenverschiebungen bei Frequenzen verschiedener Harmonischer lässt sich das binaurale Pendant zu einem harmonisch-komplexen Ton erzeugen. Bei diesem, als 'Bilsen multiple phase shift pitch' bekanntem Phänomen, lässt sich virtuelle Tonhöhe erzeugen, welche nach Bilsen auf ein 'zentrales Spektrum' zurückzuführen ist.

Ein weiterer Reiz zur Erzeugung binauraler Tonhöhe ist die Kanten-Tonhöhe (binaural edge pitch) von Klein und Hartmann (1980). Der Reiz setzt sich, wie bei additiv erzeugtem Rauschen, aus zahlreichen Sinustönen entlang des hörbaren Frequenzspektrums zusammen. Dabei wird die interaurale Phasendifferenz für Sinustöne bis zu einer festgelegten Frequenz (Kante) konstant gehalten, die Phase der übrigen Sinustöne variiert zufällig. Die feine Tonhöhe, vergleichbar mit der eines Sinustones mit der Kantenfrequenz, kann auf zentrale laterale Inhibition zurückgeführt werden (Hartmann, 2000). Wie bei dem bekannten visuellen Mechanismus der Kantenverstärkung wird hier die binaurale Phasendifferenz genutzt, um Tonhöhe zu erzeugen.

Carlyon, Demany und Deeks (2001) untersuchten die binaurale Verarbeitung von temporalen Reizen und verwendeten Signale bei denen jegliche Ortsinformation durch Filterung extrahiert wurde (unresolved). Die Ergebnisse zeigten, dass bei der Tonhöhenwahrnehmung auf der Basis von Zeitinformation die laterale Reizverarbeitung keinen Einfluss hat und somit ein zentraler, zeitlich orientierter Mechanismus angenommen werden kann.

Diese Beispiele zeigen, dass bei der Extraktion von Tonhöhe ein zentraler Mechanismus zuständig sein muss. Das schließt nicht aus, dass auch periphere Mechanismen involviert sein können. Evidenz dafür liefert die interaurale Tonhöhendifferenz. Ein binaural gehörter Sinuston ruft gewöhnlich eine einzige Tonhöhe hervor. Wird dieser jedoch monaural mal dem linken Ohr, mal dem rechten Ohr dargeboten, so kann häufig festgestellt werden, dass sich die wahrgenommenen Tonhöhen unterscheiden. Dieser, als interaurale Tonhöhendifferenz oder binauraler Diplakusis bezeichneter Effekt ist individuell sehr verschieden ausgeprägt und äußert sich in einer Links-Rechts-Abweichung von bis zu 2% der dargebotenen Reizfrequenz (Terhardt, 1998). Zum Vergleich: ein Halbton überspannt ein Intervall von 6%. Aufgrund von physiologischen Unterschieden zwischen den rechten und linken Komponenten des auditiven Systems entstehen leicht unterschiedliche proximale Reize, welche im Allgemeinen nicht als verschieden wahrgenommen werden. Dieser Effekt macht deutlich, dass bei der Bildung der Tonhöhe auch periphere Mechanismen beteiligt sind.

Unabhängig von der Ursache der lateralen Unterschiede deuten die oben dargestellten Befunde darauf hin, dass ein zentraler Mechanismus existiert,

welcher die Informationen zu einem Perzept integriert. Eine Übersicht über zeitliche Aspekte beim binauralen Hören findet sich in Carlyon et al. (2001).

Binaurale Integration - *Suppression or Fusion*

Die Frage „wo“ Tonhöhe gebildet wird wurde beantwortet. Zur Beantwortung der Frage, „wie“ die lateralen Informationen zu einem Tonhöhenperzept kombiniert werden, werden im Folgenden zwei Mechanismen diskutiert. 1.) Präferenz für ein dominantes Ohr und gleichzeitige Unterdrückung des kontralateralen Reizes (Suppression), 2.) Fusion der binauralen Information und Integration dieser zu einem Perzept (Fusion).

Abhängig von der Art des Reizes und den Bedingungen der Darbietung finden sich in der Literatur Evidenzen für beide Mechanismen. Deutsch und Roll (1976) schlagen im Zusammenhang mit der Erklärung der Oktav-Illusion ein Suppressionsmodell vor. Die Oktav-Illusion entsteht, wenn zwei Töne im Abstand von einer Oktave, jeweils mal dem linken und dem rechten Ohr und vice versa präsentiert werden. Trotzdem sowohl der höhere als auch der tiefere Ton alternierend auf beiden Ohren präsentiert werden, entsteht beim Hörer den Eindruck, dass z.B. der hohe Ton immer links und der tiefe Ton immer rechts dargeboten wird. Zur Erklärung dieses Phänomens gehen die Autoren von einem dominanten Ohr bzw. einer dominanten Seite aus, welche bei der Reizselektion präferiert wird. Die Reize im nicht dominanten Ohr werden demnach unterdrückt.

Trotz der von Deutsch und Roll (1976) vorgeschlagenen Erklärung unterstützen aktuelle Befunde eher das Modell der binauralen Fusion (Chambers, Mattingley & Moss, 2002; Arehart & Burns, 1999; Demany & Semal, 1988).

Pantev, Elbert, Ross, Eulitz und Terhardt (1996) konnten in einer MEG-Studie zeigen, dass eine Repräsentation von Tonhöhe im auditorischen Cortex durch Fusion der lateralen Signale entsteht. Die präsentierten harmonisch-komplexen Töne wurden dabei so konstruiert, dass bei monauraler Darbietung eine Tonhöhe ähnlich eines 500 Hz Sinustones wahrgenommen wurde. Im Fall der binauralen Fusion sollten bei dichotischer Darbietung die rechten und linken Subharmonischen in Kombination eine virtuelle Tonhöhe ähnlich eines 250 Hz Sinustones hervorrufen. Sollte es zu keiner Fusion kommen, so sollte die Aktivierung im auditiven Kortex dem Muster der monauralen Darbietung, also einer wahrgenommenen Tonhöhe bei 500 Hz entsprechen. Die Autoren konnten zeigen, dass das Aktivierungsmuster im auditiven Kortex bei dichotischer Darbietung harmonisch-komplexer Reize dem eines Sinustones mit einer Frequenz

von 250 Hz entspricht. Somit kann von einer Fusion der lateral präsentierten Teilreize ausgegangen werden.

Fasst man die aufgeführten Befunde zusammen, so kann davon ausgegangen werden, dass bei der Wahrnehmung von Tonhöhe ein zentraler Mechanismus operiert, welcher die Reizinformationen aus beiden Ohren fusioniert. Als allgemeine Regel lässt sich dieses Wirkprinzip leicht nachvollziehen, da wir auf diese Weise unsere Hör-Umwelt wahrnehmen. Die kleinen Differenzen zwischen dem linken und rechten Reiz sind Voraussetzung für die akustische Wahrnehmung von Entfernung und Raum. Werden diese Differenzen, wie in einigen Experimenten dargestellt, künstlich überhöht, so schlägt ggf. die Integration der lateralen Informationen fehl. Wo die Grenzen für diese Integrationsleistung liegen, wurde aufgezeigt. Auf weitere Unterschiede zwischen diotischen und dichotischen komplexen Tönen weist Bernstein und Oxenham (2003) hin und extrahiert zwei Faktoren welche für die Integration wesentlich sind, der Grad der Auflösung (resolvability) und die Anzahl der Teiltöne je Seite. Eine Übersicht über Modelle zum binauralen Hören findet sich bei Hartmann und Zhang (2003).

5 Zur Psychophysik oktavkomplexer Töne

5.1 Tonhöhe oktavkomplexer Töne

5.1.1 Tonhöhe und Chroma

Wie eingangs erwähnt, ist Tonhöhe in einschlägigen Definitionen wie auch im alltäglichen Sprachgebrauch als eine eindimensionale Größe konzeptualisiert, welche von tief nach hoch aufsteigt und im umgekehrten Fall absteigt. Alle uns umgebenden Schalle, die eine Tonhöhenempfindung auslösen, können i.d.R. auf dieser Skala eingeordnet werden (vgl. 3.1.3 zur Eindeutigkeit von Tonhöhe).

Der Eindimensionalität der Tonhöhe auf dieser Skala steht eine zirkuläre Größe gegenüber, das *Chroma*. Innerhalb der eindimensionalen Tonhöhen-skala werden Töne, die eine volle Anzahl von Oktaven auseinanderliegen, mit einer ähnlichen „Qualität“ oder „Farbe“ (griechisch Chroma) wahrgenommen (Révész, 1972). Besonders im Bereich der musikalischen Verwendung von Tonhöhe zeigt diese Skala, sei es als Klaviatur auf dem Piano oder als Griffbrett der Gitarre, eine weitere Eigenschaft: sie wiederholt sich nach einer Oktave. In der westlichen Musik teilt man eine Oktave in zwölf logarithmisch gleichabständige Frequenzintervalle sog. Halbtonschritte ein. Die dadurch entstandenen Intervallgrenzen werden mit den Notennamen C, Cis, D, Dis, E, F, Fis, G, Gis, A, B, H bzw. im Englischen mit . . . Gis, A, Ais, B bezeichnet, welche die chromatische Skala wiederkehrend über den hörbaren Frequenzbereich diskret abbilden (Hall, 1991). Ausgehend vom zweifach untergestrichenem C_0 bei 16.4 Hz bis zum siebengestrichenen C_8 bei 16744 Hz wiederholt sich der Oktavraum zehn Mal, ehe die obere Wahrnehmungsschwelle überschritten wird. Die Eigenschaft, welche alle Noten mit gleichem Namen teilen, wird als Tonverwandtschaft oder auch als Chroma bezeichnet. Alle *C*s gehören der Chromaklasse C an, alle *D*s der Chromaklasse D usw.

Ein praktischer Nutzen für solch eine Verwandtschaft liegt darin, dass Instrumente wie auch Stimmen stets miteinander harmonieren, solange sie in der gleichen Chromaklasse spielen bzw. singen. Die Gesangsstimme einer Sängerin passt z.B. zur tieferen Männerstimme, wenn beide unisono, d.h. im selben Chroma, singen. Physikalisch betrachtet liegt die Verwandtschaft darin, dass die Grundfrequenzen von Tönen gleicher Chromaklassen immer Teiler bzw. Vielfache voneinander sind, ihre Perioden passen phasengerichtet ineinander.

Die Tatsache, dass Töne im Abstand von einer Oktave musikalisch als gleichartig angesehen werden, wird als *Oktav-Generalisierung* bezeichnet. Oktav-Generalisierung ist in allen musikalisch höher entwickelten Kulturen zu finden. So unterschiedlich die Tonleitern auch sein mögen, ihre Intervalle sind immer innerhalb einer Oktave definiert (Burns & Ward, 1982).

Empirische Evidenz für Oktav-Generalisierung oder Oktaväquivalenz wurden bei Probanden mit absolutem Gehör, sog. Absoluthörern gefunden. Diese waren in der Lage den Notennamen eines Tones korrekt zu benennen, hatten aber Schwierigkeiten die korrekte Oktave anzugeben (Lockhead & Byrd, 1981). In Experimenten zum Kurzzeitgedächtnis für Tonhöhe konnte Deutsch (1973) zeigen, dass Verwechslungen bei Oktavtönen häufiger auftraten als bei anderen Intervallen.

Es lassen sich somit zwei Dimensionen von Tonhöhe unterscheiden, eine lineare sog. spektrale Tonhöhe von tief nach hoch (pitch height) und die zirkuläre Tonhöhe (pitch class). Das Zusammenspiel beider Tonhöhenaspekte kann am Spiralmodell der Tonhöhe verdeutlicht werden (siehe Abb. 5.1). Hier sind die Töne auf einer Spirale angeordnet, wobei sich die Tonhöhe (pitch height) in der Senkrechten wiederfindet und von unten nach oben ansteigt. Der Abstand der Töne zueinander definiert hier ein Maß für die chromatische Ähnlichkeit. Töne gleichen Chromas finden sich übereinander in der vertikalen Achsen wieder, es zeigt sich die Oktav-Generalisierung. Diese tonale Darstellungsform wurde eingehend von Ruckmick (1929) untersucht und geht auf Arbeiten von Drobisch (1846; zitiert in Ruckmick, 1929) zurück. In einer Studie von Ueda und Ohgushi (1987) konnten diese das Spiralmodell empirisch verifizieren. Ausgehend von Paarvergleichen von oktav-komplexen Tönen konnten die Autoren mit Hilfe der multidimensionalen Skalierung bei der Mehrheit der Probanden die beschriebene Helixstruktur reproduzieren. Eine Erweiterung des Spiralmodells wird von Shepard (1982) beschrieben.

Werden alle Töne mit gleichem Chroma zu einem Ton zusammengefasst, so erhält man eine Tonhöhenklasse (pitch class). Diese künstlich erzeugten Töne finden sich in der Literatur auch unter der Bezeichnung oktav-komplexe Töne

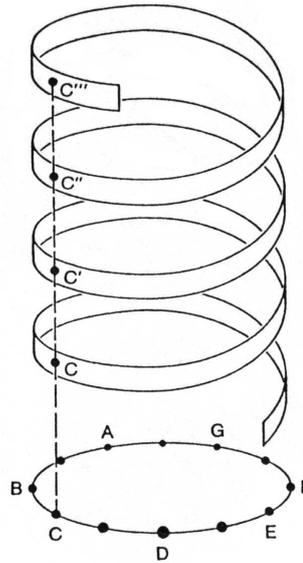


Abbildung 5.1. Im Spiralmodell ist der Zusammenhang zwischen den beiden Dimensionen der Tonhöhe dargestellt. Mit dem Anstieg der spektralen Tonhöhe in der vertikalen Richtung wiederholt sich das Chroma, verdeutlicht als Kreis am Fuße der Spirale.

ne, Tonklassen, Chromaklassen oder Shepard-Töne, benannt nach dem ersten Konstrukteur Roger Shepard (1964). Wie werden diese Töne konstruiert?

5.1.2 Reizbeschreibung

Oktav-komplexe Töne oder Tonklassen sind zusammengesetzte Töne, welche aus Sinustönen im Oktavabstand bestehen. Die Amplitude der einzelnen Sinustöne wird durch eine glockenförmige Hüllkurve definiert (siehe Abb. 5.2).

Gegeben sei eine Tonklasse A bestehend aus n Sinuskomponenten a_1, a_2, \dots, a_n , so ergeben sich die Frequenzen der einzelnen Komponenten durch $f_{ai} = 2^{i-1} f_{a1}$. Im Unterschied zu harmonisch-komplexen Tönen werden hier nicht alle Vielfache der Grundfrequenz berücksichtigt, sondern nur die Komponenten im Oktavabstand zum Grundton (hier nur A_s). Bei einer gegebenen Grundfrequenz von $a_1 = 55$ Hz erhält man 110 Hz, 220 Hz, 440 Hz, 880 Hz, 1760 Hz, 3520 Hz usw. als Obertöne zu a_1 .

Die Amplitude jeder Komponente a_i kann durch folgende Funktion berechnet werden

$$A_{ai}(f_{ai}) = 0,5 - 0,5 \cos \left[\frac{2\pi}{n} \log_2 \left(\frac{f_{ai}}{f_{\min}} \right) \right]. \quad (5.1)$$

Die so erhaltenen n amplitudengewichteten Sinuskomponenten werden ana-

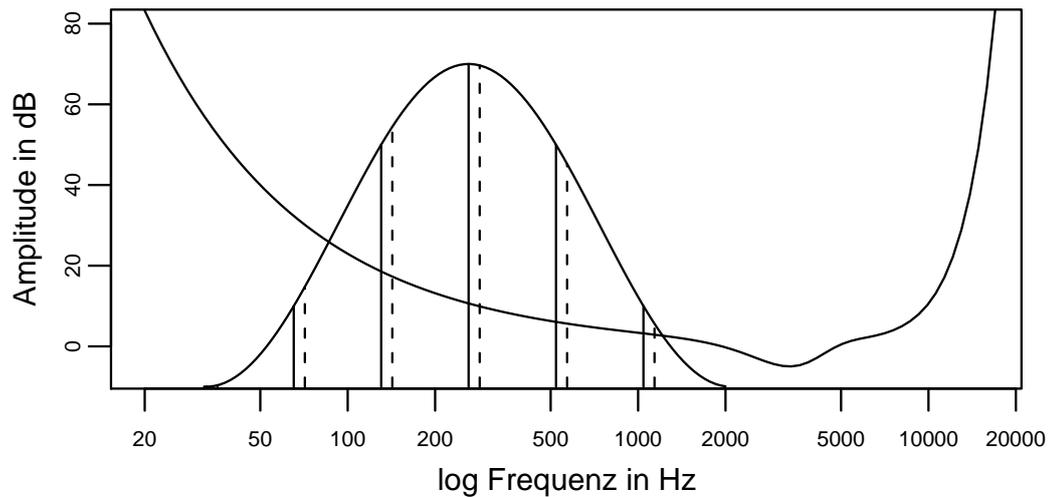


Abbildung 5.2. Die absolute Hörschwelle und die Spektren der Tonklassen C und D. Die Amplituden der Teiltöne werden durch die glockenförmige Hüllkurve definiert (C durchgezogene Linie, D gestrichelte Linie). Es wird deutlich, dass die Amplituden der Teiltöne am äußeren Rand der Hülle deutlich unter der Hörschwelle liegen.

log zur Fourier-Synthese durch Summation zu einem oktav-komplexen Ton zusammengefügt. Die Amplituden-Zeit-Funktion W_a erhält man durch

$$W_a(t) = \sum_{i=1}^n A_{ai} \sin(2\pi f_{ai}t). \quad (5.2)$$

Berechnet man für jeden der 12 Halbtonschritte den zugehörigen oktav-komplexen Ton, so erhält man ein Tonklassenset $(Z_{n,f_{\min}})$. Ein solches Set ist die Grundlage für die mit oktav-komplexen Tönen erzeugten Phänomene wie z.B. das Tritonusparadox oder der Effekt der unendlich aufsteigenden Tonhöhe (Shepard-Illusion). Die Amplituden der Teiltöne aller Tonklassen eines Sets sind durch einer bei f_{\min} festgelegten Hülle definiert. Die relativen Amplituden des Sets $Z_{6,32.7}$ oktav-komplexer Töne sind in Tabelle 5.1 dargestellt. Das ursprünglich von Shepard (1964) verwendete Set bestand aus $n=10$ Komponenten je Tonklasse und einer Grundfrequenz von $f_{\min} = 16.2$ Hz, auch unterteilte er den Oktavraum in 10 Halbtonschritte statt wie üblich in 12. Die Parameter des dargestellten Sets entsprechen den von Deutsch (1986) verwendeten Werten.

Tabelle 5.1

Amplitudengewichte und die zur Komponente a_1 gehörige kleinste Frequenz f_{a_1} am Beispiel des Tonklassensets $Z_{6,32.7}$. Für alle Komponenten dieses Sets ist die Hülle fixiert bei $f_{min} = 32.7\text{Hz}$.

Tonklasse	f_{a_1}	a_1	a_2	a_3	a_4	a_5	a_6
C	32.70	0.00	0.25	0.75	1.00	0.75	0.25
Cis	34.64	0.01	0.29	0.79	1.00	0.71	0.21
D	36.70	0.01	0.33	0.82	0.99	0.67	0.18
Dis	38.89	0.02	0.37	0.85	0.98	0.63	0.15
E	41.20	0.03	0.41	0.88	0.97	0.59	0.12
F	43.65	0.05	0.46	0.91	0.95	0.54	0.09
Fis	46.24	0.07	0.50	0.93	0.93	0.50	0.07
G	48.99	0.09	0.54	0.95	0.91	0.46	0.05
Gis	51.91	0.12	0.59	0.97	0.88	0.41	0.03
A	54.99	0.15	0.63	0.98	0.85	0.37	0.02
Ais	58.27	0.18	0.67	0.99	0.82	0.33	0.01
B	61.73	0.21	0.71	1.00	0.79	0.29	0.01

5.1.3 Tonhöhenwahrnehmung bei oktav-komplexen Tönen

Wie lässt sich die Tonhöhe von oktav-komplexen Tönen bzw. Tonpaaren beschreiben und erklären? Eine korrespondierende Frage wäre, was macht diese Töne so besonders? Phänomenologisch betrachtet klingen oktav-komplexe Töne rau, künstlich und blechern. Das Spezielle dieser Art von Tönen eröffnet sich erst in dem Vergleich mehrerer Töne miteinander. Die ersten Experimente dazu wurden von Shepard (1964) durchgeführt. Die Intention von der Shepard ausgeht, war die Demonstration der Unabhängigkeit von Chroma und spektraler Tonhöhe. Durch das Zusammenfügen von Teiltönen einer Chromaklasse bzw. das Weglassen harmonischer Komponenten bei gleichzeitiger Hüllengewichtung der Teiltöne, wird der Einfluss der spektralen Tonhöhe minimiert. Shepard konnte mit dem wiederholten Abspielen eines Sets von oktav-komplexen Tönen den Eindruck der unendlich auf- bzw. absteigenden Tonhöhe demonstrieren. Dieses Phänomen und die damit verbundene Zikularität in den Tonhöhenurteilen können am Modell des *Tonklassenkreises* veranschaulicht werden. Der in Abb. 5.3 dargestellte Tonklassen- oder Chromaklassenkreis beschreibt den Oktavraum durch die Anordnung gleichabständiger Tonklassen und bildet die

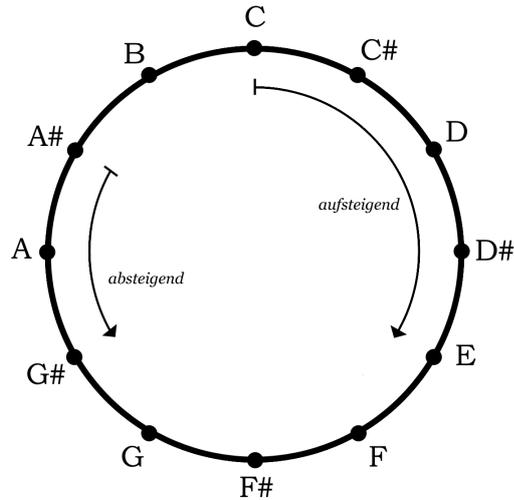


Abbildung 5.3. Der Tonklassenkreis. Die Distanz auf dem Kreis entscheidet über die wahrgenommene Tonhöhenänderung. Das Tonpaar C-E wird als aufsteigend wahrgenommen, weil die kürzeste Distanz zwischen C und E im Uhrzeigersinn verläuft. Im Beispiel A-G ist der Weg auf dem Kreis gegen den Uhrzeigersinn der kürzere. Dieses Tonpaar wird folglich als absteigend wahrgenommen.

Grundlage der von Shepard (1964) postulierten Theorie der relativen Tonhöhe von oktav-komplexen Tönen.

Werden zwei Tonklassen nacheinander dargeboten, so hängt die empfundene Richtung der Tonhöhenänderung (auf- oder absteigend) von der Distanz der Tonklassen auf dem Kreis ab. Ist der Abstand zwischen den Tonklassen im Uhrzeigersinn kürzer als gegen den Uhrzeigersinn, so wird das Tonpaar als aufsteigend wahrgenommen. Ist der Abstand gegen den Uhrzeigersinn der kürzere, so wird das Paar absteigend wahrgenommen (vgl. Abb. 5.3).

Intransitivität zirkulärer Urteile

Transitivität ist eine Eigenschaft von Relationen linearer Ordnung. Aus $x \prec y$ und $y \prec z$ folgt $x \prec z$. Übertragen auf z.B. Sinustöne kann Folgendes illustriert werden. Wird ein Ton y , welcher höher klingt als ein Ton x , mit einem Ton z verglichen, welcher wiederum höher als Ton y empfunden wird, so kann davon ausgegangen werden, dass z auch höher als x wahrgenommen wird, es gilt $x \prec y \prec z$.

Da oktav-komplexe Töne auf dem Chroma- oder Tonklassenkreis angeordnet werden können und somit die Tonhöhe nicht linear, sondern zirkulär abgebildet wird, sind die Relationen zwischen den Tönen *intransitiv*. Bietet man

ein Set von 12 Shepardtönen dar, so gilt $C \prec Cis \prec Dis \prec D \prec, \dots, \prec Ais \prec B \prec C$. Hier zeigt sich das Paradoxon der unendlich aufsteigenden Tonhöhe: die Tonklasse C ist höher als die Tonklasse C.

5.1.4 Das Tritonus Paradox

Die Shepardsche Theorie macht Aussagen über ungleiche Wege im bzw. gegen den Uhrzeigersinn. Somit lässt sich auch die als unendlich auf- bzw. abwärts wahrgenommene wiederholte Abfolge der 12 Tonklassen vorhersagen. Aufeinander folgende Töne liegen im Intervall genau eines Halbtonschrittes, je nach Richtung, im Uhrzeigersinn bzw. gegen den Uhrzeigersinn, wird die Tonfolge als auf- bzw. absteigend wahrgenommen. Im Falle des Halboktav- oder Tritonusintervalls sind die Distanzen im und gegen den Uhrzeigersinn gleich. Gemäß der Theorie sollten somit auch derartige Tonpaare mal auf- und mal absteigend wahrgenommen werden. Zur Erklärung führt Shepard an, dass es sich um einen ambivalenten Reiz handle, welcher ähnlich der Illusion des Necker Würfels mal von oben und mal von unten wahrgenommen werden kann. Shepard selbst konnte zeigen, dass hier auf- bzw. absteigende Urteile im Mittel bei 50% liegen. In seiner Studie mittelt Shepard die Tonhöhenurteile für das Tritonusintervall und erhält konsistent mit seiner Vorhersage ein Tonhöhenurteil bei 50%. Spätere Untersuchungen, vor allem um Diana Deutsch, konnten zeigen, dass Hörer systematisch und stabil bestimmte Tritonuspaare als aufsteigend andere als absteigend beurteilten (Deutsch, 1986, 1987, 1988, 1991, 2007; Deutsch, North & Ray, 1990; Deutsch & Dolson, 2004).

In der Musik wird ein Intervall von sechs Halbtönen oder einer halben Oktave als Tritonus bezeichnet. Folgende Tonkombinationen ergeben Tritonusintervalle: C-Fis, Cis-G, D-Gis, Dis-A, E-Ais, F-B, Fis-C, G-Cis, Gis-D, A-Dis, Ais-E, B-F. Das als Tritonusparadox bekannte Phänomen gewinnt seine Paradoxie erst aufgrund der von Shepard postulierten Theorie zur Vorhersage der Tonhöhenänderung bei oktav-komplexen Tönen. Demnach ist die Tonhöhenänderung abhängig von der Distanz zweier Töne auf dem Tonklassenkreis. In zahlreichen Studien konnte gezeigt werden, dass die Sicherheit der Tonhöhenurteile mit zunehmender Distanz auf dem Tonklassenkreis abnimmt. Je näher ein Paar sich dem Halboktavintervall nähert, je unsicherer werden die Urteile. Bei Paarvergleichen im Tritonusintervall zeigt sich hingegen, dass Hörer bestimmte Tonpaare als sicher auf- bzw. sicher absteigend wahrnehmen. Dies erscheint auf den ersten Blick paradox und widersprüchlich zur Theorie. Welche Paare dabei auf- bzw. absteigend gehört werden, ist zwar individuell verschieden, folgt

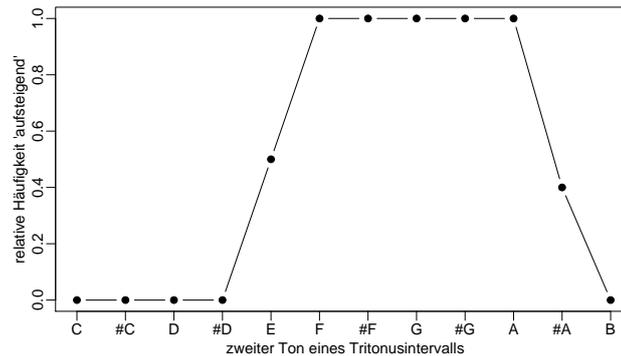


Abbildung 5.4. Antwortmuster einer Versuchsperson bei einem Paarvergleich von oktav-komplexen Tönen im Halboktavintervall. Abgetragen sind die relativen Häufigkeiten der Antwort 'Tonhöhe ist aufsteigend' im Paarvergleich als Funktion des zweiten Tones eines Paares.

jedoch im Allgemeinen der Regel, dass sechs zusammenhängende Tonklassen gegenüber den übrigen sechs stets als höher beurteilt werden. Die Ergebnisse eines Paarvergleichs einer Versuchsperson¹ sind in Abbildung 5.4 dargestellt. Die am höchsten wahrgenommenen Tonklassen sind hierbei zentriert um die Tonklasse G. Ähnliche Antwortmuster bei der Beurteilung der Tonhöhe von Tritonus-Paaren finden erstmals Erwähnung in der Arbeit von Deutsch (1986). Repp (1994) schreibt über seine Reaktion auf die Tritonusaufgabe „In the present experiment, however, I heard all tritone pairs in the $C\#_5$ set and all but two in the G_4 set as falling in pitch“ (S. 651). Hartmann (1998) erwähnt Fälle, in denen die Antwortmuster einer Gleichverteilung mit einer Wahrscheinlichkeit bei 0.5 entsprechen, die betreffenden Personen also nicht zwischen auf- und absteigend diskriminieren können. Wenngleich eine starke Heterogenität der Befunde besteht, wird in der Mehrheit der Studien das eingehend erwähnte 6-auf-6-ab-Muster beschrieben. Auch zur Erklärung dieser Befunde wurden mehrere Theorien postuliert.

Theorien zum Tritonus Paradox

Sprachsozialisation: Nach der von Deutsch postulierten Hypothese besitzen Personen eine mentale Langzeitrepräsentation der von ihnen gesprochenen Sprache. Diese ist innerhalb einer Sprachgemeinschaft stabil und wird

¹Die dargestellten Daten beruhen auf Antworten des Autors in einem der zahlreichen Selbstversuche. Die Stimuli wurden hierbei über die internen Lautsprecher eines Notebooks (IBM T42) dargeboten. Das dargestellte Antwortmuster stellt eine konsistente Beurteilung von Tonhöhenänderungen hinsichtlich Sicherheit und Symmetrie der Urteile dar.

durch Kommunikation erlangt und durch Medien verbreitet. Die Verbindung zum Tritonusparadox gestaltet sich so, dass, der Theorie von Deutsch folgend, ein mentaler Tonklassenkreis mit einer festen Orientierung angenommen wird. Weiterhin wird postuliert, dass die am höchsten wahrgenommene Tonklasse mit der Frequenz bzw. dem Chroma der angenommenen Sprachrepräsentation korrespondiert.

Deutsch (1991) untersuchte Kalifornier und Engländer mit der Tritonusaufgabe und fand konträre Muster in der Verteilung der am höchsten bewerteten Tonklassen beider Gruppen. Die englischen Probanden bewerteten die Tonklassen zwischen Fis und G am häufigsten mit als höchste, während die Kalifornier die Tonklassen zwischen B und Dis als höchste wahrnahmen.

Giangrande (1998) konnte zeigen, dass Hörer aus Florida ähnlich auf dargebotene Tritonusintervalle reagierten wie die von Deutsch untersuchten kalifornischen Probanden. Weitere Untersuchungen finden Unterschiede zwischen Schweden und Briten (Chalikia & Leinfelt, 2000) und Koreanern und Amerikanern (Chalikia, Miller & Vaid, 2001). Krüger und Lukas (2002) vergleichen die Urteile deutscher und japanischer Hörer und finden heterogene Antwortmuster. Zwei Probanden zeigen einen klaren Hülleneffekt, zwei weisen ein über die Hüllenbedingungen stabiles Antwortmuster auf. In der Mehrzahl kommt es zu einer Verschiebung der am höchsten wahrgenommenen Tonklasse.

Auch innerhalb einer Sprache zeigen sich vom Dialekt abhängige Unterschiede in der Bewertung der Tritonusaufgabe, z.B. zwischen Hörern aus Texas und Kalifornien (Chalikia & Vaid, 1999b). Unterschiede zwischen zweisprachig aufgewachsenen und einsprachig aufgewachsenen Probanden untersuchte Chalikia und Vaid (1999a). Die Autoren finden bei in Texas aufgewachsenen Hörern Unterschiede zwischen solchen, die spanisch und englisch sprechen und solchen, die nur englisch sprechen.

In einer Studie von Deutsch et al. (1990) untersuchten die Autoren das Frequenzspektrum von gesprochener Sprache, um dieses mit der Präferenztonklasse beim Tritonustest zu vergleichen. Während die Probanden spontan redeten, wurde die Stimme aufgezeichnet. Die Aufzeichnungen wurden in kurze Sequenzen unterteilt, von den jeweils die fundamentale Frequenz bestimmt wurde. Danach wurde das Oktavband jeder Person ermittelt, in dem die meisten fundamentalen Frequenzen lagen. Bei 8 von 9 Personen stimmten die Oktavgrenzen mit dem Chroma der vorher erhobenen höchsten Tonklasse überein.

In einem Experiment von Deutsch und Dolson (2004) konnten diese zeigen, dass die Sprache, mit der eine Person sozialisiert wird, entscheidend ist für die Wahrnehmung der Tonhöhe bei Tritonuspaaren. Weiter wird angenommen,

dass diese Präferenz für eine bestimmte Chromaklasse in einer, für den Spracherwerb sensiblen Phase angelegt wird und bis ins Erwachsenenalter überdauert. Dies konnte u. a. anhand der Übereinstimmung der höchsten Tonklassen bei Müttern und ihren Nachkommen gezeigt werden (Deutsch, 2007).

Kritikpunkte

Der dargestellte Zusammenhang zwischen Sprache und Tritonusparadox gilt aufgrund der zahlreichen Untersuchungen als fundiert. Gleichsam gibt es Ergebnisse, welche diesen Einfluss kritisch betrachten bzw. widerlegen. Wie oben bereits angedeutet, gibt es interindividuelle Unterschiede in der Fähigkeit, die Tonhöhe zweier oktav-komplexer Töne zu bewerten. So führt Repp (1994) an, dass zahlreiche Personen bereits kleine Intervalle nicht sicher bewerten können. Für den Fall der ambivalenten Tonhöhenänderung beim Tritonus sind somit oftmals keine sicheren Urteile zu erwarten. Für diese Fälle kann die Sprachsozialisationshypothese nicht sicher überprüft werden und somit auch nicht verworfen bzw. angenommen werden.

Ein weiterer Kritikpunkt bezieht sich auf einen spezifischen Parameter beim Reizmaterial, die Position der Hüllkurve. Wenn, wie von Deutsch (1987) postuliert, davon ausgegangen wird, dass Personen eine sprachabhängige mentale Repräsentation der höchsten Tonklassen haben, so sollten auch Tritonuspaare stets in gleicher Weise bewertet werden, unabhängig von der Position der Hülle.

In ihrer Untersuchung verwendet Deutsch (1991) vier unterschiedliche Sets von Tonklassen. Diese Sets unterscheiden sich in der Frequenz bei der die spektrale Hülle zentriert wurde, nämlich 262 Hz (C_4), 370 Hz ($F\#_4$), 523 Hz (C_5) und 740 Hz ($F\#_5$). In der Studie kombiniert Deutsch die Daten verschiedener Hüllenbedingungen mit dem Ziel, potentielle Einflüsse der Hülle zu nivellieren. Über den Einfluss der Hülle macht Deutsch keine Aussagen. Dawe, Platt und Welsh (1998) verwenden in ihrem Experiment die gleichen vier Hüllenparameter und finden keinen signifikanten Unterschied bei der Verteilung der Antworten zwischen den Hüllenbedingungen. Zwei von 20 Personen zeigten einen maximalen Hüllen-Effekt unter allen vier Bedingungen, nämlich eine Verschiebung der peak pitch classes um sechs Halbtonschritte. Bei fünf Probanden zeigte sich diese Verschiebung nur bei einer der vier Bedingungen. Die Mehrheit der getesteten Personen zeigte einen klaren Tonklassen-Effekt, also eine Bewertung der Tonhöhe in Abhängigkeit von der präsentierten Tonklasse. Giangrande (1998) untersucht die Verteilung der peak pitch classes unter den

gleichen vier Hüllen und findet einen Hülleneffekt bei weniger als 15% ihrer Probanden.

Repp (1994) verwendet bei sonst gleichen Parametern Hüllen zentriert bei 440 Hz (A_4) sowie 622 Hz ($D\#_5$) und findet bei mehr als der Hälfte der Personen einen ausgeprägten Hülleneffekt. Auch in weiteren Untersuchungen beschreibt Repp (1997) eine Umkehr des Tritonusurteils durch die Verwendung eines um 6 Halbtonschritte verschobenen Tonklassensets und somit einen signifikanten Hüllen- bzw. Kontexteffekt. Ein ausgeprägter Hülleneffekt wird auch von Chalikia und Vaid (1999b) berichtet. Repp und Thompson (2009) konnten die Ergebnisse von Repp (1997) nicht replizieren und finden in jüngeren Untersuchungen einen geringer ausgeprägten Hülleneffekt.

Auch zu der von Deutsch postulierten Annahme des mental repräsentierten Kreises, welcher in seiner Ausrichtung abhängig von der Sprachsozialisati-on orientiert ist, finden sich keine Belege oder einschlägige neuroanatomische Entsprechungen. Jedoch wird in psychoakustischen Erklärungsansätzen eine Alternative zur Sprachhypothese vorgeschlagen.

Psychoakustische Ansätze

Eine Erklärung zum Tritonusparadox wird von Terhardt und Grubert (1988) anhand der Theorie der virtuellen Tonhöhe geliefert (vgl. Abschnitt 4.2.2). Dabei wird davon ausgegangen, dass Tonhöhe, wie auch Frequenz, ein eindimensionales Konstrukt darstellt und demnach die Tonhöhe aller komplexen Töne ambivalent ist. Werden komplexe Töne dargeboten, existieren stets mehrere potentiell wahrnehmbare Tonhöhen.

Unter den komplexen Tönen stellen die oktav-komplexen Töne eine Extremform hinsichtlich der Ambivalenz der Tonhöhe dar. Sie enthalten keine zusätzlichen Subharmonischen, die als eindeutiges Vielfaches einer bestimmten Frequenz auszumachen wären. Alle enthaltenen Frequenzen sind Vielfache oder Teiler aller übrigen Frequenzen.

Terhardt erklärt die beim Vergleich von Tritonuspaaren gefundenen Antwortmuster durch eine Bevorzugung bestimmter Frequenzbereiche um etwa 300-500 Hz (Terhardt, 1989). Dies führt er zurück auf die in früheren Arbeiten postulierte spektrale Dominanzregion. Die im oktav-komplexen Ton enthaltenen Frequenzen um etwa 300 Hz (D_4) werden stärker gewichtet und fungieren somit eher als Kandidaten für die primär empfundene Tonhöhe.

In einem Abgleichsexperiment ließen Terhardt und Grubert (1988) Probanden die Frequenz von Sinustönen so einstellen, dass sie der Tonhöhe eines

präsentierten oktav-komplexen Tones gleich. Es konnte gezeigt werden, dass vermehrt die Frequenzen der spektralen Dominanzregion eingestellt wurden.

Eine Erweiterung des Modells der virtuellen Tonhöhe wird von Normann, Purwins und Obermayer (2001) vorgeschlagen. Dabei wird aus den Spektren zweier oktav-komplexer Töne ein Differenzspektrum (spektrum of pitch difference), ähnlich der von Terhardt vorgeschlagenen subharmonischen Koinzidenzdetektion (vgl. 4.2.1), abgeleitet. Mit ihrem Experiment bestätigen die Autoren einmal mehr die Zirkularität in der Tonhöhenwahrnehmung bei oktav-komplexen Tönen, auch wenn die Reize eine unterschiedliche Teiltonanzahl und unterschiedliche Amplituden aufweisen. Zur Modellierung der Zirkularität werden die Differenzspektren verwendet. Über die Vorhersage des Modells bei Tritonusintervallen wird nichts berichtet.

Parncutt und Tan (2005) beziehen sich ebenfalls auf die spektrale Dominanzregion zur Erklärung des Tritonusparadoxes. Sie führen die interindividuellen Unterschiede bei der Tritonusaufgabe auf die unterschiedlichen Formen und Position der Dominanzregionen zurück. Auch werden zwei Hörergruppen unterschieden, analytische und holistische Hörer. Dabei beziehen sich die Autoren auf die individuelle Präferenz eher spektrale Tonhöhen oder eher virtuelle Tonhöhen wahrzunehmen, ebenfalls ein Faktor zur Erklärung interindividueller Unterschiede. In Bezug auf intraindividuelle Unterschiede bzw. die Unsicherheit im Tonhöhenurteil führen die Autoren aus, dass die Tonhöhe aller komplexen Töne mehrdeutig sei.

Fazit

Vergleicht man beide Erklärungsansätze, so lassen sich viele Annahmen widerspruchsfrei miteinander kombinieren. So kann der von Deutsch postulierten peak pitch class des mentalen Tonklassenkreises mit der von Terhardt eingeführten spektralen Dominanzregion verglichen werden. Beide erhalten ihre individuelle Prägung aufgrund von Hörgewohnheiten. Im Gegensatz zu den Ausführungen von Deutsch, sagt das Modell der virtuellen Tonklasse einen klaren Hülleneffekt vorher (Terhardt & Grubert, 1988). Überschaute man die Literatur, so erscheint eine Verbindung zwischen Sprache und Tritonusaufgabe als wahrscheinlich, wenngleich die Variation und der Effekt der Einflussgrößen weiterhin diskursiv betrachtet werden können.

Im Folgenden werden die Einflussfaktoren bei der Wahrnehmung der Tonhöhe von oktav-komplexen Tönen im Allgemeinen und bei der Wahrnehmung von Tritonusintervallen im Speziellen dargestellt.

5.1.5 Einflussfaktoren

Sequenzen oktav-komplexer Töne

Die von Shepard postulierte Theorie macht Vorhersagen über die Wahrnehmung von Tonpaaren. Mit zunehmender Distanz im Uhrzeigersinn zwischen den Tönen, sinkt die Wahrscheinlichkeit, ein Paar aufsteigend wahrzunehmen. Überträgt man diese Theorie auf Tonsequenzen, so wird klar, warum ein wiederholt präsentiertes Set oktav-komplexer Töne als unendlich auf- bzw. absteigend empfunden wird. Unter bestimmten Bedingungen kann jedoch gezeigt werden, dass Sequenzen von oktav-komplexen Tönen starken Reihenfolge- oder Kontexteffekten unterliegen.

Giangrande, Tuller und Kelso (2003) präsentierten Sequenzen von Paaren, bestehend aus einem ersten konstanten Ton und einem zweiten Ton, welcher mit jeder Wiederholung chromatisch auf- bzw. abstieg (z.B. C-Cis, C-D, C-Dis, C-E, ...). Aufgabe der Teilnehmer war es, jedes Paar als auf- bzw. absteigend einzuschätzen. Es konnte gezeigt werden, dass es bei einer Darbietung von sequentiell auf- bzw. absteigenden Paaren zu einer signifikanten Verschiebung der Tonhöhenurteile kommt. Bei zufällig dargebotenen Paaren liegt die mittlere Schwelle von aufsteigenden zu absteigenden Urteilen etwa sechs Halbtonschritte vom Startton entfernt (Shepard, 1964). Bei aufsteigenden Paaren verschiebt sich diese um etwa drei Halbtonschritte in Richtung des Uhrzeigersinns, bei absteigenden Paarsequenzen drei Halbtonschritte entgegen dem Uhrzeigersinn. Die Versuchspersonen tendieren dazu, ein gegebenes Urteil beizubehalten.

Anders hingegen verhält es sich, wenn chromatisch auf- bzw. absteigende Skalen vor einem Tritonusintervall dargeboten werden. Dawe et al. (1998) präsentierten ihren Versuchspersonen vor jedem Tritonuspaar 3 Sekunden lang eine auf- bzw. absteigende chromatische Tonfolge bestehend aus Shepardtönen. Das zu beurteilende Tritonusintervall wird häufiger als absteigend bewertet, wenn die vorangestellte Sequenz aufsteigend war und häufiger aufsteigend, wenn die Sequenz gegen den Uhrzeigersinn, also absteigend, dargeboten wurde. Als Erklärung führen Dawe et al. (1998) einen negativen Bewegungsnacheffekt an. Demnach führt eine vorangestellte Sequenz zur Adaptation von Neuronen, welche für die Detektion und Repräsentation von Tonhöhenänderung verantwortlich sind, so dass ein ambiges Tonhöhenurteil, wie es bei einigen Tritonuspaaren der Fall ist, leicht beeinflusst werden kann.

Repp (1997) konnte zeigen, dass die Wahrnehmung von Tritonusintervallen sich durch vorweg präsentierte oktav-komplexe Töne umkehren lässt. So wird das Tonpaar C-Fis eher als aufsteigend wahrgenommen, wenn vorher ein Dis

präsentiert wird. Wird hingegen die Tonklasse A vorangestellt, so tendieren Hörer eher dazu, das Intervall als absteigend zu bewerten. Als Begründung führt Repp die Distanz auf dem Tonklassenkreis zwischen dem ersten und dem letzten Ton an. Der Primingeffekt des vorangestellten Tones auf die Bewertung des Tritonuspaares kann substantiell reduziert werden, wenn das selbe Tritonusintervall wiederum diesem Triplet vorangestellt wird, z.B. C-Fis-A-C-Fis.

Repp und Thompson (2009) präsentierten Tritonuspaare bestehend aus in der Tonhöhe eindeutigen harmonisch-komplexen Tönen vor Tritonuspaaren bestehend aus oktav-komplexen Tönen. Die Autoren finden, entgegen den Ergebnissen von Repp (1997), bei der Hälfte der Probanden keinen Primingeffekt sondern ein stabiles Antwortmuster auf Tritonuspaare. Bei den übrigen Probanden findet sich bei der einen Hälfte ein positiver Primingeffekt und bei der anderen Hälfte ein negativer Primingeffekt.

Ein Vergleich ohne Vergleichston wurde von Ragozzine (2001) durchgeführt, um den Kontexteffekt auszuschließen. Dieser beschreibt einen Versuchsaufbau, bei dem er statt Paaren oktav-komplexer Töne einzelne Töne präsentierte. Die Hörer wurden angehalten, die Tonhöhe des Tritonustones zu schätzen (höher/tiefer), ohne dass dieser dargeboten wurde. Ausgehend von der Theorie des mentalen Tonklassenkreises nach Deutsch (1986) wurde erwartet, dass die einzelnen Töne analog zu ihrem Kreismodell der Höhe nach eingeschätzt wurden und somit bestimmte einzelne Töne eher tiefer als der nicht vorhandene Vergleichston und andere eher höher als der Vergleichston beurteilt wurden. Die Ergebnisse der Einzeltonbedingung sind vergleichbar mit den Ergebnissen der im Vorfeld durchgeführten Tritonusaufgabe.

Diese Befunde machen deutlich, dass es bei der Wahrnehmung von oktav-komplexen Tönen nicht allein auf die Tonpaare ankommt, sondern dem Kontext eine besondere Rolle zufällt.

Einfluss der Hülle

In Abschnitt 5.1.4 zum Tritonusparadox wurde der Hülleneffekt bereits erwähnt. Dabei ist der Einfluss der Position der Hülle auf die Bewertung von Tritonuspaaren, speziell der subjektiv höchsten Tonklasse, gemeint. Im Folgenden soll im Allgemeinen auf die Auswirkungen der Parameter der Hülle näher eingegangen werden.

Wie in Abbildung 5.2 bereits dargestellt wurde, werden die Amplituden der Oktavtöne eines Shepardtones durch eine Hüllenfunktion begrenzt. Die Hülle hat die Funktion eines Band-Pass-Filters, welcher nur Frequenzen ei-

nes bestimmten Bereichs passieren lässt. Die Form der Hülle und somit die Ausprägung der Teiltonamplituden variiert in den verschiedenen Studien. Neben der üblichen glockenförmigen Hüllkurve, wie sie von Shepard (1964) und Deutsch (1986) verwendet wurde, lassen sich unterschiedlichste Formen in der Literatur finden. Nakajima, Tsumura, Matsuura, Minami und Teranishi (1988) verwenden eine trapezförmige Hüllkurve mit identischen Maximalamplituden im Bereich von 250 Hz bis 2 kHz und Steigungen von 20dB/Oktave an den Flanken.

In seinen Experimenten variiert Pollack (1978) unter anderem die Form der Hülle. Ausgehend von einer dreieckigen Hüllkurve über gleiche Teiltonamplituden bis hin zu einer Hüllkurve mit Maxima im oberen und unteren Frequenzbereich erhebt der Autor Tonhöhenänderungen zwischen zwei dargebotenen oktav-komplexen Tönen. In den Ergebnissen zeigen sich Ambivalenzen in Tonhöhenwahrnehmung, welche der Autor als „Shepard-Tonhöhe“ interpretiert. Demnach ist die glockenförmige Hüllkurve nicht notwendig, um die typischen Tonhöhenänderungen zu erzeugen. Die von Pollack vorgestellte Studie ist dahingehend problematisch, da im genannten Experiment die Position der Hülle mit der Frequenz der einzelnen Tonklassen variiert. Position und Form der Hülle sind konfundiert und daher nur bedingt mit den einschlägigen Shepardreizen vergleichbar.

Ein wesentlicher Wirkfaktor bei allen verwendeten Hüllen scheint das Ein- bzw. Ausschleichen der Amplituden zu sein, so dass die Teiltonamplituden im oberen und unteren Frequenzbereich die Absolutschwelle unterschreiten. In einem der vorgestellten Experimente 6.5 wurde u.a. dieser Effekt überprüft. Welche Auswirkung hat das Weglassen der Hülle auf die Bewertung der Tonhöhenurteile bei oktav-komplexer Töne?

Ein Experiment bei dem alle Teiltöne die gleiche Amplitude aufwiesen wurde von Parncutt (1993) durchgeführt. Dieser bot seinen Probanden in drei Experimenten Akkorde dar, indem er simultan 1, 2, 3 oder 4 Shepardreize präsentierte und z.B. erfragte, wie viele Töne in einem Akkord hörbar seien. Ziel war die Erklärung von Akkordverwandtschaften anhand von Tonhöhenähnlichkeit und -gemeindsamkeit von Skalen. Bei den insgesamt drei vorgestellten Experimenten wurden keine Shepardtöne direkt bzgl. ihrer Tonhöhe verglichen. Auch wurde die Zusammensetzungen von Akkorden im Tritonusintervall explizit vermieden. Es wurde argumentiert, dass die Töne zwar mit flacher Hülle konstruiert wurden, bei der anschließenden Verstärkung diese jedoch „abgerundet“ wurden, so dass sie mit Shepard-Tönen vergleichbar waren.

Eine weitere Variation oktav-komplexer Töne wird von Purwins (2005) vor-

gestellt. Dieser verwendet eine sinusförmige Hülle, welche mit jedem oktav-komplexen Ton entgegen der Frequenz verschoben wird. Steigt die Frequenz der Komponenten so steigt auch die Amplitude der tieferen Teiltönen. Der Effekt ist, dass die Tonhöhe steigt, während die Klangfarbe dunkler wird. Auch hier bleibt Zirkularität erhalten.

Einfluss des Teiltonabstandes

Pollack (1978) weist darauf hin, dass, um den Effekt der unendlichen Tonhöhe zu erzielen, eine Mindestanzahl von 3 Teiltönen je oktav-komplexen Ton vorhanden sein muss. Diese müssen vom ersten zum nachfolgenden Ton so nahe zueinander positioniert sein, dass ein direkter Vergleich der Komponenten ermöglicht wird bzw. das Gesetz der Nähe (proximity effect) wirksam wird.

Shepard demonstrierte die Zirkularität der Tonhöhe mit oktav-komplexen Tönen, also Töne mit Teiltönen im Oktavabstand. Burns (1981) konstruierte inharmonische, komplexe Töne mit einem Teiltonabstand kleiner als eine Oktave (z.B. 10 Halbtonschritte) und konnte mit seinem Experiment zeigen, dass auch mit diesen Reizen zirkuläre Urteile erzielt werden können, wenn sie sukzessiv abgespielt werden, wenngleich die Illusion einer unendlich auf- bzw. absteigenden Tonfolge deutlich weniger ausgeprägt empfunden wurde, als bei Shepardtönen. Bemerkenswert war weiterhin, dass sich die Tonfolgen aufgrund ihrer Konstruktion nicht wie üblich nach einer Oktave wiederholte, sondern nach dem nun verminderten Oktavintervall. Die Shepard-Ton Phänomene sind somit nicht auf die Oktave beschränkt.

Teranishi (2001) stellt eine Variante oktav-komplexer Töne vor, bei der die Teiltöne innerhalb einer Oktave einen Dur-Dreiklang bilden. Die Teiltonamplituden eines insgesamt zehn Oktaven umfassenden oktav-komplexen Tones wird dabei durch eine trapezförmige Hülle definiert. Nach Aussage der Autoren lassen sich die so konstruierten Töne mit einem elektrischen Klavier erzeugen. Eine veränderte Variante wird von Nakajima et al. (1988) präsentiert. Zum Eindruck des dadurch erzeugten Phänomens der unendlich auf-/absteigenden Tonhöhe wird angemerkt, „Sometimes, however we need to adopt a special attitude in order to obtain the impression of an endlessly ascending scale“ (Nakajima et al., 1988, S. 3).

Eine kontinuierliche Variante dieser Illusion wurde von Risset (1971) konstruiert. Mit Hilfe analoger Sinusgeneratoren konnte eine stufenfreie, fließende Tonhöhenänderung erzielt werden. Statt einzelne Töne zu synthetisieren, wird

ein „sweep“ erzeugt. Im Gegensatz zur Version von Shepard wird diese Illusion durch das hörbare „Einschleichen“ der vorerst nicht hörbaren Frequenzen geschmälert. Die Pausen zwischen Shepard-Tönen unterstützen den Fokuswechsel ähnlich dem Phänomen der „change blindness“ (Rensink, O’Regan & Clark, 1997).

Schroeder (1986) präsentiert eine weitere mit oktav-komplexen Tönen erzeugte Illusion. Er konstruiert, ausgehend von den Shepard-Tönen, Reize, welche bei doppelter Geschwindigkeit abgespielt, einen Halbton tiefer empfunden werden als bei einfacher Geschwindigkeit. Die Umkehrung gilt für die halbe Geschwindigkeit. Zur Erklärung bezieht sich der Autor auf die hohe Selbstähnlichkeit dieser Reize und verweist auf die fraktale Geometrie. Werden die Frequenzen eines Shepard-Tones bzw. die Bandgeschwindigkeit verdoppelt, ändert sich das grundlegende Frequenzmuster nicht, da jede tiefere Oktavkomponente an die Stelle der nächsthöheren tritt, der Kern des Signals bleibt erhalten.

5.2 Psychometrische Modellierung

Bevor auf die experimentellen Arbeiten näher eingegangen wird, sollen an dieser Stelle zwei Modelle beschrieben werden, welche zur Auswertung der Daten aus den noch folgenden Tonhöhenexperimenten (Kapitel 6) verwendet wurden. Die beiden hier vorgestellten psychometrischen Modelle, das *Modell des subjektiven Tonklassenkreises* von Lukas und Krüger (2001a,b) und das *Vier-Parameter-Modell (P_4)* von Krüger (2003), wurden speziell für Paarvergleiche bei oktav-komplexen Tönen konzipiert und ermöglichen die Vorhersage von Reaktionen einer Versuchsperson bei einem gegebenen Reizpaar.

Anhand der geschätzten Modellparameter lassen sich Aussagen über hypothetische Einflussgrößen bei der Tonhöhenwahrnehmung ableiten. Beide Modelle erwarten als Input ein geordnetes Paar von Ziffern, welche das dargebotene Tonpaar definieren, z.B. (1,3) für das Tonpaar C-D. Abhängig von den gewählten bzw. geschätzten Parametern liefern die Modelle die Wahrscheinlichkeit, ein gegebenes Tonpaar in der Tonhöhe als 'aufsteigend' oder eben komplementär als 'absteigend' zu bewerten.

5.2.1 Das Modell des subjektiven Tonklassenkreises

Das Modell von Lukas und Krüger (2001a) beruht auf der Annahme, dass das von Shepard vorgeschlagene Kriterium, nämlich der Abstand auf dem Tonklassenkreis, die Beurteilung der Tonhöhe aller Tonklassenpaare (inklusive Tritonuspaare) erschöpfend beschreiben und vorhersagen kann. Dazu wird zwischen dem physikalischen und dem psychologischen oder subjektiven Tonklassenkreis unterschieden. Im Gegensatz zum gleichabständigen physikalischen Kreis, der die Töne so repräsentiert wie sie konstruiert sind, stellt der subjektive Tonklassenkreis ein Modell für die Struktur der mentalen Repräsentation dar (siehe Abb. 5.5). Angenommen, ein Ton eines Tritonusintervalls wäre um einen bestimmten Betrag im oder gegen den Uhrzeigersinn verschoben, so würde dies bedeuten, dass es einen kürzeren Weg auf dem Kreis gäbe. Das Distanzkriterium aus der Theorie von Shepard würde somit hinreichen, die Wahrnehmung von oktav-komplexen Tönen beliebiger Intervalle vorherzusagen. Mehr noch könnten qualitative Aussagen im Sinne der Vorhersage von Wahrscheinlichkeiten über die Wahrnehmung von Shepardtönen getätigt werden.

Grundlage des Modells ist die folgende logistische Funktion des Abstandes

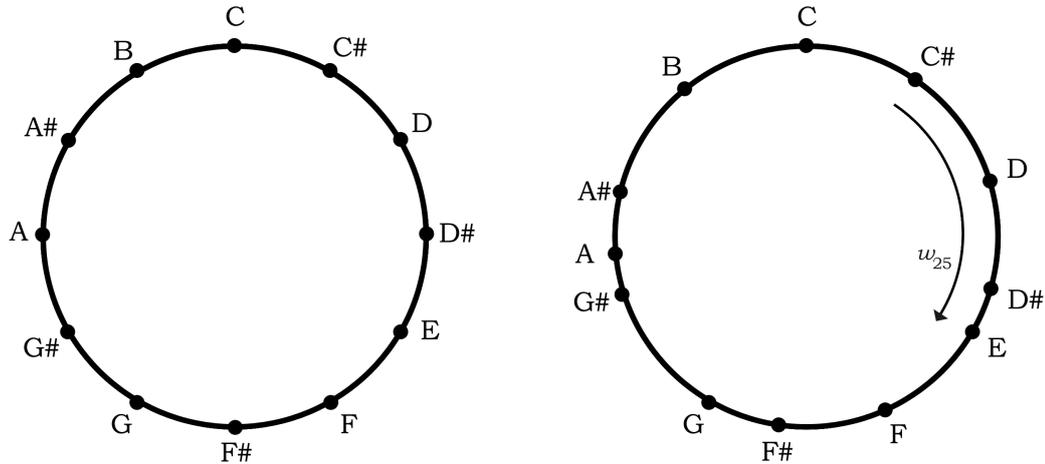


Abbildung 5.5. Der physikalische Tonklassenkreis (links) und der subjektive Tonklassenkreis (rechts). Beispielhaft ist die Distanz auf dem Kreis für das Tonpaar Cis - E illustriert, welches zum Urteil 'absteigend' führt.

zweier Töne auf dem subjektiven Tonklassenkreis

$$p(t_j \succ t_i) = f(w_{ij}) = \frac{1}{1 + \exp^{\frac{w_{ij}-6}{b}}}, \quad (5.3)$$

wobei t_i der i -ten Tonklasse entspricht (konventionell ist $t_1 = C$, $t_2 = Cis$, ... , $t_{12} = B$). Die Position der Tonklasse t_i auf dem Kreis ist durch den aus den Daten geschätzten Parameter v_i definiert. Die Distanz im Uhrzeigersinn zwischen zwei Tonklassen t_i und t_j erhält man mit folgender Funktion

$$w_{ij} = (v_j - v_i + 12) \bmod 12 \quad w_{ij} \in [0, 12[. \quad (5.4)$$

Der Grundgedanke, eine logistische Funktion zu verwenden, ist der, dass bei einem Paarvergleich zweier Tonklassen mit zunehmender Distanz im Uhrzeigersinn (1 bis 11) die Wahrscheinlichkeit, den zweiten Ton als höher wahrzunehmen, gegen Null geht. Die logistische Funktion wird somit hier verwendet, um die Wahrnehmungsschwelle im Bereich der Tritonusdistanz 6 zu beschreiben.

Ein weiterer Vorteil dieser Funktion liegt darin, dass es, genau wie beim Antwortmuster eines Probanden, einen 'sensiblen' mittleren Bereich gibt. In diesem Bereich—genauer gesagt um den Wendepunkt der Funktion—ändert sich die Steigung stärker als in den Randbereichen, und da die Funktion Werte im Intervall $]0,1[$ liefert, ändert sich an dieser Stelle der als Wahrscheinlichkeiten interpretierte Wertebereich am schnellsten.

Die hier verwendete psychometrische Funktion hat zwei Parameter, den Anstieg und die Position entlang der X-Achse. Die Position ist determiniert durch den Wendepunkt der Funktion, welcher bei $f(x)=0.5$ liegt. Dieser auch als Punkt subjektiver Gleichheit bezeichnete Parameter hat im Falle der oben dargestellten Funktion den Wert 6 (siehe Gleichung 5.3). Ein Intervall aus Tönen, die auf dem Kreis genau gegenüberliegen (Distanz 6), würde folglich mit einer Wahrscheinlichkeit von 0.5 als auf- bzw. absteigend bewertet werden. Da Tritonusintervalle in der Regel nicht einheitlich mit der relativen Häufigkeit 0.5 als auf- bzw. absteigend wahrgenommen werden, sondern meist stabil als eines von beiden, werden die jeweiligen Tonpaare so auf dem Kreis zu positionieren, dass der Abstand zwischen ihnen die Antwortmuster auf die Tritonusaufgabe adäquat beschreibt. Verallgemeinert man diesen Sachverhalt auf alle übrigen Distanzen, so erhält man ein Modell zur Beschreibung von Reaktionen auf oktav-komplexe Töne.

Die Schätzung der Wahrscheinlichkeit mit der eine Person einen Ton a höher wahrnimmt als einen Ton b ist somit die transformierte Distanz im Uhrzeigersinn zwischen Tonklassen eines subjektiven, nicht zwingend gleichabständigen Kreises. Das Modell von Lukas und Krüger (2001a) macht genau diese Voraussetzung. Kennt man den subjektiven Tonklassenkreis einer Person, so kennt man auch die Wahrscheinlichkeit mit der diese Person ein bestimmtes Tonpaar in Bezug auf die Tonhöhe beurteilt.

Eigenschaften des Modells

Das Modell des subjektiven Tonklassenkreises besitzt verschiedene Vor- und Nachteile. In der Untersuchung von Lukas und Krüger (2001a) konnten bei einzelnen Probanden hohe Anpassungen erzielt werden. So reizvoll die klassisch-psychophysische Gegenüberstellung von einem gleichabständigen und einem individualisiertem Tonklassenmodell auch zu sein scheint, so schwierig gestaltet sich die praktische Umsetzung. In vorangehenden Studien wurde deutlich, dass die Schätzung der 23 freien Parameter (11 Positionen auf dem Kreis, Starton C ist fix, + 12 Steigungsparameter) ein nicht zu unterschätzendes Anpassungsproblem darstellt (Krüger, 2003). So scheiterte die Funktionsanpassung oftmals an Konvergenzproblemen bei Vorgabe von Daten mit erhöhter Unsicherheit der Hörer bei der Beurteilung der Tonhöhenaufgabe.

Weiterhin lassen sich die geschätzten Parameter nur bedingt interpretieren, wodurch Rückschlüsse auf die Eigenschaften der dargebotenen Reize oder die Fähigkeiten eines Hörers erschwert werden. Die präferierte subjektiv höchste

Tonklasse kann zwar post hoc aus den Schätzungen rekonstruiert werden, eine direkte Ableitung aus den Kreispositionen ist jedoch nicht möglich.

5.2.2 Das P4-Modell

Das P4-Modell schließt an das von Lukas formulierte Modell des subjektiven Tonklassenkreises an und kann als Weiterentwicklung dessen verstanden werden. Ziel dabei war es, die Vielzahl der Schätzparameter auf ein Mindestmaß an interpretierbaren Parametern zu reduzieren.

Zu jedem Startton (C, Cis,..., B) wird auch bei diesem Modell eine logistische Funktion entlang der Distanz (1-11) angenommen. Die Positionen der 12 logistischen Funktionen werden nun durch eine Sinuskurve vorgegeben. Die Parameter dieser Sinusfunktion sind die Amplitude, die Phase und die Position der gesamten Kurve auf der Distanzachse. Der vierte Parameter moduliert die Steigung der logistischen Funktion, welche für alle Starttöne konstant ist (siehe Abb. 5.6).

Im Modell wird davon ausgegangen, dass die Wahrscheinlichkeit, ein gegebenes Tonklassenpaar (t_i, t_j) als aufsteigend zu bewerten, eine Funktion der Distanz zwischen den Klassen auf dem gleichabständigen Tonklassenkreis ist. Geht man dabei von der gerichteten Distanz im Uhrzeigersinn vom ersten zum zweiten Ton aus, so lässt sich diese Distanz d_{ij} wie folgt berechnen

$$d_{ij} = (j - i + 12) \bmod 12 . \quad (5.5)$$

Die Argumente i und j sind dabei die Indizes der einzelnen Chromaklassen $(t_1, t_2, t_3, \dots, t_{12})$, vergleichbar mit den Positionen der Tonklassen auf dem Tonklassenkreis. Gemäß der bisherigen Konvention entspricht auch hier t_1 der Chromaklasse C. Die Distanz im Uhrzeigersinn d_{ij} ist aufgrund des Modulo Operators auf das Intervall $[0,12[$ begrenzt. Zum Beispiel entspricht die Distanz des Tonpaares Dis-A mit $d_{39} = 6$ einem Tritonusintervall. Das Paar E-D ist ein Intervall einer Sekunde, mit $d_{4,2}$ entspricht es jedoch 10 Halbtonschritten im Uhrzeigersinn.

Als Funktion, welche die Distanz eines Tonpaares als Wahrscheinlichkeit und somit auf das Intervall $]0,1[$ abbildet, wurde die folgende zweiparametrische logistische Funktion gewählt

$$p(t_i \prec t_j) = f(a_i, d_{ij}) = \frac{1}{1 + \exp \frac{d_{ij} - a_i}{\beta}} , \beta \neq 0. \quad (5.6)$$

Der Parameter β korrespondiert mit der Steigung der Funktion, je kleiner β desto steiler die Funktion. Der Parameter a_i gibt die Lage der Funktion

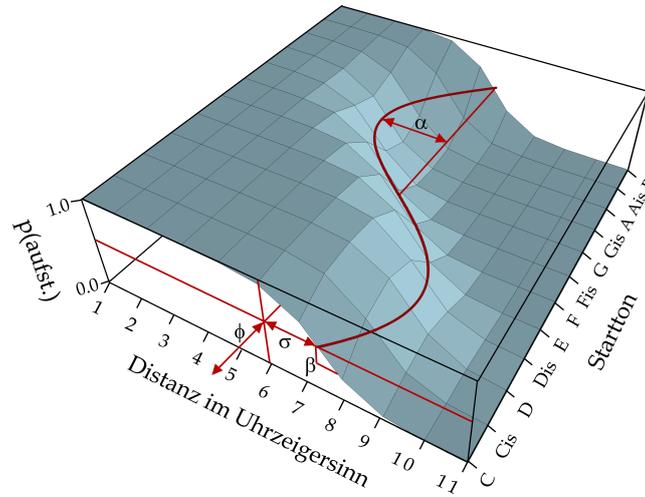


Abbildung 5.6. Darstellung der Parameter des P4-Modells. Abgetragen ist die Häufigkeit, ein Tonpaar als aufsteigend zu bewerten als Funktion des Starttones und der Distanz im Uhrzeigersinn für die Parameter $\alpha=2$, $\beta=1$, $\phi=2\pi/12$, $\sigma=0$.

entlang der Distanzachse an. Er ist eine Funktion des Starttones t_i und wird durch folgende Sinusfunktion moduliert

$$a_i = 6 + \sigma + \alpha \sin\left(\frac{2\pi(i + \phi)}{12}\right). \quad (5.7)$$

Setzt man a_i in Gleichung 5.6 auf der vorherigen Seite ein, so erhält man ein Modell zur Vorhersage der Antwort beim Paarvergleich der Shepard-Töne t_i und t_j . Das Modell behält seine Gültigkeit nur für den Paarvergleich von verschiedenen Tönen. Für den Fall, dass $i = j$, also ein Ton mit sich selbst verglichen wird, erhält man eine Wahrscheinlichkeit $p(t_i < t_j)$ nahe 1. Die Bedingung $i \neq j$ gilt sowohl für das P4-Modell als auch für das Modell des subjektiven Tonklassenkreises.

Ein Vorteil des Modells ist, dass die vier freien Parameter α , β , ϕ und σ nicht nur der besseren Anpassung dienen, sondern auch inhaltlich interpretierbar sind.

Interpretation der Parameter

α (Amplitude): Die Amplitude der Sinusfunktion wird durch den Parameter *Amplitude* beeinflusst. Er gibt das Ausmaß der maximalen Auslenkung entlang der Distanzachse in Halbtönen an. Bildlich lässt sich die Wirkung des Parameters α in einer mal mehr, mal weniger „gewellten Fläche“ beschreiben. Sinnvolle Werte lassen sich im Bereich von 0 Halbtönen

(keine Auslenkung) bis 6 Halbtonschritten (maximale Auslenkung) realisieren. Der Term $2\pi/12$ gewährleistet dabei, dass innerhalb der 12 möglichen Starttöne eine Periode durchlaufen wird und somit positive als auch negative Auslenkung gleichermaßen vorkommt.

ϕ (**Phase**): Der Parameter *Phase* gibt Auskunft darüber, welche Tonklassen bezüglich der Tonhöhe präferiert werden, also im Vergleich zu den übrigen Tonklassen als eher auf- bzw. eher absteigend wahrgenommen werden. Er steht im engen Zusammenhang mit dem Parameter α (Amplitude), welcher das Ausmaß der Präferenz angibt bzw. den Effekt des Tritonusparadoxes wiedergibt. Er korrespondiert mit dem von Deutsch (1986) vorgeschlagenen Maß der *peak-pitch-class* beim Tritonusparadox und kann in dieses überführt werden.

β (**Steigung bzw. Slope**): Der Parameter *Slope* korrespondiert mit der Steigung der Modellfunktion entlang der Distanz-Achse und kann als Maß für die Reliabilität der Urteile interpretiert werden. Er gibt Auskunft über die Sicherheit, mit der ein Hörer die präsentierten Tonpaare als auf- bzw. absteigend beurteilt. Je kleiner der Parameterwert, je sicherer bzw. reliabler erfolgen die Tonhöhenurteile. Dieser Parameter kann Werte im Bereich größer Null bis unendlich annehmen. Zwar sind auch negative Ausprägungen möglich, was die Funktion entlang der Distanzachse um den Wendepunkt spiegeln würde, diese wurden jedoch durch Restriktionen im Schätzalgorithmus unterbunden. Empirisch sinnvolle Werte liegen hier im Bereich von etwa 0.01 (sehr steile Funktion) bis 5 (sehr flache Funktion).

σ (**Bias**): Der Parameter *Bias* gibt Auskunft über die Antwortverzerrung hinsichtlich einer Aufsteigend- oder Absteigend-Antwort-Tendenz und wird in Halbtonschritten angegeben. Beurteilt ein Hörer mehr Tonpaare aufsteigend als absteigend, so äußert sich das in einem positiven Biaswert. Werden vermehrt Tonpaare als absteigend beurteilt, so führt dies zu einem negativen Bias. Sinnvolle Werte finden sich bei diesem Parameter im Bereich kleiner ± 6 , wobei Schätzungen zufolge meist sehr geringe Ausprägungen unter einem Halbtonschritt auftreten.

Zusammengefasst, ϕ beschreibt, *wo* ein Effekt auftritt, α beschreibt, *wie groß* dieser Effekt ist, β beschreibt, *wie sicher* geantwortet wird und σ wie groß der *Unterschied zwischen auf- und ab-Urteilen* ist.

5.3 Modellierung der Tonhöhe oktav-komplexer Töne

Im vorangehenden Abschnitt wurde die Vorhersage des Antwortverhaltens speziell beim Paarvergleich von oktav-komplexen Tönen beschrieben. Im vorliegenden Abschnitt werden die Vorhersagen der beiden in Kapitel 4 eingeführten Tonhöhenmodelle, das Modell der virtuellen Tonhöhe und das Unitary Pitch Modell, bei oktav-komplexen Reizen vorgestellt.

Vorhersage des Modells der virtuellen Tonhöhe

Als Vertreter der Ortstheorie wurde das Modell der virtuellen Tonhöhe mit dem von Terhardt (2004) entwickelten Programm VPITCH realisiert. Dieses ist u.a. auf der Internetseite des Autors erhältlich und liegt in der Version 0.4 als Quelltext der Programmiersprache C vor. Als Eingabe verlangt das Modell eine Liste von Frequenzen und den zugehörigen Amplituden, welche zusammengesetzt das Signal beschreiben. Hier wird deutlich, dass bereits beim Eingangssignal ein Frequenzspektrum erwartet wird und somit eine Zerlegung des ursprünglichen Signals, analog einer Fouriertransformation, bereits vorausgesetzt wird.

Die Ausgabe der Schätzung erfolgt ebenfalls als Liste von frequenzäquivalenten Tonhöhen in pu (pitch units) und deren Gewichtung. Hierbei unterscheidet der Autor zwei Arten von Tonhöhen, die nominale und die wahre Tonhöhe (nominal pitch, true pitch). Die nominale Tonhöhe entspricht der regulären Ausgabe von VPITCH. Zusätzlich zu dieser wird bei der wahren Tonhöhe dem Effekt der Tonhöhenverschiebung (pitch shift) Rechnung getragen (z.B. die Änderung der Tonhöhe bei unterschiedlichen Lautstärken, vgl. Abschnitt 3.1.5). Nominale und wahre Tonhöhe unterscheiden sich im Allgemeinen nur um wenig Prozent voneinander (Terhardt, 2004). Weiterhin wird zu jeder Tonhöhe ausgegeben, ob es sich um eine spektrale oder virtuelle Tonhöhe handelt, also ob die Tonhöhe eine äquivalente Frequenz im Signal findet (spektral) oder ob sie als zusätzliche Subharmonische berechnet wurde (virtuell).

In Abbildung 5.7 (links) ist das Ergebnis der Tonhöhenschätzung des oktav-komplexen Tones C ($f_0 = 32.7Hz$) mit 6 Teiltönen dargestellt. Abgebildet sind die ersten 8 Tonhöhen mit der höchsten Gewichtung (> 0.1). Die prominentesten Tonhöhen zeigen sich bei 130 pu und 261 pu mit jeweils einer Gewichtung von 0.82. Abgesehen von geringen Variationen sind beide Vielfache der Grundfrequenz f_0 und entsprechen damit ebenfalls der Chromaklasse C. Auch wird

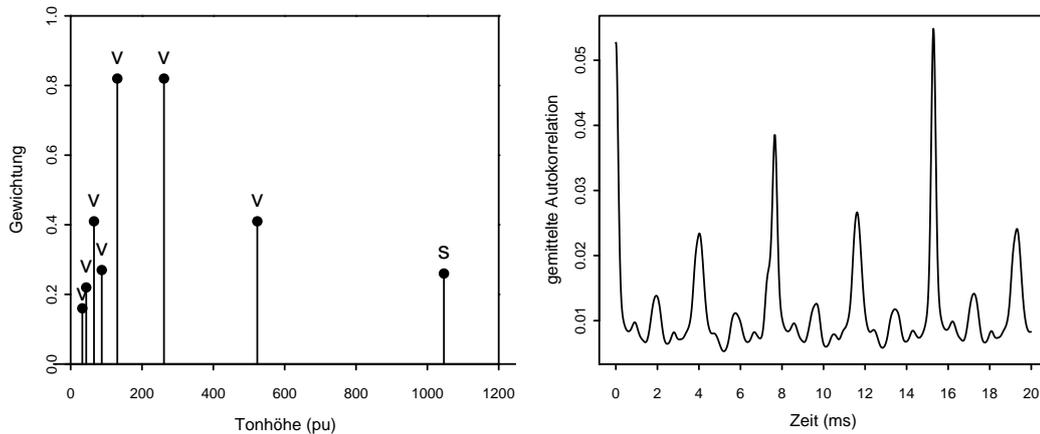


Abbildung 5.7. Die Vorhersage des Modells der virtuellen Tonhöhe (links) und des Unitary Pitch Modells (rechts) bei Eingabe eines oktav-komplexen Tones mit 6 Teiltönen und einer Frequenz f_0 von 32.7 Hz. (C)

deutlich, dass alle Tonhöhen, bis auf die bei 1046 pu, virtuelle Tonhöhen sind. Erwartungsgemäß entsprechen diese Schätzungen den von Terhardt und Grubert (1988) berechneten Tonhöhen.

Vorhersage des Unitary Pitch Modells

Das Unitary Pitch Modell von Meddis und Hewitt (1991b) und Meddis und O'Mard (1997) ist im Vergleich zum Modell der virtuellen Tonhöhe deutlich komplexer. Die Autoren entwickelten kein isoliertes Modell, sondern ein modular aufgebautes Programmpaket, DSAM (Development System for Auditory Modelling). Dieses ermöglicht die Integration unterschiedlicher Teilmodelle (Gehörgang, Cochlea, Haarzelle, etc.) und die freie Wahl aller modellspezifischen Parameter. Neben dem Unitary Pitch Modell wurde auch das Auditory Image Modell von Patterson et al. (1995) mit DSAM realisiert. Die DSAM-Bibliotheken wurden in der Programmiersprache C entwickelt und liegen für verschiedene Plattformen aktuell in der Version 2.8 vor. Daneben wird die grafische Bedienoberfläche AMS (Auditory Modelling System) zum Testen und Visualisieren der Modelle bereitgestellt.

Für die Vorhersage der Tonhöhe von oktav-komplexen Tönen wurde die von O'Mard und Meddis (2007) erstellte Simulationsdatei `unitary.sim` verwendet, welche 17 weitere Parameterdateien (z.B. das Filterbankmodell von Moore und Glasberg (1983): `filtIIRMoore.par`, das Modell der Basilarmembran: `BM-DRNLHuman.par` oder das Modell der inneren Haarzellen von Sumner, Lopez-Poveda, O'Mard und Meddis (2002): `IHC-Meddis2000HSR.par` zu einem

Gesamtmodell kumuliert. Die einzelnen Parameterdateien enthalten wiederum etwa 10 bis 20 weitere modellspezifische, frei wählbare Parameter.

Als Vertreter der Zeit-Theorien erwartet das Modell die Wellenform eines beliebigen Signals als Eingabe, welche durch eine Audiodatei bereitgestellt wird. Die Ausgabe des letzten Moduls, hier die summierte Autokorrelation, erfolgt als Liste von Werten, welche als Funktion der Zeit dargestellt werden können. Das Modellsystem kann aufgrund der Modularität und der hohen Freiheitsgrade sehr flexibel eingesetzt werden. So kann z.B. nach jedem Teilmodell eine Ausgabe der Schätzungen erfolgen.

Die Ausgabe des Unitary Pitch Modells bei Vorgabe des oktav-komplexen Tones C ($f_0 = 32.7$ Hz, $N=6$) ist in Abbildung 5.7 (rechts) dargestellt. Hier finden sich in den ersten 20 ms mehrere deutliche Maxima. Das bei 0 ms entspricht der Korrelation des Signals mit sich selbst. Das nächste relevante Maximum liegt bei 15.3 ms und entspricht, nach Bildung des reziproken Wertes, einer Tonhöhe von 65 pu. Die nächsten Maxima liegen bei 7.7 ms und bei 11.7 ms und sagen potentielle Tonhöhen bei 130 pu und 85 pu vorher.

Vergleich der beiden Schätzungen

Beide Modelle liefern mehrere potentielle Tonhöhenkandidaten, welche weitgehend identisch sind. Vergleicht man die Vorhersagen, so lassen sich unter den zwei jeweils gewichtigsten Kandidaten Schätzungen bei 130 pu finden. Diese korrespondieren mit dem dritten Teilton, dem Vierfachen der kleinsten Frequenz bei 130.8 Hz. Auch liegen die Tonhöhen bei 261 pu (3.8 ms) beim Modell der virtuellen Tonhöhe und bei 257 pu (3.9 ms) beim Unitary Pitch Modell nahe des dritten Oktavtones bei 261.6 Hz. Einige Tonhöhen, wie z.B. bei 613 pu, finden keine Entsprechung im anderen Modell.

Betrachtet man die Tonhöhen mit den höchsten Gewichten, so zeigt sich, dass beide Modelle eine große Schnittmenge an potentiellen Tonhöhenschätzungen aufweisen, das Unitary Pitch Modell jedoch tiefere Kandidaten (65 pu, 85 pu) vorhersagt als das Modell der virtuellen Tonhöhe (261 pu, 523 pu).

Vorhersage bei einem Set von Tonklassen

In Tabelle 5.2 sind die Tonhöhenschätzungen eines vollständigen Sets von Tonklassen der beiden vorgestellten Modelle und deren Gewichtungen aufgelistet. Zusätzlich wurde diesen Modellen die Schätzung der in Abschnitt 4.2.1 vorgestellten Kernalgorithmen, der Subharmonischen Koinzidenzdetektion und der

Tabelle 5.2

Die Tonhöhenschätzungen der beiden vorgestellten Modelle (VP: Virtual Pitch, Unit: Unitary Pitch), der Kernalgorithmen (Sub: Subharmonische Koinzidenzdetektion, ACF: Autokorrelationsfunktion) und deren relative Gewichtungen für alle Tonklassen eines Sets.

Ton	F_{min}	Sub	VP	VP-W	ACF	ACF-W	Unit	Unit-W
C	32.7	130.8	130.0	0.82	271.7	0.82	65.4	0.055
Cis	34.6	138.6	138.2	0.87	289.9	0.81	69.4	0.100
D	36.7	146.8	146.8	0.90	310.5	0.80	73.4	0.082
Dis	38.8	155.5	155.8	0.92	156.2	0.79	77.6	0.098
E	41.1	164.8	165.3	0.93	165.8	0.79	82.3	0.090
F	43.6	174.6	175.4	0.94	175.7	0.80	87.1	0.056
Fis	46.2	185.0	186.0	0.94	186.6	0.80	92.2	0.093
G	48.9	196.0	197.2	0.93	198.0	0.80	97.5	0.065
Gis	51.9	207.6	209.1	0.91	210.5	0.81	51.9	0.090
A	54.9	220.0	221.6	0.89	223.7	0.81	54.9	0.069
Ais	58.2	233.1	235.0	0.88	238.6	0.81	58.2	0.082
B	61.7	246.9	249.2	0.87	254.4	0.82	61.7	0.094

Autokorrelationsfunktion gegenübergestellt. Ein graphischer Vergleich der jeweils gewichtigsten Tonhöhe ist in Abbildung 5.8 dargestellt.

Über die Tonklassen hinweg betrachtet wird deutlich, dass die Tonhöhe meist linear mit der Frequenz ansteigt. So erhält man z.B. beim Vergleich der Frequenz f_0 mit der Tonhöhenschätzung des Modells der virtuellen Tonhöhe eine Korrelation nahe 1.0. Dieses Ergebnis könnte man auch bei Vorhersagen der Tonhöhe von Sinustönen oder harmonisch-komplexen Tönen im vergleichbaren Frequenzbereich erwarten.

Eine Besonderheit von oktav-komplexen Tönen wird beim Blick auf die Tonhöhenschätzung des Unitary Pitch Modells oder der Autokorrelationsfunktion deutlich, hier zeigen sich 'Brüche' in der Vorhersage. Trotz zunehmender Grundfrequenz kommt es zum plötzlichen Abfall der geschätzten Tonhöhe um eine Oktave zwischen D und Dis bzw. zwischen G und Gis.

Vergleicht man die vier Methoden, so zeigt sich, dass die Schätzungen der subharmonischen Koinzidenzdetektion stark den Schätzungen des Modells der virtuellen Tonhöhe ähneln. Die Vorhersage des Unitary Pitch Modells liegt

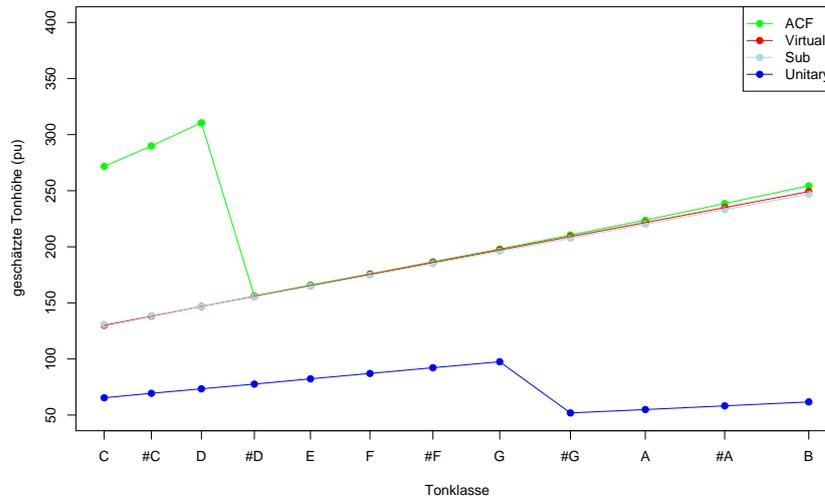


Abbildung 5.8. Die graphische Darstellung der in Tabelle 5.2 präsentierten Tonhöhenschätzungen der vier Schätzmethoden. Abgebildet ist die prominenteste Tonhöhe jeder Tonklasse eines Sets.

meist eine Oktave unter den Schätzungen des Modells der virtuellen Tonhöhe. Zwei Methoden sagen 'Brüche' im Verlauf der Tonhöhe über die Tonklassen hinweg voraus.

Hier sei erwähnt, dass die 'Brüche' in allen dargestellten Modellen vorkommen. In weiteren, hier nicht dargestellten Modelltests mit anderen Reizparametern wurde deutlich, dass z.B. auch das Modell der virtuellen Tonhöhe einen spontanen Abfall der Tonhöhe vorhersagt. Es kann vermutet werden, dass an diesen Stellen die Amplitude einer kleineren Frequenz aufgrund der Hüllenfunktion soweit ansteigt, dass ihre Gewichtung die der vormals primären Tonhöhe übersteigt und somit den Platz der primären Tonhöhe einnimmt. Bei welchen Frequenzen sich dieser Wechsel vollzieht, liegt in den Eigenarten bzw. den Parametern des jeweiligen Modells begründet und lässt sich nicht auf eine bestimmte Chromaklasse reduzieren.

Von der primären Tonhöhe zum Paarvergleich

Mit den vorgestellten Modellen wurde die Tonhöhe von oktav-komplexen Tönen geschätzt. Es konnte gezeigt werden, dass die Vorhersagen, wenn auch nicht immer identisch, so doch vergleichbar waren. Die Phänomene von oktav-komplexen Tönen beruhen nicht auf Einzeltönen, sondern auf Paaren oder Tonreihen, wie z.B. beim Phänomen der unendlich aufsteigenden Tonhöhe. Wie lassen sich diese Phänomene anhand der primären Tonhöhe erklären? Auch

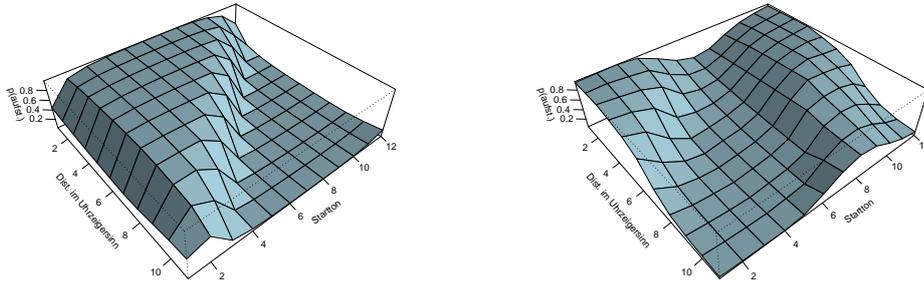


Abbildung 5.9. Die Vorhersage der Antworten von vollständigen Paarvergleichen von oktav-komplexen Tönen mit dem Modell der virtuellen Tonhöhe. Dargestellt sind zwei Schätzungen mit unterschiedlichen Reizparametern (Rechts: 6 Teiltöne, $f_{min} = 32.7$ Hz; Links: 10 Teiltöne, $f_{min} = 16.3$ Hz). Abgetragen ist jeweils die Wahrscheinlichkeit der Antwort 'Tonhöhe ist aufsteigend' als Funktion des Starttones eines Paares und der Distanz im Uhrzeigersinn.

stellt sich die Frage, ob die 'Brüche' in der primären Tonhöhe zur Erklärung der Zirkularität beitragen können?

Allein aus logischen Gesichtspunkten muss die gestellte Frage mit nein beantwortet werden. Anhand der einen dargestellten, primären Tonhöhe ist es nicht möglich, eine zirkuläre Relation abzuleiten. Da die relative Tonhöhen bei oktav-komplexen Tönen eine Funktion des Abstandes zwischen den Tönen auf dem Tonklassenkreis ist, sollten die Tonpaare D und Dis bzw. G und Gis aufsteigend bewertet werden. Dies steht im Widerspruch zur Vorhersage, da hier ein 'Bruch' vorhergesagt wird. Gleiches gilt für die 'Brüche' zwischen B und C in den übrigen Schätzungen.

Dies impliziert nicht, dass die Vorhersage der einschlägigen Modelle falsch ist oder hier nicht anwendbar ist. Zirkularität kann nicht mittels nur einer primären Komponente vorhergesagt werden, sondern bedarf mehrerer Tonhöhen. Dieser Gedanke findet sich bereits in der von Terhardt und Grubert (1988) formulierten Funktion zur Vorhersage der Antworten im Paarvergleich

$$p(t_1 \succ t_2) = \frac{1}{\pi} \arctan[(W_{11} - W_{12})(W_{21} - W_{22})K \operatorname{sign}(H_{11} - H_{21})] + 0.5, \quad (5.8)$$

wobei W_{ij} das Gewicht der j ten Tonhöhe des i ten Tones bezeichnet. H_{ij} bezeichnet die j te Tonhöhe des i ten Tones. K entspricht lt. den Autoren einer willkürlich gewählten Konstanten (hier 250). Mittels dieser Funktion lässt sich die Wahrscheinlichkeit der Antwort 'Tonhöhe ist absteigend' berechnen. Da-

bei beziehen die Autoren zwar nicht mehrere Tonhöhen mit ein, wohl aber die Gewichtung der ersten beiden Komponenten.

Wendet man die Gleichung 5.8 an, so lassen sich die Wahrscheinlichkeiten für die Antworten in einem vollständigen Paarvergleich berechnen. In Abbildung 5.9 sind zwei Paarvergleiche mit unterschiedlichen Reizparameteren dargestellt. In der linken Abbildung sind die Antwortwahrscheinlichkeiten des Sets aus Abbildung 5.8 bzw. Tabelle 5.2 dargestellt. Dabei enthält jeder oktav-komplexe Ton 6 Teiltöne, die Hülle ist fixiert bei $f_{min} = 32.7$ Hz. Es zeigt sich, dass die Antworten stark dem Muster eines harmonisch-komplexen Tones folgen, da die Urteile fast ausschließlich durch das untere und obere Dreieck der Startton-Distanz-Matrix definiert sind. Die meisten Distanzen 1 im Uhrzeigersinn werden als aufsteigend bewertet und die meisten Distanzen 11 als absteigend. Auch das Muster bei Distanz 6 ähnelt der Antwort in der Tritonusaufgabe (6-Paare-absteigend vs. 6-Paare-aufsteigend).

Davon abweichend präsentiert sich das weniger deterministische Antwortmuster in der rechten Abbildung. Hierbei enthält jeder Ton 10 Teiltöne bei einer Frequenz $f_{min} = 16.3$ Hz. Auch zeigt sich, dass hier kleinere Distanzen eher als aufsteigend und größere Distanzen eher als absteigend bewertet werden, ohne scharfe Kontraste zwischen einzelnen Tonklassen hervorzurufen.

Fazit

Diese beiden Beispiele sollen verdeutlichen, dass anhand der geschätzten primären Tonhöhe bereits zahlreiche Effekte von oktav-komplexen Tönen annähernd modelliert werden können. Die Ausprägung von Urteilen aufgrund der Distanz (z.B. alle Distanzen 1-4 stets aufsteigend), wie sie bei experimentellen Untersuchungen zu finden sind (vgl. Abschnitt 6.4), werden nicht erreicht. Es kann somit festgehalten werden, dass die alleinige Berücksichtigung der primären Tonhöhe für die Modellierung der Zirkularität, wie sie beim Tonhöhenvergleich von oktav-komplexen Tönen auftritt, nicht hinreichend ist. Ein Algorithmus, welcher mehrere Teiltonhöhen inkludiert und gleichzeitig Allgemeingültigkeit behält, im Sinne der Anwendbarkeit auf z.B. harmonisch-komplexe Töne, steht weiterhin aus.

6 Experimentelle Untersuchungen

In diesem Kapitel werden die folgenden, im Rahmen des Dissertationsprojektes durchgeführten Experimente, beschrieben, ausgewertet und diskutiert.

1. Vorexperiment I: Hörschwellenaudiometrie
2. Vorexperiment II: Tonhöhendiskrimination
3. Isolierte Präsentation von Tritonusintervallen (Tritonusaufgabe)
4. Vollständiger Paarvergleich: Einfluss der Teiltonanzahl
5. Vollständiger Paarvergleich: Einfluss der Hüllenposition und -form
6. Experiment I zur dichotischen Darbietung oktav-komplexer Töne
7. Experiment II zur dichotischen Darbietung, Erweiterung um das Tritonusintervall

Die sieben Experimente lassen sich grob zu drei Kernthemen zusammenfassen. Die Experimente 1 und 2 dienen der Audiometrie. Mit dem Ziel der Qualitätssicherung wurden diese Tests von allen Hörern absolviert, unabhängig, an welchem Experiment noch teilgenommen wurde. In den nächsten drei Experimenten wird die Wahrnehmung oktav-komplexer Töne anhand einfacher Paarvergleiche untersucht. Die Variation bezieht sich dabei auf isoliert dargebotene Tritonusintervalle (Exp. 3), auf die Anzahl der Komponenten eines oktav-komplexen Tones (Exp. 4) und auf das Vorhandensein und die Position der Hülle zur Gewichtung der Teiltonamplituden (Exp. 5). In den Experimenten 6 und 7 soll mit Hilfe dichotisch dargebotener oktav-komplexer Töne u.a. die Frage beantwortet werden, ob die Tonhöhenwahrnehmung zentralen oder peripheren Mechanismen unterliegt. Aufgrund der Heterogenität der einzelnen Fragestellungen werden diese nicht allgemein an dieser Stelle behandelt, sondern zu Beginn jedes Experimentes.

Methodische Vorbemerkungen

In einschlägigen Publikationen zu oktav-komplexen Tönen wird in aller Regel dem Umstand der Zirkularität der Chroma-Dimension nicht gerecht. Dies äußert sich insbesondere darin, dass Statistik für lineare Daten auf zirkuläre Dimensionen angewandt wird. Um Rechenfehler zu vermeiden, wird somit oft auf schwächere Skalenniveaus oder die Einführung zusätzlicher Faktoren zurückgegriffen, was üblicherweise ein Verlust an Information bedeutet. In der vorliegenden Arbeit wird daher neben den üblichen Verfahren bei Vorliegen zirkulärer Daten, wie z.B. den Chromaklassen, *zirkuläre Statistik* verwendet. Die Möglichkeiten, zirkuläre Statistik zu rechnen wird im Allgemeinen durch Limitierungen der verwendeten Software begrenzt. Die hier durchgeführten Berechnungen basieren auf dem Zusatzpaket `circular` der freien Statistiksoftware R (R Development Core Team, 2010).

Exkurs: Zirkuläre Statistik

Der folgende Exkurs soll die Notwendigkeit zirkulärer Statistik am Beispiel des Mittelwertes bei zirkulären Daten demonstrieren. Zirkuläre oder auch directionale Daten liegen immer dann vor, wenn die Ausprägungen der zugrundeliegenden Dimension regelmäßige Wiederholungen aufweist, wie z.B. bei Wochentagen, Winkelmaßen oder den Chromaklassen (Fisher, 1993). Aufgrund dieser Eigenschaft würde die Anwendung der sonst üblichen linearen Statistik zu falschen Resultaten führen. Die folgende, kurze Rechnung soll diesen Sachverhalt verdeutlichen.

Beispiel: Arithmetisches Mittel bei zirkulären Daten

Gesucht sei der Mittelwert der Winkel 10, 20 und 30 Grad. Bei Verwendung des sonst üblichen arithmetischen Mittels, erhielte man einen korrekten Wert von 20 Grad. Verschiebt man jedoch die Datenpunkte entlang des Kreises um 15 Grad gegen den Uhrzeigersinn auf 355, 5 und 15 Grad, so würde als arithmetisches Mittel ein Wert von 125 Grad resultieren. Dieses Ergebnis ist falsch, da das Zentrum der drei Werte bei 5 Grad liegt. Da auch bei zirkulärem Mittelwert die Schwerpunkteigenschaft erhalten bleibt, lässt sich dies z.B. leicht durch eine Zeichnung verdeutlichen.

Zur Errechnung des zirkulären Mittelwertes $\bar{\theta}$ benötigt man im ersten Schritt den Sinus- \bar{s} und den Cosinusmittelwert \bar{c} , welche wie folgt berechnet werden

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n \sin(x_i) = \frac{1}{3} \left(\sin(355) + \sin(5) + \sin(15) \right) = 0.086 \quad (6.1)$$

$$\bar{c} = \frac{1}{n} \sum_{i=1}^n \cos(x_i) = \frac{1}{3} \left(\cos(350) + \cos(10) + \cos(20) \right) = 0.986 \quad (6.2)$$

Ausgehend von diesen beiden Werten errechnet sich der zirkuläre Mittelwert durch Einsetzen der ermittelten Werte in folgende Funktion

$$\bar{\theta} = \begin{cases} \arctan\left(\frac{\bar{s}}{\bar{c}}\right) & \bar{s} > 0, \bar{c} > 0 \\ \arctan\left(\frac{\bar{s}}{\bar{c}}\right) + \pi & \bar{c} < 0 \\ \arctan\left(\frac{\bar{s}}{\bar{c}}\right) + 2\pi & \bar{s} < 0, \bar{c} > 0. \end{cases} \quad (6.3)$$

Da Sinus- und Cosinusmittelwert entsprechend Gleichung 6.1 und 6.2 positiv sind, ergibt sich daraus folgender Mittelwert

$$\bar{\theta} = \arctan\left(\frac{0.086}{0.986}\right) = \arctan(0.087) = 5. \quad (6.4)$$

Ähnlich wie in diesem Beispiel dargestellt, verhält es sich auch mit der übrigen Statistik. Sowohl die deskriptive Statistik als auch die Inferenzstatistik muss für die Anwendung bei zirkulären Daten angepasst werden.

6.1 Vorexperiment I: Hörschwellenaudiometrie

Zur Überprüfung der Hörfähigkeit absolvierte jede Versuchsperson zwei Vorexperimente. In diesem ersten Experiment wurde die absolute Hörschwelle ermittelt. Hierfür wurde mit einer adaptiven Methode die monaurale Hörschwelle für beide Ohren ermittelt. Der Vergleich der ermittelten Hörschwelle mit dem Normwert stellt sicher, dass alle getesteten Versuchspersonen über normale Hörfähigkeit verfügen.

Bevor auf das konkrete Vorexperiment eingegangen wird, soll an dieser Stelle die verwendete adaptive Methode eingeordnet werden. Die vorgestellte Methode zur Hörschwellenmessung wurde speziell für das Dissertationsprojekt konzipiert und programmiert.

6.1.1 Zur adaptiven Messmethodik

In der Psychophysik wird bei der Messung von Wahrnehmungsschwellen davon ausgegangen, dass das Antwortverhalten einer Person auf einen Reiz durch eine psychometrische Funktion beschrieben werden kann. Generell lassen sich zwei Arten von Messungen unterscheiden: die Lokation eines Punktes entlang der Reizdimension (z.B. PSE-Schwelle, 75%-Leistung) und der Anstieg der Funktion als Maß der Reliabilität des Antwortverhaltens (Strasburger, 2001). Diese Funktionen sind in der Regel monoton im Anstieg und kontinuierlich entlang der Reizdimension (z.B. logistisch, Weibull, Arc-Tangens).

Im Gegensatz zu klassischen psychophysischen Methoden mit vorher festgelegter Reizabfolge wird beim adaptiven Testen die Abfolge der Reize erst während des Experimentes entschieden. Ziel ist es, auf der Grundlage bereits erhobener Reaktionen der Versuchsperson einen möglichst schwelennahen Reiz auszuwählen. Diese Flexibilität ist der Grund für die erhöhte Effizienz adaptiver Messmethoden. Deutlich über- bzw. unterschwellige Reize sind zur Anpassung der psychometrischen Funktion zwar hilfreich, liefern aber wenig Information über die gesuchte Schwelle. Auf eine mehrfache Darbietung dieser Reize kann somit verzichtet werden.

Die hier verwendete adaptive Messmethode ist eine Modifikation der maximum likelihood adaptive staircase Methode von Harvey (1997). Sie ist eine Kombination aus der PEST-Methode (parameter estimation by sequential testing) von Taylor und Creelman (1967) und der Maximum Likelihood Prozedur

von Hall (1981). Bei PEST-Methoden wird die Richtung der Reizänderung (z.B. lauter/leiser) und die Schrittweite auf der Stimulusache (z.B. 10dB) systematisch variiert. Die Algorithmen, nach denen die Reize ausgewählt werden, unterscheiden sich stark hinsichtlich ihrer Komplexität. Diese reicht z.B. von einfachen „Up/Down“- Methoden bis hin zur Schwellenschätzung in Echtzeit. Einen Überblick über adaptive Messmethoden bietet z.B. Leek (2001).

Bei Maximum Likelihood Methoden wird zusätzlich nach einer definierten Anzahl von Trials eine psychometrische Funktion an die bislang erhobenen Daten angepasst. Nach jedem Trial erhält man somit eine neue Schwellenschätzung. Der Wendepunkt der geschätzten Funktion liefert die Schätzung für die gesuchte Schwelle. Ist die Abbruchbedingung erfüllt, entspricht die letzte Schätzung unmittelbar dem gesuchten Schwellenwert.

Bei der vorgestellten Methode ist eine wesentliche Änderung zu PEST und ML-PEST Methoden, dass die Reizdimension nicht kontinuierlich, sondern diskret, also in Reizstufen (staircase-methode), repräsentiert ist. Dies hat den Vorteil, dass für einen Reizwert mehrere Antworten vorliegen und somit eine Mittelung möglich ist.

Die konkrete Anwendung, die Parameter und die Ergebnisse der adaptiven Methodik werden im Rahmen des folgenden Experimentes zur Hörschwellenaudiometrie beschrieben.

6.1.2 Methode

Versuchspersonen

Die hier vorgestellten Daten beruhen auf audiometrischen Messungen bei 36 Versuchspersonen im Alter von 17 bis 53 Jahren ($M = 27.1$, $SD = 8.3$). Alle Personen nahmen freiwillig an der Untersuchung teil. Alter, Geschlecht, Händigkeit und Musikalität¹ aller Teilnehmer sind in Tabelle 6.2 im Ergebnisteil zusammengefasst. Die Probanden erhielten für die Teilnahme an den Vorexperimenten keine Aufwandsentschädigung.

Aufgabe der Versuchsperson

Die Versuchspersonen erhielten die Information, dass im Folgenden sehr leise Töne dargeboten werden, sobald ein Ton wahrgenommen wird, solle dies durch das Drücken der Maustaste angezeigt werden.

¹Die Operationalisierung von Musikalität wird in Experiment II (Abschnitt 6.2.1) beschrieben.

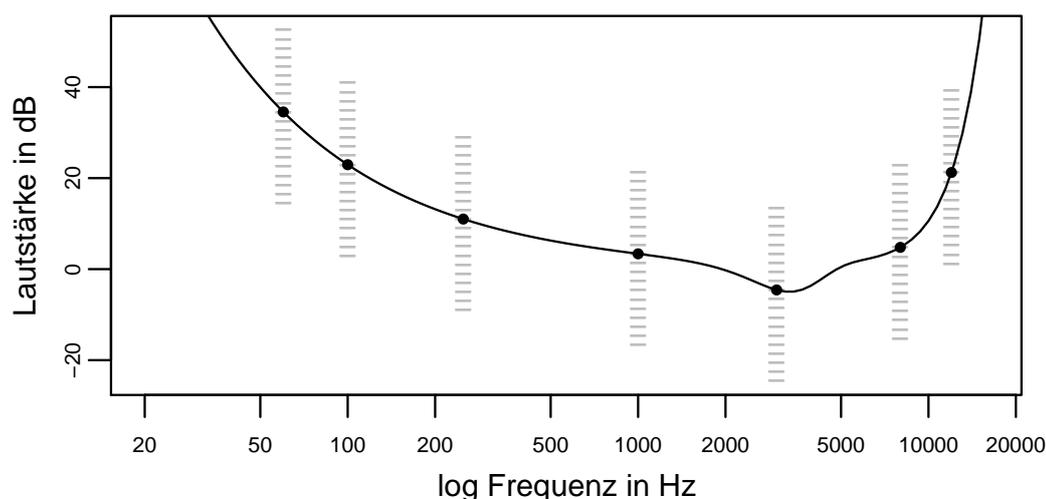


Abbildung 6.1. Die Reizstufen zentriert um die theoretische Hörkurve nach Terhardt (1998).

Stimuli und Versuchsplan

Die abhängige Variable war die Lautheit von Sinustönen, operationalisiert durch die Entdeckungsleistung von in der Lautstärke variierten Sinustönen. Es wurden Töne in sieben Frequenzen dargeboten (60 Hz, 100 Hz, 250 Hz, 1000 Hz, 3000 Hz, 8000 Hz, 12000 Hz). Jede Frequenz wurde minimal 8 mal und maximal 18 mal pro Ohr dargeboten. Die exakte Anzahl der Darbietungen war abhängig von der Erfüllung der adaptiven Abbruchbedingung. Jeder Ton hatte eine Dauer von einer Sekunde inklusive 10 ms langen linearen Ein- und Ausschwingphasen. Alle Töne wurden mit einer Samplingrate von 48 kHz und einer Samplingtiefe von 16 Bit berechnet.

Die Amplitude eines Sinustones wurde adaptiv in 20 Stufen um den Startwert variiert (siehe Abb. 6.1). Der Abstand zwischen den Stufen betrug 2 dB. Die Startwerte wurden nach der von Terhardt (1998) vorgeschlagenen theoretischen Hörkurve berechnet

$$SPL = 3,64 \cdot f^{-0,8} - 6,5 \cdot \exp^{-0,6 \cdot (f-3,3)^2} + 10^{-3} \cdot f^4. \quad (6.5)$$

Diese Funktion beschreibt die monaurale, absolute Hörschwelle für normalhörende Erwachsene. Sie gibt den Lautstärkepegel (SPL, sound pressure level) in dB an bei dem ein Sinuston mit der Frequenz f in kHz eben noch wahrnehmbar ist.

Reizauswahl und adaptive Methodik

Die adaptive Reizauswahl basierte auf einer um die aktuell geschätzte Schwelle gelegten Normalverteilung, aus der zufällig Werte gezogen wurden. Die Reizstufe (staircase), welche dem gezogenen Wert am nächsten lag, lieferte den Parameter für Reizdarbietung im nachfolgenden Trial. Der Mittelwert der Normalverteilung war die jeweils letzte (beste) Schätzung der Hörschwelle und variierte somit von Trial zu Trial. Die Standardabweichung der Normalverteilung, der Bereich aus dem in ca. 70 % der Ziehungen die Reizstufe hervorging, wurde im Vorab auf 5 dB festgelegt. Nach jedem Trial wurde unter Berücksichtigung aller bislang erhobenen Antworten (Reiz bei gegebener Lautstärke detektiert oder nicht) die Schwelle neu geschätzt. Dazu wurde einer logistische Funktion an die Antworten angepasst. Die zweiparametrische Funktion gab hier die Wahrscheinlichkeit an, einen Ton mit gegebener Lautstärke (SPL) zu detektieren. Die Funktion ist wie folgt definiert

$$p(\text{detektiert}) = f(\text{SPL}) = \frac{1}{1 + \exp \frac{\text{SPL} - \alpha}{\beta}}. \quad (6.6)$$

Als Schätzer der Schwelle wird der Parameter α gewählt. Er entspricht dem Wendepunkt der logistischen Funktion und dem Lokationspunkt entlang der Reizachse. Da die Wahrscheinlichkeit $p(\text{detektiert})$ bei einer Lautstärke von α den Wert von 0.5 annimmt, wird α auch als Punkt subjektiver Gleichheit (point of subjective equality, PSE) bezeichnet. Der Parameter β , welcher mit der Steigung der Funktion am Wendepunkt korrespondiert, wurde ebenfalls geschätzt. Er kann als Maß der Reliabilität der Detektionsleistung interpretiert werden.

Da zu Beginn des Experimentes noch keine Schwellenschätzung vorlagen, um welche die Normalverteilung gelegt werden konnte, wurde als Startwert jeder Frequenz die in Gleichung 6.5 definierte theoretische Hörkurve gewählt. Dieser Normwert stellt die beste a priori Schätzung für die Hörschwelle eines Teilnehmers dar.

Ausgehend von diesem Startwert werden weitere Reize zufällig aus einer Normalverteilung gezogen, bis die zur Schätzung notwendigen Bedingungen erfüllt sind (mind. zwei unterschiedliche Antworten auf verschiedenen Reizstufen). Um dies zu beschleunigen und mögliche Falschantworten bei schwelennahen Reizen zu vermeiden, wurden zu Beginn jedes Experimentes Reize auf Extremstufen dargeboten. Dies führte unabhängig von der normalverteilten Zufallsziehung zur Darbietung der kleinsten und größten Reizstufe, was die Wahrscheinlichkeit erhöhte, zwei verschiedene Antworten auf unterschiedlichen

Reizstufen zu erhalten.

Als Abbruchbedingung wurde eine minimale (8) und eine maximale (18) Anzahl von Trials festgelegt. Wurde die minimale Trialanzahl erreicht, so wurde die Standardabweichung der jeweils letzten 5 Schwellenschätzungen berechnet und das Experiment beendet, sobald ein vorher festgelegter Wert (hier 0.5) unterschritten wurde, da nun weitere Trials die Schätzung kaum noch verändern. Unterschreitet die Standardabweichung nicht den festgelegten Grenzwert, so endete das Experiment, nachdem die maximale Trialanzahl erreicht wurde.

Zur Anpassung der Lautstärke des im Labor verwendeten Equipments an die Werte der Normhörkurve wurden Vortests durchgeführt. Mit einem Schallpegelmessgerät² (Center SL322) wurde die Lautstärke eines Referenztones (1000 Hz) auf mehreren Stufen gemessen, um den Referenzschalldruck (Amplitudengewicht für einen Sinuston mit 0 dB) zu ermitteln. Die so erhaltene Konstante ($I_0 = 5 \cdot 10^{-5}$) wurde zur Berechnung der Amplitudengewichte aller Frequenzen verwendet. Mit Hilfe der Funktion

$$I = 10^{\left(\frac{L_{SPL}}{10}\right)} \cdot I_0 \quad (6.7)$$

wurde der Gewichtungsfaktor I für einen Sinuston mit einer Lautstärke von L_{SPL} in dB ermittelt.

Apparatur

Die Töne wurden mit einem Standard-PC (1.1 GHz Athlon) berechnet und mit einer Audigy 2 ZS Pro Soundkarte generiert. Die Darbietung der Reize erfolgte binaural über einen geschlossenen Kopfhörer (Sennheiser HD 265). Alle Experimente wurden im schallgedämmten Raum des Wahrnehmungslabors am Institut für Psychologie Halle (Saale) durchgeführt.

Zur Steuerung der Experimente wurde die von Hanke und Krüger (2005) entwickelte Software **PsychoExp** verwendet. Diese auf der Programmbibliothek Qt basierende Software umfasst zahlreiche in C++ geschriebene Klassen zur Generierung, Auswahl und Präsentation auditiver und visueller Reize. Neben einem verbesserten Schätzalgorithmus, durch Einbindung des Programms **Psignifit** von Wichmann und Hill (2001), stellt die Darbietung von Extremreiz-

²Das verwendete Messgerät gibt den gemessenen Schalldruck in Dezibel nur mit den Gewichtungen A oder C wieder. Demnach müssten alle hier präsentierten Lautstärkewerte mit dBA oder dBC gekennzeichnet sein. Da bei der verwendeten Eichfrequenz von 1000 Hz alle Gewichtungen (A,B,C,D) gleich Null sind, kann auf einen Zusatz verzichtet werden. Es wird davon ausgegangen, dass die berichteten Maße ungewichteter dB SPL Skalierung entsprechen.

stufen vor der Schätzung eine Neuerung bei den implementierten adaptiven Methoden dar.

Durchführung

Die Teilnehmer bekamen die Instruktion, dass im Folgenden Töne unterschiedlicher Frequenz auf dem linken oder auf dem rechten Ohr präsentiert werden. Aufgabe der Versuchspersonen war es, die Maustaste zu drücken, sobald sie einen Ton hörten. Als letzte Instruktion wurden die Versuchspersonen aufgefordert, sich etwa eine Minute an die Stille des Raumes zu gewöhnen, bevor sie das Experiment durch Tastendruck starteten. Während des etwa 10-minütigen Experimentes wurde kein visuelles Feedback gegeben. Nach jedem Ton folgte eine variable Pause mit einer zufällig gewählten Dauer von minimal 500 ms und maximal 2000 ms. Aufgrund dieser variablen Pause konnte auf Leer- bzw. Kontrolldurchgänge verzichtet werden. Die Darbietung der Töne für das rechte und linke Ohr sowie die Reihenfolge der Frequenzen erfolgte vollständig randomisiert. Die Auswahl der Lautstärken für jede Frequenz wurde adaptiv durch ein Maximum-Likelihood Stufenverfahren festgelegt.

6.1.3 Ergebnisse

Die Ergebnisse der 14 Parameterschätzungen für eine Versuchsperson (SK) sind in Abbildung 6.2 dargestellt. Die Platzierung der adaptiv ausgewählten Reizstufen lassen sich aus den Häufigkeitsverteilungen der jeweiligen Frequenz entnehmen. Hier zeigt sich, dass die Häufigkeit der dargebotenen Reizstufen der erwarteten Normalverteilung um den Schwellenwert folgt, welche an den Enden des Reizkontinuums durch die dargebotenen Extremwerte eingefasst wird.

Als individuelle Hörschwelle des linken und rechten Ohres wurde der geschätzte Lokationsparameter der psychometrischen Funktion α gewählt. Er gibt die Lautstärke in dB an, bei der ein Reiz in 50% der Fälle detektiert wird und somit eben noch wahrnehmbar ist. Aufgrund der Funktionsanpassung kann dieser Schwellenwert beliebig genau geschätzt werden und ist nicht mehr an diskrete Reizstufen gebunden. Ein weiterer Vorteil der Schätzung kommt dem zweiten freien Parameter β zugute. In Abbildung 6.2 zeigt sich z.B. bei den Frequenzen 60-Rechts und 3000-Rechts, dass Antwortmuster mit größerer Streuung durch eine entsprechende Steigung der psychometrischen Funktion modelliert werden. Ein fester Steigungsparameter würde bei derartigen Antwortmustern zwangsläufig zu höherer Varianz der Schwellenschätzung führen.

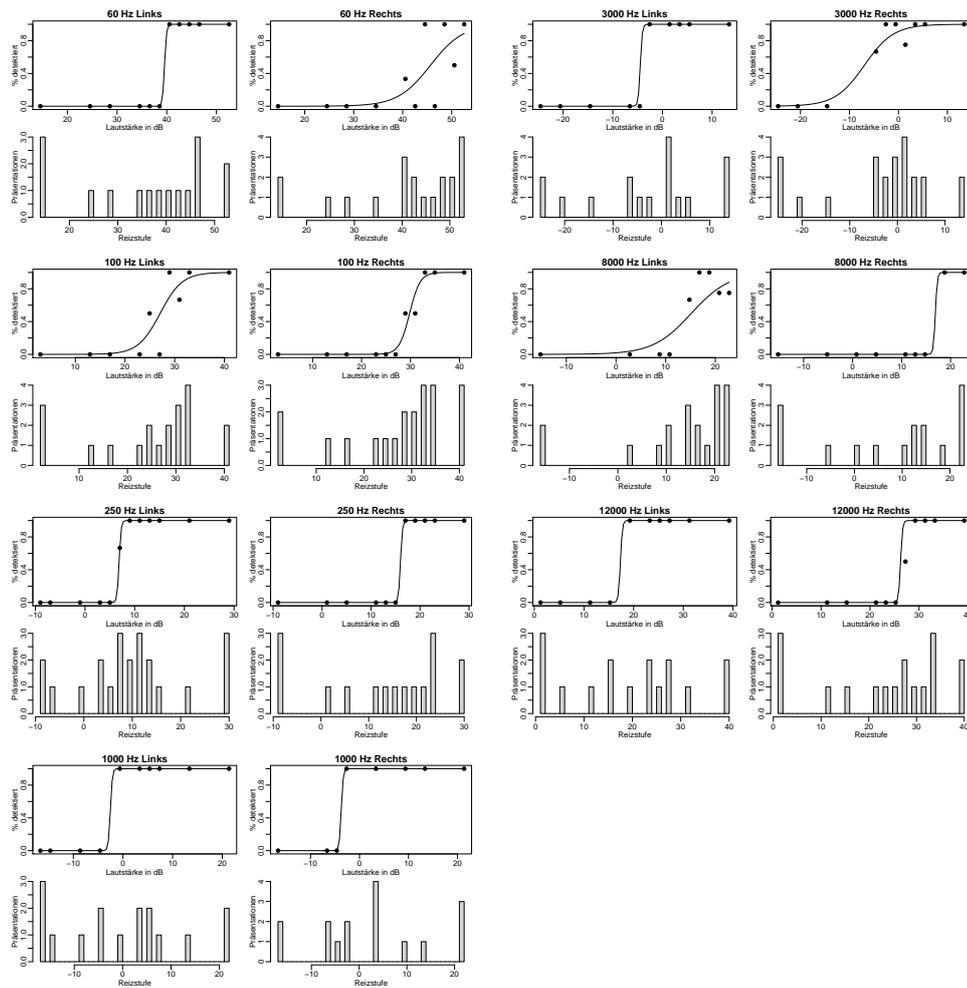


Abbildung 6.2. Darbietungshäufigkeiten der Reizstufen, Detektionsleistungen und geschätzte psychometrische Funktionen einer Versuchsperson abgetragen für alle Frequenzen und beide Ohren.

Von den insgesamt 504 ($36 \times 7 \times 2$) geschätzten Alpha-Parametern konnten 26 nicht angepasst werden, da auch bei maximaler Intensität kein Reiz detektiert wurde. In diesen Fällen wurde als Schätzung für die Hörschwelle die nächst höhere Schwellenstufe (22dB über Normalniveau) angenommen.

In Abbildung 6.3 auf der nächsten Seite sind die geschätzten Hörschwellen gemittelt über alle Teilnehmer abgetragen. Die dargestellte mittlere Hörschwelle korrespondiert mit einer durchschnittlichen Abweichung von 4.0 dB (Min = 0.2 dB bei 12kHz, Max = 8.5 dB bei 8kHz) gut mit der theoretische Hörkurve. Die eingezeichnete gepunktete Kurve markiert die binaurale—die eigentliche—Hörschwelle, d.h. die Lautstärke, bei der ein Ton vom besseren Ohr detektiert wird. Diese Schwelle liegt durchschnittlich 2.2 dB unter der mittleren monau-

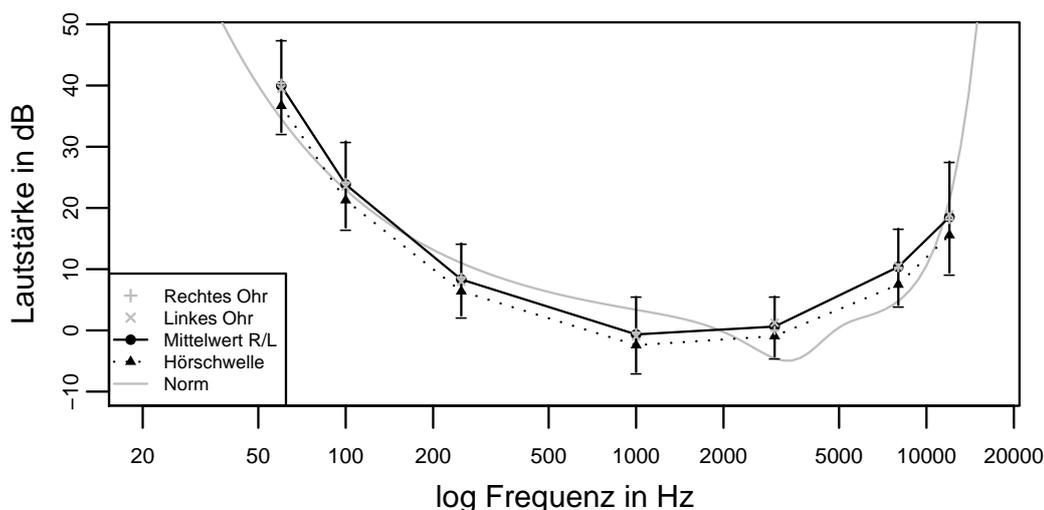


Abbildung 6.3. Die geschätzte mittlere Hörschwelle (Alphawerte) abgetragen je Frequenz für beide Ohren und deren Mittelwert. Die eingezeichneten Intervalle um den Mittelwert markieren jeweils eine Standardabweichung. Die gepunktete Linie beschreibt die mittlere Hörschwelle bei Berücksichtigung des individuellen, monauralen Minimums (kleinster individueller Alphawert). Die graue durchgezogene Kurve entspricht der Normhörkurve nach Gleichung 6.5.

ralen Hörschwelle.

Um zu überprüfen, ob sich die Schwellen des linken und des rechten Ohres unterscheiden, wurde über die geschätzten Alphawerte eine zweifaktorielle Varianzanalyse mit Messwiederholung mit den Faktoren Frequenz und Seite (rechts/links) gerechnet. Die Ergebnisse zeigen deutlich, dass sich die Schwellen für das linke und das rechte Ohr nicht substantiell unterscheiden ($F_{1,490} = 0.15, p = 0.64$). Wie bereits aufgrund der U-förmigen Hörkurve zu erwarten war, gibt es signifikante Unterschiede zwischen den Stufen des Faktors Frequenz ($F_{6,490} = 237.5, p < 0.001$). Eine Interaktion beider Faktoren konnte nicht festgestellt werden ($F_{6,490} = 0.36, p < 0.90$), was darauf hindeutet, dass die Schwellenschätzungen frequenzunabhängig sind. In Abbildung 6.4 auf der nächsten Seite ist die Detektionsleistung aller Teilnehmer für beide Ohren abgetragen. Hier lässt sich zeigen, dass die gewählten Reizstufen im sensiblen Bereich liegen, die Reizlautstärke hat über- als auch unterschwellige Ausprägungen. Gleichwohl wird deutlich, dass insgesamt 44% aller dargebotenen Töne detektiert wurden.

Alterseffekte. Die Häufigkeiten der nicht detektierten Reize je Frequenz sind in Tabelle 6.1 abgetragen. Sieben Fehldetektionen fallen allein der Versuchsperson SL zu und jeweils vier den Hörern JL, SS und JF. Alle übrigen Hörer weisen

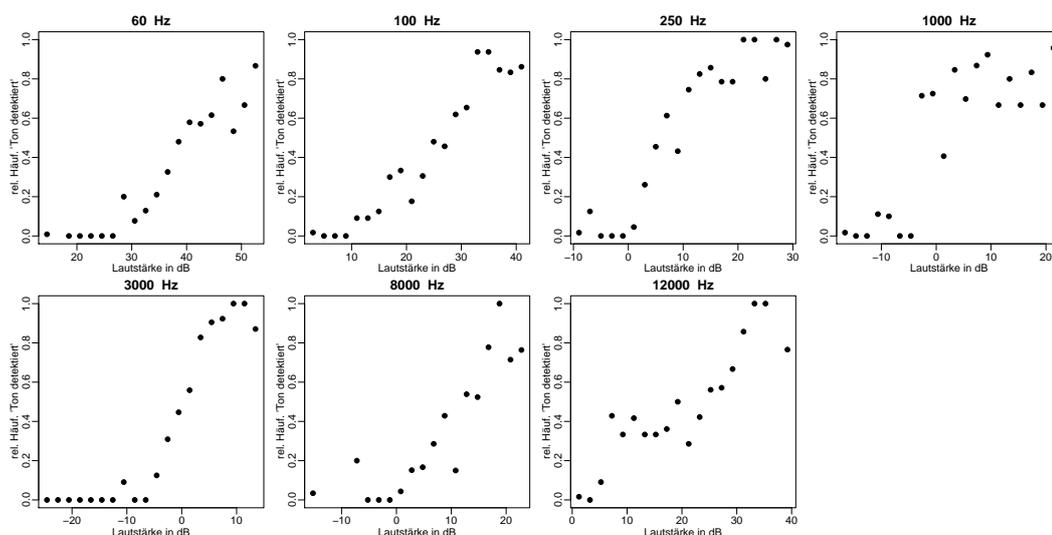


Abbildung 6.4. Die relative Häufigkeit der entdeckten Reize als Funktion der Reizlautstärke, abgetragen für alle dargebotenen Frequenzen.

Tabelle 6.1

Häufigkeit der Tests bei denen kein Reiz detektiert wurde, abgetragen für beide Seiten und alle Frequenzen.

	Frequenzen in Hz						
Seite	60	100	250	1000	3000	8000	12000
Rechts	2	1	0	0	2	5	5
Links	0	1	0	0	2	4	4
Summe	2	2	0	0	4	9	9

drei oder weniger Fehldetektionen auf. Auffällig ist, dass drei Teilnehmer mit den häufigsten Fehlern alle älter als 43 Jahre sind. Zur Überprüfung, ob dieser Unterschied auf das Alter der Hörer zurückgeführt werden kann, wurde die Korrelation für die Variablen Alter und Anzahl der Fehldetektionen berechnet und ein hoch signifikanter Zusammenhang ($r = 0.76, p < 0.001$) gefunden.

Einzelfalldiagnostik. Ab welcher Schwelle kann von einem beeinträchtigten Hörvermögen ausgegangen werden? Für eine derartige Diagnose können unterschiedliche Kriterien herangezogen werden. Terhardt (1998) weist auf abweichendes Hörverhalten ab einer Differenz von 10 dB hin, während Moore (2003) einen Bereich von 20 dB über und unter der Normschwelle als normal beschreibt.

Das ICD-10 (WHO, 1992) klassifiziert unter H90-95 (Hörverlust durch Schalleitungs- oder Schallempfindungsstörung) neben verschiedenen Schädi-

gungen des Mittel- und Innenohrs auch unterschiedliche Stufen des Hörverlustes. Demnach liegt eine leichtgradige Schwerhörigkeit dann vor, wenn der Hörverlust im Hauptsprachbereich einen Wert von 25 dB auf dem besseren Ohr überschreitet. Mittelgradige und hochgradige Schwerhörigkeit kann bei Überschreitungen von 40 dB bzw. 85 dB diagnostiziert werden. Von ähnlichen Kategorien geht auch Goodman (1965) aus. Er schlägt als Maß für den Hörverlust (PTA, pure tone average) die mittlere Schwelle bei 500 Hz, 1000 Hz und 2000 Hz vor.

Geht man von diesen Kriterien aus und operationalisiert den Hörverlust im Hauptsprachbereich durch das arithmetische Mittel der Normabweichungen bei den Frequenzen 250 Hz, 1000 Hz und 3000 Hz, so erreicht keiner der getesteten Hörer dieses Kriterium.

An dieser Stelle könnte argumentiert werden, dass es bei der gewählten Lautstärkespanne (± 20 dB) gar nicht möglich sei, den kritischen Wert von 25 dB zu überschreiten. Hierbei sollte jedoch beachtet werden, dass die durchgeführten Hörschwellenmessungen, gerade bei geringer Anzahl von Trials, stark variieren. Dieser *Messfehler* sollte bei einer abschließenden Diagnose berücksichtigt werden.

Zur Bestimmung des Standardmessfehlers S wird die Reliabilität des vorliegenden Tests und dessen Standardabweichung benötigt

$$S = SD \sqrt{1 - r_{RL}}. \quad (6.8)$$

Da im hier berichteten Vorexperiment keine Retests durchgeführt wurden, wird als Maß für die Reliabilität die Korrelation der Hörschwellen des linken und des rechten Ohrs berechnet. Bei dieser Form der Paralleltestreliabilität kann zwar von einzelnen systematischen Rechts-Links-Abweichungen ausgegangen werden, über alle Versuchspersonen hinweg betrachtet, sind diese jedoch vernachlässigbar (siehe obige ANOVA).

Bei einer Reliabilität von $r_{RL} = 0.76$ und einer Standardabweichung von $SD = 5.04$ erhält man nach Formel 6.8 einen Standardmessfehler von $S = 2.47$. Um Wahrscheinlichkeitsaussagen treffen zu können, wird ein Vertrauensbereich CI um die erhobene Schwellenschätzung α mit folgender Funktion berechnet,

$$CI = \alpha \pm Z_{1-\alpha/2} S. \quad (6.9)$$

Bei einer angenommenen Irrtumswahrscheinlichkeit von 0.05 ($Z_{1-\alpha/2} = 1.96$) erhält man einen Konfidenzbereich von $\alpha \pm 4.9$ dB. Es kann somit davon ausgegangen werden, dass in 95% der Fälle die wahre Hörschwelle in einem Bereich von 5 dB um dem ermittelten Schwellenwert liegt. Auch unter Einbezug des

Messfehlers erreicht kein Versuchsteilnehmer die kritische Grenze von 25 dB im mittleren Hauptsprachbereich.

Um abweichende Hörfähigkeit zu diagnostizieren, sollte neben der ermittelten Hörschwelle auch das Alter der Hörer berücksichtigt werden. Auch sollten gerade bei Experimenten in denen dichotische Reize präsentiert werden (siehe 6.6) Rechts-Links-Differenzen stärker gewichtet werden. Die Ergebnisse der Schwellenschätzungen, die mittleren Abweichungen vom Normwert sowie die Rechts-Links-Differenz getrennt für jeden Teilnehmer sind in Tabelle 6.2 aufgelistet.

Die mittlere Abweichung aller Hörer von der Norm nach Terhardt (1998) beträgt 2.38 dB (SD = 4.9, Min = -5.3, Max = 15.3). Die durchschnittliche Abweichung im Hauptsprachbereich nach ICD-10 (250 Hz, 1000 Hz, 3000 Hz) liegt bei 0.04 dB (SD = 4.5, Min = -6.5, Max = 12.3). Die mittlere Differenz zwischen der Hörschwelle des rechten und linken Ohres beträgt 2.8 dB (SD = 2.5, Min = 0.3, Max = 12.2).

6.1.4 Diskussion

In dem vorliegenden Experiment wurde die monaurale Hörschwelle mit einer erweiterten adaptiven Stufenmethode nach Harvey (1997) erhoben. Alle Probanden erfüllten die im ICD-10 definierten Kriterien für eine normgerechte Hörfähigkeit. Zwar konnten Alterseffekte nachgewiesen werden, welche aufgrund der Frequenzselektivität vernachlässigbar erscheinen, da diese nicht bei Frequenzen im Hauptsprachbereich liegen.

Weiterhin konnte gezeigt werden, dass mit wenigen Trials (≤ 18) die Hörschwelle reliabel an sieben Frequenzpunkten erhoben werden kann, was die Methode als effizientes Screeningverfahren auszeichnet. Dass insgesamt 96% aller Tests zu Schwellenschätzungen führten, spricht für die Wahl der theoretischen Hörkurve als Startlautstärke. Die 26 Subtests, bei denen keine Funktionsanpassung möglich war, da die Hörschwelle der Teilnehmer über der höchsten Reizausprägung lag, können durch eine zukünftige Erweiterung der Reizspanne optimiert werden. Hierbei sollten auch die unterschiedlichen Varianzen auf den sieben Frequenzen berücksichtigt werden. Eine Anpassung der Stufen- bzw. Schrittweite (hier 2dB) in Abhängigkeit zur Frequenz (bzw. theoretischen Absolutschwelle) könnte so zu größeren Schritten bei den Höhen und Tiefen und zu kleineren Schritten im mittleren Frequenzbereich führen. Dies würde auch der, durch den Alterseffekt hervorgerufenen, höheren Streuung bei den hohen Frequenzen Rechnung tragen.

Tabelle 6.2

Initialien, Alter, Geschlecht, Händigkeit, Musikalität, Schätzungsfehler, Schwelle nach ICD-10, mittlere Abweichung von der Norm und Rechts-Links-Differenz aller Teilnehmer an den beiden Vorexperimenten.

Nr	Vp	Alter	Geschl.	Hand	Musik	Fehler	ICD10	Normdiff.	RLDiff
1	am	25	m	r	-	0	-1.81	0.25	0.38
2	ap	25	m	r	-	0	-1.32	-0.85	2.88
3	aw	26	w	r	-	0	4.26	3.65	4.69
4	ch	23	m	r	x	0	0.54	1.30	2.58
5	dl	21	m	r	x	0	-3.93	-3.05	0.62
6	hd	27	m	r	x	0	-3.42	-2.33	1.99
7	in	22	w	r	x	0	-2.45	-1.02	1.88
8	jf	49	w	r	x	4	9.66	15.39	6.32
9	jh	18	w	r	-	0	-1.99	-1.50	2.01
10	jl	53	m	r	x	4	-4.65	2.21	0.95
11	jm	22	w	r	-	0	4.78	5.27	0.53
12	js	26	w	r	x	0	2.21	2.99	2.96
13	ka	23	w	r	x	0	-2.47	-3.30	0.71
14	ma	29	m	r	-	1	-1.13	3.08	0.33
15	md	32	m	r	x	0	-5.48	-1.34	1.29
16	mh	25	m	r	-	0	-6.54	-3.21	9.12
17	mi	29	m	r	x	0	-1.35	3.56	4.37
18	mm	19	w	r	x	0	-1.54	2.15	0.59
19	mo	40	m	r	-	2	-2.16	4.35	0.71
20	mr	20	w	r	-	0	3.86	4.69	3.90
21	nm	17	w	r	-	0	-5.50	-5.37	2.67
22	rz	29	m	r	x	1	3.19	5.61	5.30
23	sb	21	w	r	-	0	3.17	6.43	1.48
24	sk	30	m	r	x	0	-2.40	2.59	3.57
25	sl	43	m	r	-	7	10.80	14.49	4.25
26	sn	24	w	r	x	0	-4.91	-2.08	3.90
27	so	21	w	r	x	0	-4.30	-1.99	2.87
28	ss	31	m	r	-	4	12.33	10.78	2.88
29	st	18	w	r	x	0	0.94	2.30	0.43
30	tb	22	m	r	x	0	-1.78	-0.28	2.97
31	tg	25	m	l	x	0	-1.69	-0.57	1.42
32	th	32	m	r	x	1	3.64	5.89	12.21
33	tm	25	w	r	x	0	0.89	1.37	3.73
34	tr	20	w	r	x	2	-0.14	10.35	1.01
35	ts	36	m	r	-	1	3.79	7.03	0.71
36	wo	26	m	r	x	0	-1.64	-3.12	1.08

Als weitere Verbesserungen der adaptiven Methode sollte es möglich sein, die Varianz der Normalverteilung, welche die Wahrscheinlichkeit der Ziehung der nächste Reizstufe festlegt, mit steigender Trialzahl zu schmälern. Dies würde die Reizsektion in Schwellennähe und somit die Iteminformation schrittweise erhöhen. Als weitere Trial-Einsparungen könnten die eingeführten Extremwertpräsentationen von drei auf eine Präsentation verringert werden.

Um in Zukunft genauere Messungen durchführen zu können, wäre es auch notwendig die absoluten Lautstärken zu messen, um die Apparatur (Soundkarte, Verstärker, Kopfhörer) unter Berücksichtigung der Frequenzselektivität zu eichen.

6.2 Vorexperiment II: Tonhöhendiskrimination

Im folgenden zweiten Vorexperiment wurde die Diskriminationsleistung für die Tonhöhe von Sinustönen untersucht. Hintergrund ist, wie auch bei dem ersten Vorexperiment, die Sicherstellung der Hörfähigkeit, speziell hier die Fähigkeit zur Differenzierung von Tonhöhen. Nicht alle Personen sind gleich gut in der Lage, Tonhöhen zu unterscheiden, was verschiedene Autoren veranlasste ähnliche Voruntersuchungen durchzuführen. So berichtet z.B. Deutsch (1991), dass ihre Probanden jeweils nicht mehr als 6 Fehler bei 48 Halboktav-Paarvergleichen (auf-/absteigend) von Sinustönen aufwiesen. Dies entsprach einem Anteil von 12.5 % Fehlern. Es wird nicht berichtet, wie viele Personen das gesetzte Kriterium nicht erreichten und folglich vom Experiment ausgeschlossen worden. In eigenen Voruntersuchungen ermittelten Giangrande et al. (2003), dass 40% der Teilnehmer nicht in der Lage waren, 85% der dargebotenen Sinustonpaare korrekt als auf- oder absteigend zu klassifizieren.

Warum Sinustöne? Ein wesentlicher Grund liegt in der Einfachheit der Reize. Die Tonhöhe von Sinustönen ist primär durch die Frequenz definiert. Anders als bei komplexen Tönen haben Sinustöne nur eine Tonhöhe, welche im Paarvergleich mit der eines zweiten Sinustones verglichen werden muss. Diese Eindeutigkeit ist der Vorteil gegenüber anderen komplexeren Reizen, welche mehrere Tonhöhen hervorrufen, was zu Ambiguitäten bei der Beurteilung der Tonhöhe führen kann. Sinustöne sind eindeutiger als komplexe Töne, aber zugleich schwerer zu beurteilen als harmonisch komplexe Töne. Würde dieser Test mit natürlichen Klängen, z.B. am Klavier durchgeführt werden, so würde die Vergleichsaufgabe vermutlich leichter werden. Der Test würde jedoch auch an Trennschärfe verlieren (vgl. Abschnitt 3.1.3 zur Eindeutigkeit der Tonhöhe).

6.2.1 Methode

Versuchspersonen und Apparatur

An dem Experiment nahmen alle 36 Hörer teil, welche auch das erste Vorexperiment absolvierten. Daher sei in Bezug auf die Kennwerte zu den Personen (Alter, Geschlecht, Händigkeit, Musikalität) sowie der verwendeten Apparatur auf Tabelle 6.2 sowie auf Abschnitt 6.1 verwiesen.

Die Musikalität der Teilnehmer wurde zu Beginn der Vorexperimente erfasst. Ein Hörer wurde als musikalisch eingestuft, wenn er angab, mindestens

ein Jahr ein Instrument zu spielen. Da es hier primär um die Fähigkeit geht, Tonhöhe zu unterscheiden wurden nur Melodieinstrumente gewertet (z.B. wurde Schlagzeug nicht als Melodieinstrument gezählt). Gesang wurde ebenfalls als Melodieinstrument gewertet. Bei Mehrfachangaben (z.B. Gesang, Gitarre) wurde der Proband als musikalisch eingestuft, wenn mindestens ein Melodieinstrument bzw. Gesang angegeben wurde.

Aufgabe der Versuchsperson

Die Aufgabe der Versuchspersonen war, zu entscheiden, ob der zweite Ton eines vorher dargebotenen Tonpaares höher oder tiefer war als der erste Ton. Das Urteil der Versuchsperson (aufsteigend oder absteigend) wurde durch Tastendruck erhoben, wobei die Tastaturbelegung kongruent mit den zu beurteilenden Empfindungen war, d.h. bei aufsteigend empfundenem Perzept sollte die Pfeiltaste-Hoch gedrückt werden und bei absteigendem Perzept die Pfeiltaste-Runter.

Stimuli und Versuchsplan

Die Töne eines Paares hatten jeweils eine Dauer von 400 ms. Jeder Ton enthielt lineare Ein- und Ausschwingphasen im Bereich der ersten und letzten 10 ms.

Die Frequenzen der Tonpaare wurden wie folgt bestimmt. Die Frequenz des ersten Tones wurde zufällig aus einem Bereich von 500 Hz bis 2500 Hz gezogen. Ausgehend von dieser wurde die Frequenz des zweiten Tones ebenfalls zufällig aus dem Intervall von ± 11 Halbtonschritten bestimmt. Die Bedingung dabei war, dass auf- und absteigende Tonpaare gleich häufig auftraten und alle Intervalle (1 bis 11 Halbtonschritte) mindestens einmal enthalten waren.

Jeder Teilnehmer beurteilte 100 Reizpaare. Die Reihenfolge der dargebotenen Tonpaare war für alle Teilnehmer gleich. Um die Reliabilität des Verfahrens zu überprüfen, wurde bei einigen Probanden eine zweite Testung durchgeführt. Für diese Retests wurde die Reihenfolge neu permutiert, diese war jedoch wiederum für alle Teilnehmer gleich.

Durchführung

Die Reizpaare wurden den Teilnehmern in einem schallgedämmten Raum binaural über geschlossene Kopfhörer mit 65 dBA präsentiert. Die Reizgenerierung und -präsentation sowie das Erfassen der Antworten erfolgte am Computer. Zu Beginn des knapp 10 minütigen Experimentes erhielten die Teilnehmer die Instruktionen (siehe Anhang) auf dem Bildschirm. Bei dem 2AFC-Paradigma

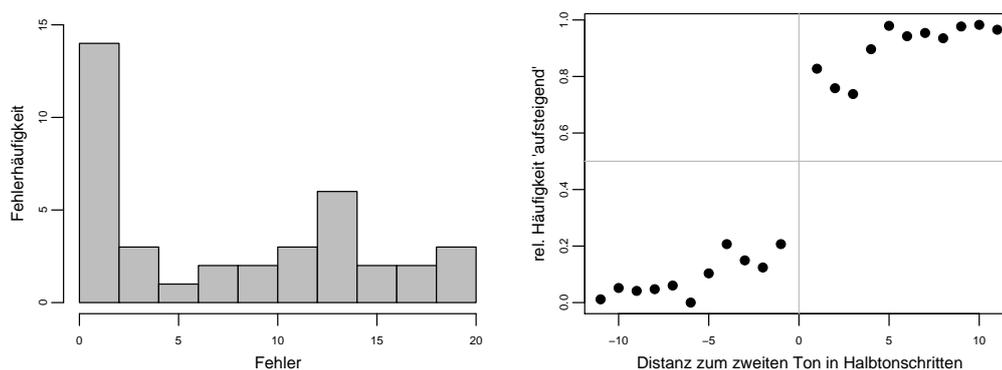


Abbildung 6.5. Die relative Häufigkeit der Antwort 'Tonpaar ist aufsteigend' abgetragen für alle dargebotenen Intervalle in Halbtönen.

wurden die Teilnehmer nach der Präsentation der Reize aufgefordert, die relative Tonhöhenänderung (auf- oder absteigend) durch Tastendruck (Pfeil-Hoch-Taste oder Pfeil-Runter-Taste) zu signalisieren.

6.2.2 Ergebnisse

Im Gegensatz zu oktav-komplexen Tönen lässt sich bei einfachen Sinustönen anhand der Frequenz (f) eindeutig feststellen, ob ein Reizpaar als auf- oder absteigend bewertet werden sollte. Ist $f_1 < f_2$ so liegt eine aufsteigende Tonfolge vor, ist $f_1 > f_2$ sollte das Paar mit absteigend bewertet werden. Alle anderen Antworten wurden als Fehler gewertet.

Insgesamt wurden im Mittel 8 Tonpaare (8%) falsch beurteilt (SD = 6.86, Min = 0, Max = 20). Davon entfallen 4.2% Fehler auf aufsteigende Paare, welche fälschlicherweise als absteigend beurteilt wurden und 3.8% Fehler auf absteigende Paare, welche als aufsteigend wahrgenommen wurden. Aufsteigende Reizpaare sind demnach schwerer zu beurteilen als absteigende. Bei insgesamt 3600 Paarvergleichen erweist sich der Unterschied zwischen auf- und absteigenden Paaren mit 162 zu 140 Fehlern als statistisch nicht bedeutsam ($\chi^2 = 1.59$, $df = 1$, $p = 0.21$). Die Fehlerhäufigkeiten sowie die relative Fehlerhäufigkeit je Intervall sind in Abbildung 6.5 dargestellt. Wie dem Histogramm entnommen werden kann, verteilen sich die Fehler mit Ausnahme der Häufungen bei 0-1 und 12-13 Fehlern relativ gleichmäßig in einem Bereich von 0 bis 20. In Abbildung 6.5 rechts wird deutlich, dass die Fehlerhäufigkeit bei auf- als auch absteigenden Reizpaaren mit kleiner werdenden Intervallen ansteigt.

Einfluss musikalischer Vorbildung. Geht man davon aus, dass „ein gutes Gehör“ erlernbar ist, wie es z.B. bei der musikalischen Gehörbildung vorausge-

setzt wird, dann sollten sich auch bei der Tonhöhendiskrimination Unterschiede zwischen Musikern und Nicht-Musikern zeigen lassen.

In der Voruntersuchung wurden insgesamt 22 Probanden als musikalisch eingestuft, d.h. diese gaben an mindestens ein Jahr ein Melodieinstrument gespielt zu haben. Die am häufigsten genannten Instrumente waren Tasteninstrumente (Klavier, Keyboard, Orgel, Akkordeon) gefolgt von Gitarre, Gesang, Cello, Geige und Klarinette. Eine Person, welche angab Schlagzeug zu spielen, wurde im o.g. Sinne als nicht musikalisch eingestuft.

Insgesamt wurden durchschnittlich 4.4 Fehler von den Musikern ($N = 22$, $SD = 5.3$) und 12.4 Fehler von den Nicht-Musikern ($N = 14$, $SD = 5.9$) verursacht. Um den Einfluss von Musikalität zu prüfen, wurde die punktbiserial Korrelation von Musikalität und der Anzahl der Fehler bei der Tonhöhendiskrimination berechnet. Der negative, hoch signifikante Zusammenhang ($r = -0.58$, $p < 0.001$) bestätigt die Hypothese, dass die Diskrimination von Tonhöhe einem Lerneffekt unterliegt.

Innerhalb der Musiker lässt sich eine weitere Gruppe in Abhängigkeit der erlernten Instrumente ausmachen. So zeigt sich ein Vorteil für Musiker mit Instrumenten, bei denen ein Ton erst gesucht werden muss (z.B. Geige, Cello, Gesang) gegenüber denen mit Instrumenten mit vordefinierten Intervallgrenzen (z.B. Klavier, Gitarre). Die Stichprobe derer, die bundlose Instrumente beherrschen, ist jedoch mit $N = 3$ zu gering, um generalisierbare Aussagen zu treffen.

Einzelfalldiagnostik. Um entscheiden zu können, wie genau ein Hörer Tonhöhe differenzieren kann, soll wiederum der Einfluss des Messfehlers berücksichtigt werden. Zur Bestimmung der Reliabilität wurde der Test vier Teilnehmern (CH, SK, SS, TH) wiederholt vorgelegt und die Fehlersummen des ersten und des zweiten Tests korreliert. Mit der so ermittelten Retestreliabilität von $r_{tt} = 0.98$ wurde nach der Gleichung 6.8 ein Standardmessfehler von 0.99 errechnet. Nach Formel 6.9 erhält man bei einer Irrtumswahrscheinlichkeit von 0.05 einen Vertrauensbereich um einen Messwert von ± 2 Fehlerpunkten. Hierbei sollte bedacht werden, dass die Verteilung der Fehler gerade bei den Musikern aufgrund vieler korrekter Antworten linksschief ist und die Anwendung des Vertrauensbereiches sich somit nur bei mittleren Fehlerzahlen empfiehlt.

6.2.3 Diskussion

Im vorliegenden Experiment wurde die Fähigkeit zu Tonhöhendiskrimination anhand eines Paarvergleiches von Sinustönen untersucht. Diese Fähigkeit ist

bereits bei sehr kleinen Intervallen vorhanden. So konnten Wier, Jesteadt und Green (1977) zeigen, dass der kleinste, eben noch wahrnehmbare Frequenzunterschied (difference limen for frequency, DLF) bei 0.2% bis 0.3% liegt. Dieser gibt an, bei welcher Differenz ein Tonpaar mit einer Wahrscheinlichkeit von 0.75 korrekt als auf- bzw. absteigend erkannt wird. Das kleinste im vorliegenden Experiment dargebotene Intervall von einem Halbtonschritt entspricht einer Differenz von 6%. Bei einem 1000 Hz Sinuston beträgt das DLF demnach 2-3 Hz, zum Vergleich, ein Halbtonschritt entspricht ca. 60 Hz. Somit könnte erwartet werden, dass Hörer in der Lage sind, Differenzen im Bereich der 20fachen DLF mit hoher Sicherheit unterscheiden zu können. Dennoch scheint die gestellte Aufgabe allgemein schwerer zu sein als erwartet. Aus den Ergebnissen wird deutlich, dass selbst Intervalle von 3 Halbtonschritten nur mit einer relativen Häufigkeit von rund 0.80 richtig klassifiziert werden.

Hier stellt sich die Frage nach einem geeigneten Kriterium für diesen Test. Ab welcher Fehlerzahl wird einer Person noch die Fähigkeit zur Diskrimination zugesprochen? Diese Frage soll auf zwei Arten beantwortet werden, empirisch und statistisch.

Vergleicht man die Ergebnisse mit den eingangs erwähnten Fehlerquoten von z.B. Giangrande et al. (2003), so weisen die hier getesteten Hörer bessere Ergebnisse bei der Beurteilung von Sinustonpaaren auf. Selbst unter Berücksichtigung des Messfehlers kann davon ausgegangen werden, dass die Fähigkeit zur Tonhöhendiskrimination im Vergleich zu einschlägigen Studien bei allen getesteten Probanden vergleichbar bis besser ausgeprägt ist.

Zur statistischen Beantwortung der Frage nach einem geeigneten Kriterium könnte die Nullhypothese „Person rät“ ($p = 0.5$) gegen die Alternativhypothese „Person kann sicher zwischen auf- und absteigend unterscheiden“ ($p \neq 0.5$) getestet werden. Bei der gegebenen Binomialverteilung (auf-/absteigend), einer Anzahl von $N=100$ Versuchen und einer Irrtumswahrscheinlichkeit von 0.05 erhält man ein Vertrauensintervall von ± 10 Fehlerpunkten. Somit kann mit einer Sicherheit von 95% davon ausgegangen werden, dass Personen mit weniger als 40 Fehlern nicht zufällig antworten. Auch unter Berücksichtigung des Messfehlers lagen alle getesteten Personen deutlich unterhalb dieses Kriteriums.

Auf ein Problem bei dem Vergleich von Sinustönen weist Henning (1966) hin. So sind Paarvergleiche von besonders hohen und tiefen Frequenzen mit der Lautheitsänderung konfundiert. Teilnehmer könnten demnach durch den bloßen Vergleich der Lautheit beider Töne zur richtigen Antwort gelangen. Für den hier gewählten Frequenzbereich (250 Hz - 5000 Hz) scheint dieses Problem marginal zu sein, sollte jedoch bei Folgeexperimenten bedacht werden.

Weiterhin konnte mit diesem Vorexperiment die naheliegende Hypothese bestätigt werden, dass das Spielen eines Melodieinstrumentes einen Einfluss auf die Diskriminationsleistung bei Sinustönen hat. Es zeigen sich große individuelle Unterschiede in der Anzahl der Fehler, die zum Teil auf die unterschiedliche musikalische Vorbildung zurückgeführt werden können. Auch legt der signifikante Unterschied zwischen den Musikern und den Nicht-Musikern nahe, für beide Gruppen ein separates Vertrauensintervall zu berechnen. Dies könnte zu einer erhöhten Genauigkeit der Messung führen. Im vorliegenden Fall wurde aufgrund der geringen Fallzahlen davon abgesehen. Es kann davon ausgegangen werden, dass das instrumentelle Üben zu einer feineren Repräsentation von Tonhöhe führt, welche exaktere Operationen (z.B. Vergleich, Abruf) erlaubt. Gerade bei den komplexen und mehrdeutigen Reizen wie den Shepardtönen ist ein „genauereres Hören“ die Grundlage für stabile Urteile. Die Diskriminationsfähigkeit ist daher eine wesentliche Voraussetzung für die Interpretation der Ergebnisse der folgenden Experimente.

6.3 Zur Stabilität der Urteile bei der Tritonusaufgabe

Im folgenden Abschnitt wird eine Untersuchung vorgestellt, bei der die Probanden die in Abschnitt 5.1.4 eingeführte „Tritonusaufgabe“ bearbeiten. Dabei werden im Paarvergleich oktav-komplexe Töne im Intervall des Tritonus dargeboten.

6.3.1 Fragestellung

Der von Deutsch und Mitarbeitern postulierten Theorie (siehe Abschnitt 5.1.4) folgend, wird von Personen einer Sprachgemeinschaft meist eine Chromaklasse in Bezug auf die Tonhöhe präferiert (Deutsch, 1986, 1987). Das vielfältig untersuchte Tritonusparadox beschreibt die Stabilität bei Tonpaaren, bei denen gemäß der Shepardschen Theorie uneindeutige oder eher zufällige Urteile zu erwarten wären (Deutsch, 1988, 1991, 2007; Deutsch et al., 1990; Deutsch & Dolson, 2004). Mit dem folgenden Experiment soll überprüft werden, ob sich die Stabilität der individuellen Urteile replizieren lässt. Im Nachgang kann erwartet werden, dass sich eine präferierte Tonklasse finden lässt, da alle Hörer der deutschen Sprachgemeinschaft angehören.

Wie bereits dargestellt wurde, unterliegen Paarvergleiche von oktav-komplexen Tönen starken Reihenfolgeeffekten. So konnte gezeigt werden, dass z.B. ein vorangestelltes eindeutiges Tonpaar die Beurteilung eines Tritonusintervalls beeinflusst (Dawe et al., 1998; Giangrande et al., 2003). In den genannten Untersuchungen werden meist ausschließlich Tritonuspaare dargeboten und davon ausgegangen, dass auf diese Weise der Kontext- oder Reihenfolgeeffekt aufgrund der Ambivalenz dieser Paare minimiert wird. Die Annahme, dass diese Intervalle weniger eindeutig sind, widerspricht bei genauer Betrachtung den Befunden zum Tritonusparadox, da die Urteile selbst innerhalb einer breiten Gemeinschaft als stabil angesehen werden. Dies führt zu der Frage: Gibt es Reihenfolgeeffekte bei der Tritonusaufgabe? Hier gilt es zu überprüfen, ob ein vorangestelltes Tritonus-Tonpaar die Antwort des darauf folgenden beeinflusst.

6.3.2 Methode

Versuchspersonen

An dem etwa viertelstündigen Experiment nahmen 21 Versuchspersonen freiwillig teil. Die 9 weiblichen und 12 männlichen Hörer im Alter von durchschnittlich 28 Jahren ($SD=9.6$) waren vorwiegend Studenten und Mitarbeiter des Institutes für Psychologie Halle. Vier Teilnehmer, ebenfalls Studenten der Martin-Luther-Universität, kamen aus anderen Fachrichtungen. Die 14 Studenten der Psychologie erhielten als Aufwandsentschädigung für die Teilnahme am Experiment und den beiden Vortests eine halbe Versuchspersonenstunde gutgeschrieben. Alle übrigen Teilnehmer bekamen keine Aufwandsentschädigung.

Aufgabe der Versuchsperson und Versuchsplan

Aufgabe der Versuchsperson war es, Tonpaare im Hinblick auf die wahrgenommene Tonhöhenänderung als auf- oder absteigend zu beurteilen. Die Antworten der Teilnehmer wurden durch das Drücken einer von zwei Tasten registriert.

Stimuli

Das verwendete Tonklassenset bestand aus den üblichen 12 Tonklassen in Halbtonschritten (C, Cis, D, . . . ,A, Ais, B), welche jeweils aus sechs Teiltönen zusammengesetzt waren und von einer Hüllkurve, fixiert bei C ($f_{min} = 32.7$ Hz), überspannt wurde. Die Berechnung der Reiz-Parameter erfolgte nach den in Kapitel 5 dargestellten Formeln. Alle 12 Tritonuspaare (C-Fis, Cis-G, D-Gis, . . . ,A-Dis, Ais-E, B-F) wurden 10 mal präsentiert. Jeder oktav-komplexe Ton wurde mit einer Dauer von 400 ms, inklusive einer 10 ms Ein- und Ausschwingphase, synthetisiert. Die Reihenfolge der Darbietung wurde für jeden Hörer erneut vollständig randomisiert. Die Bedingung dabei war, dass kein Paar zweimal hintereinander dargeboten wurde. Die Stimuli wurden mit einer Lautstärke von 65 dBA über geschlossene Kopfhörer präsentiert.

Apparatur

Die Durchführung des Experimentes erfolgte unter den in Abschnitt 6.1 auf Seite 73 dargestellten technischen und räumlichen Bedingungen.

Tabelle 6.3

Die absoluten Häufigkeiten der Tonhöhenurteile „aufsteigend vs. absteigend“, abgetragen für den zweiten Ton eines Paares.

Urteil	zweiter Ton eines Paares											
	C	Cis	D	Dis	E	F	Fis	G	Gis	A	Ais	B
absteigend	110	98	105	102	101	120	116	133	131	125	128	119
aufsteigend	100	112	105	108	109	90	94	77	79	85	82	91

Durchführung

Vor Beginn eines Experimentes wurde, so nicht bereits geschehen, die Hörfähigkeit jeder Versuchsperson mit zwei audiometrischen Vorexperimenten (siehe Abschnitt 6.1 auf Seite 73) erfasst. Nach der Präsentation der Instruktion auf dem Bildschirm sowie kurzer mündlicher Instruktion durch den Versuchsleiter wurde das Experiment gestartet.

Ein Trial setzte sich zusammen aus dem Reizpaar (400ms Ton1, 20ms Pause, 400ms Ton2) gefolgt von der Antwort der Versuchsperson und einer Pause von 2 Sekunden zum nachfolgenden Trial. Zur Beurteilung eines Reizpaares wurde kein Zeitlimit festgelegt. Während der insgesamt 120 Trials wurde keine Pause angeboten.

6.3.3 Ergebnisse

Die Ergebnisse des Paarvergleiches sind in Tabelle 6.3 dargestellt. Von den insgesamt 2520 (10 Vergleiche \times 12 Tonpaare \times 21 Personen) dargebotenen Intervallen wurden 1132 (44.9%) als aufsteigend und 1388 (55.0%) als absteigend bewertet. Dieser Antwortbias zugunsten der absteigend-Urteile erwies sich als statistisch bedeutsam ($\chi^2 = 25.80$, $df = 1$, $p < 0.001$).

Stabilität der Urteile. In Abb. 6.7 sind die Antwortmuster aller Teilnehmer als relative Häufigkeit getrennt nach Chromaklassen abgetragen. Es zeigt sich, dass die Antworten sowohl in der Sicherheit der Urteile als auch hinsichtlich der Tonhöhenpräferenz stark variieren. Von den insgesamt 21 Hörern bewerteten nur acht Teilnehmer (DL, JL, MA, MD, SK, SO, ST, TG) mindestens eine Tonklasse immer als aufsteigend und mindestens eine immer als absteigend. Dies bedeutet im Umkehrschluss, dass knapp zwei Drittel der Probanden keines der 12 Reizpaare sicher klassifizieren konnten. Auch konnte keine Person alle Tonpaare reliabel klassifizieren.

Als ein Maß für die Sicherheit der Urteile wurde die mittlere Varianz der

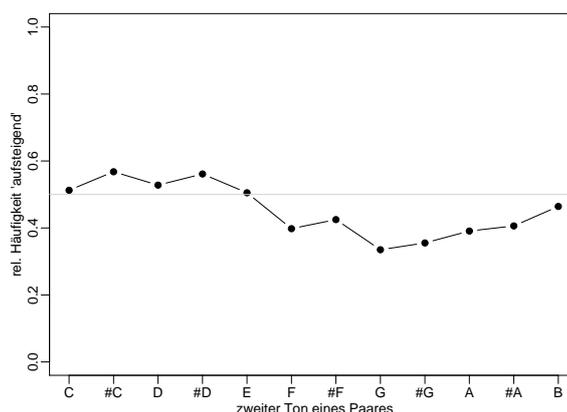


Abbildung 6.6. Die relative Häufigkeit der Antwort *aufsteigend* aller Hörer, abgetragen für den zweiten Ton der Tritonusintervalle.

Urteile für jedes Reizpaare und für jede Person berechnet. Bei einer mittleren Varianz von 0.15 wurden starke individuelle Unterschiede deutlich. Die individuellen Varianzen variierten von 0.07 bis 0.24 und überspannten somit einen weiten Bereich zwischen dem minimal möglichen Wert 0 und dem Maximum 0.25. Zur statistischen Überprüfung der Unterschiede wurde eine einfaktorische Varianzanalyse mit Messwiederholung mit den arcsin transformierten Varianz-Mittelwerten des Faktors Tonpaar berechnet. In den Resultaten zeigt sich über die 12 Tritonuspaaren hinweg kein signifikanter Haupteffekt ($F_{11,232} = 0.70, p = 0.73$). Trotz der hohen interindividuellen Variation der Sicherheit bei der Beurteilung der Tonpaare, lassen sich keine systematischen Einflüsse nachweisen. Die individuellen Antwortmuster sind in Abbildung 6.7 dargestellt.

Einfluss der Chromaklasse. In Tabelle 6.3 wird weiterhin deutlich, dass die Antworten in Abhängigkeit von der Chromaklasse variieren. So wird das Tonpaar Cis-G (133/77) am häufigsten als absteigend und damit konsistent das Tonpaar G-Cis (98/112) am häufigsten als aufsteigend wahrgenommen. Demnach wird, betrachtet man die Daten aller Personen, die Tonklasse Cis am höchsten wahrgenommen. Zur Beantwortung der Frage, ob das Chroma einer Tonklasse einen Einfluss auf die Tonhöhenurteile hat, bestimmte Tonklassen also als höher bzw. tiefer eingeschätzt werden als andere, wurde ein Chi-Quadrat-Test auf Gleichverteilung gerechnet. Die Ergebnisse weisen auf einen substantiellen Einfluss der Chromaklasse hin ($\chi^2 = 55.87, df = 11, p < 0.001$). Auch in den über alle Teilnehmer gemittelten Antwortmustern (siehe Abb. 6.6) lässt sich eine systematische Verschiebung der 'höher'-Urteile zugunsten der Chromaklassen Cis, D gegenüber den Tonklassen G, Gis erkennen.

Die *'peak-pitch-class'*. Da die Bestimmung einer am höchsten wahrgenommenen Tonklasse (peak-pitch-class, PPC) aufgrund von Mehrdeutigkeiten nicht in allen Fällen möglich ist, wurde für jeden Teilnehmer die höchste Tonklasse mit dem P4-Modell (siehe Abschnitt 5.2) geschätzt. Dazu wurden die vier Parameter des P4-Modells mittels Kleinst-Quadrat-Schätzung an die Daten angepasst. Die Parameterschätzung wurde mit der freien Statistik-Software R (Version 2.2.0, R Development Core Team, 2010) durchgeführt. Die resultierenden theoretischen Antwortwahrscheinlichkeiten sowie die geschätzten Parameter sind in Abb. 6.7 abgetragen. Das üblicherweise für zwei vollständige Dimensionen ausgelegte Modell wurde hier auf das Tritonusintervall (Distanz 6 im Uhrzeigersinn) beschränkt. Die Güte der Modellanpassung wurde mit einem Chi-Quadrat-Anpassungstest überprüft (vgl. Abb. 6.7). Dabei zeigt sich, dass bis auf einen Fall (JH) gute bis sehr gute Anpassungen erzielt wurden ($\bar{p} = 0.92$).

Um die peak-pitch-class zu berechnen, wurde der geschätzte Parameter Phase P mit folgender Funktion transformiert

$$P_{ppc} = 12 - \left((P - 3) \bmod 12 \right). \quad (6.10)$$

Der auf diese Weise berechnete Wert P_{ppc} beschreibt die am höchsten wahrgenommene Tonklasse. Dabei handelt es sich um die Schätzung der Tonklasse auf einer stetigen, zirkulären Dimension, dem Chroma. Anders als in den Arbeiten von z.B. Deutsch (1986) oder Giangrande et al. (2003) muss dieser Parameter nicht zwangsläufig mit einer bestimmten Tonklasse korrespondieren, sondern kann beliebig genaue Zwischenwerte annehmen³.

Durch die Funktion 6.10 wird das Maximum der Sinusfunktion berechnet, welche im Modell die Position der psychometrischen Funktionen festlegt. Die in Abb. 6.7 dargestellte Modellfunktion beschreibt die Wahrscheinlichkeit der Antwort 'Tonpaar ist aufsteigend' bei dem gegebenen zweiten Ton eines Paares. Dieser Graph korrespondiert mit der Sinusfunktion, das Maximum entspricht dem der Sinusfunktion bei $\pi/2$ (hier Phase $P - 3$ Halbtonschritte). Der Modulo-Operator gewährleistet vergleichbare Werte im Intervall $[0,12[$ (Chromaklassen $0 = C$ bis $11 = B$). Statt der am höchsten wahrgenommenen Tonklasse hätte auch die tiefste Tonklasse oder ein beliebig anderer Lageparameter gewählt werden können. Ausgehend von der P_{ppc} lässt sich z.B. die tiefste Tonklasse leicht durch einen Phasenversatz von 6 Halbtonschritten ermitteln.

³Die Bezeichnung „Tonklasse“ trifft hier streng genommen nicht mehr zu, da die geschätzten Wert zwischen den dargebotenen Tönen der chromatischen Tonleiter liegen. Aus Gründen der Einfachheit wird jedoch die übliche Bezeichnung beibehalten.

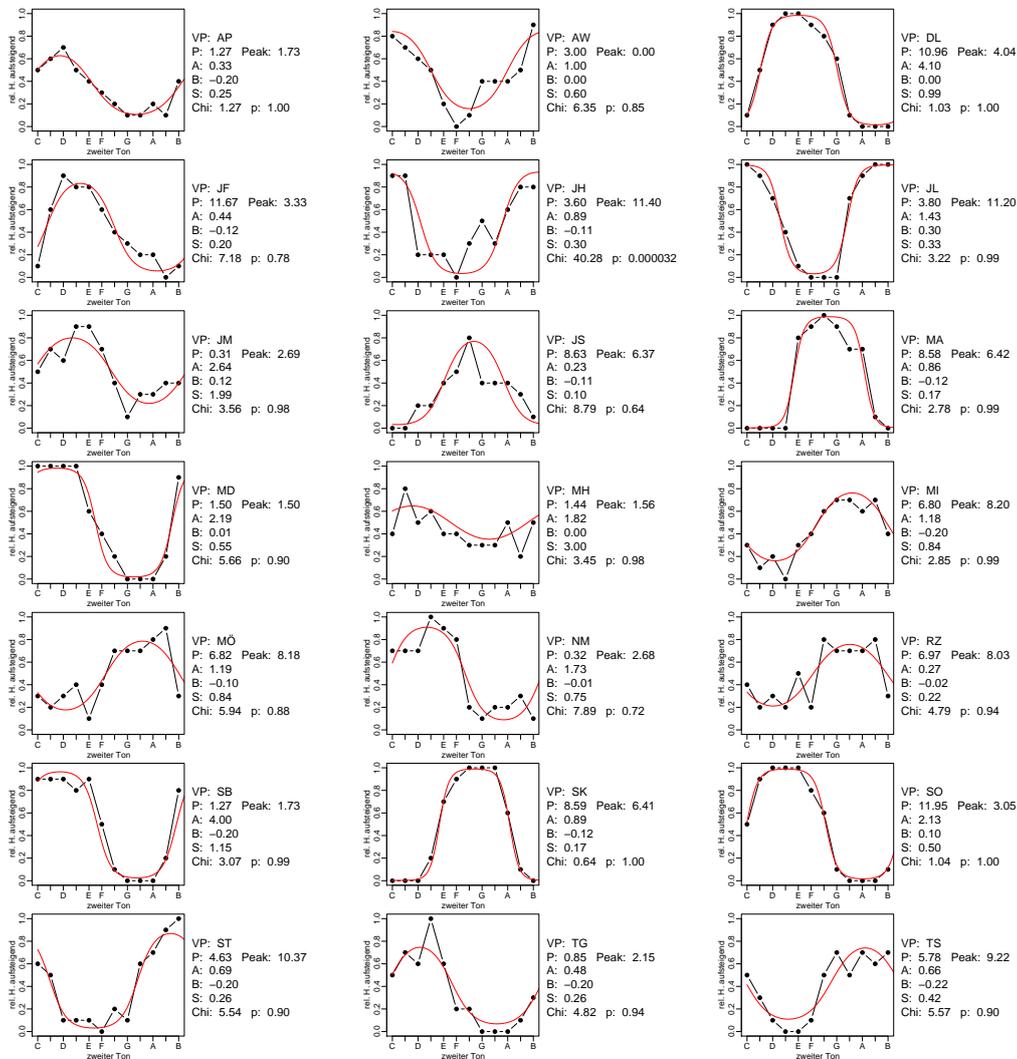


Abbildung 6.7. Anpassungen des P4-Modells an die individuellen Antwortmuster. Die relative Häufigkeit der Antwort *aufsteigend*, abgetragen für den ersten Ton eines Paares je Versuchsteilnehmer. Die durchgezogene Linie beschreibt die an die Daten angepasste theoretische Antwortwahrscheinlichkeit. Weiterhin sind für jeden Hörer die vier geschätzten Parameter, die Position der höchsten Tonklasse (Peak) und die Statistik der Modellanpassung abgetragen.

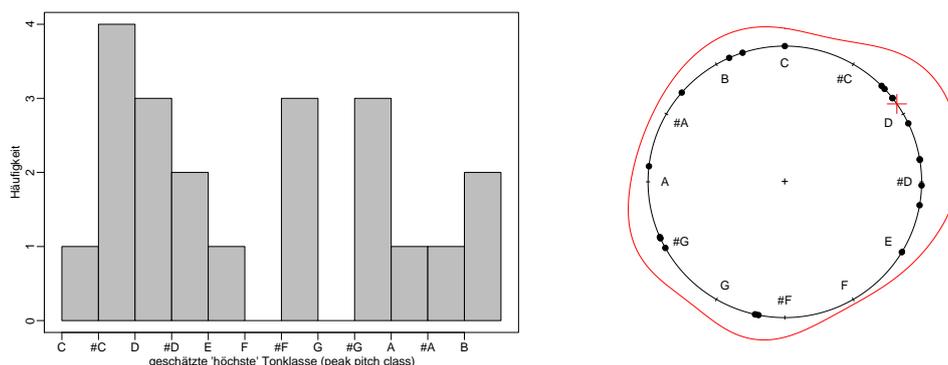


Abbildung 6.8. Die mit dem P4-Modell geschätzten höchsten Tonklassen (peak-pitch-class) aller Hörer als Histogramm und in der Kreisdarstellung. Die Kreisdarstellung beinhalten weiterhin den zirkulären Mittelwert aller Schätzungen (markiert durch ein \times) und die Verteilung der Schätzungen auf dem Kreis (zirkuläre Dichtefunktion aus R 2.2.0 mit den Glättungsparametern $bw=25$, $adjust=0.4$).

Die individuellen Schätzungen für die peak-pitch-class sind in Abbildung 6.7 notiert. Eine graphische Darstellung aller geschätzten Tonklassen findet sich in Abb. 6.8. Im Histogramm zeigt sich, dass die höchsten Tonklassen, bis auf die Intervalle zwischen F und Fis und zwischen G und Gis, im gesamten Chromaspektrum zu finden sind. Eine geschlossene Häufung tritt um die Tonklassen Cis und D auf, wobei häufigere Ausprägungen auch bei den zirkulär gegenüberliegenden Intervallen Fis-G und Gis-A zu finden sind. Der Mittelwert aller Schätzungen liegt bei 1.83 und entspricht damit einer theoretisch höchsten Tonklasse zwischen den Klassen Cis und D.

Die zirkuläre Varianz der individuellen höchsten Tonklassen beträgt 0.83. Diese Statistik für speziell zirkuläre Daten nimmt Werte im Intervall $[0,1]$ an (Jammalamadaka & SenGupta, 2001). Beispielsweise wird eine maximale Varianz bei einer Gleichverteilung, also bei stets gegenüberliegenden Datenpunkten erreicht, z.B. ergibt der Vektor $(0, \pi/2)$ oder der Vektor $(0, \pi/2, \pi, 3\pi/2)$ eine Varianz von 1. Im vorliegenden Fall zeigt sich, dass die peak-pitch-classes über einen sehr weiten Bereich, nahe der maximalen Streuung, variieren. Zur statistischen Überprüfung, wurde ein Rayleigh-Test auf zirkuläre Homogenität durchgeführt. Dieser Test prüft die Gleichverteilung zirkulärer Daten gegen die Alternativhypothese einer unimodalen Verteilung mit unbekanntem Mittelwert und unbekannter Varianz. Es zeigt sich keine signifikante Abweichung von einer zirkulären Gleichverteilung ($R = 0.17$, $p = 0.54$).

Reihenfolgeeffekte. Zur Überprüfung, ob ein systematischer Einfluss eines

präsentierten Tonpaares auf das Urteil im nachfolgenden Trial besteht, wurde ein Chi-Quadrat-Test auf Gleichverteilung gerechnet. Wird vorausgesetzt, dass es keinen Reihenfolgeeffekt gibt, so sollten sich die Antworten der 12 Tonpaare in Bezug auf die Antwort im nachfolgenden Trial nicht voneinander unterscheiden. Es zeigt sich ein gering ausgeprägter, nicht signifikanter Zusammenhang ($\chi^2 = 32.03$, $df = 11$, $p = 0.073$). Der größte Effekt wird dabei bei der Präsentation der Tonpaare F - B (130/80) und D - Gis (129/81) erreicht. Der Einfluss betrifft keine bestimmten chromatisch ähnlichen, auf dem Kreis nebeneinander positionierten Tritonuspaare, sondern erscheint unsystematisch einzelne Klassen zu betreffen. Auch lässt sich ein einfacher, autokorrelativer Zusammenhang zwischen zwei aufeinander folgende Trials nicht findet ($r_{lag1} = -0.014$).

Korrelate zur subjektiv höchsten Tonklasse. In den Vorexperimenten wurden neben den demographischen Daten der Versuchspersonen auch die Hörschwelle und die Diskriminationsfähigkeit erhoben. Die Frage, ob diese Variablen einen Einfluss auf die Beantwortung der Tritonusaufgabe haben, muss verneint werden. Es zeigten sich keine signifikanten Zusammenhänge der subjektiv höchsten Tonklassen mit dem Alter ($r=-0.01$, $p=0.96$), mit der Händigkeit ($r=0.09$, $p=0.36$), mit dem Geschlecht ($r=0.03$, $p=0.88$) oder der Musikalität ($r=-0.19$, $p=0.39$) der Probanden.

Auch ließen sich aus den Ergebnissen vom Vorexperiment zur Audiometrie keine Zusammenhänge zur Tonhöhenpräferenz ableiten. Ein mäßiger, nicht signifikanter Zusammenhang fand sich bei der Korrelation der subjektiv höchsten Tonklasse mit der Abweichung von der Normhörschwelle (Variable *Norm-Diff*, siehe Tabelle 6.2, $r=0.39$, $p=0.062$). Ebenfalls fand sich kein Zusammenhang mit der Normabweichung im Hauptsprachbereich gem. *ICD-10*, ($r=0.23$, $p=0.34$) und der Rechts-Links-Abweichung der Hörschwelle (*RLDiff*, $r=0.33$, $p=0.12$).

Im Vorexperiment 2 wurde die Diskriminationsfähigkeit für Tonhöhe getestet, auch für diese Fehlerrate konnte kein Zusammenhang mit der Tritonusaufgabe nachgewiesen werden ($r=-0.02$, $p=0.92$).

6.3.4 Diskussion

Im vorgestellten Experiment beurteilten 21 Personen die relative Tonhöhe von oktav-komplexen Tönen im Tritonusintervall. Überschautes man die Studien zum Tritonusparadox, so kann festgestellt werden, dass die im vorliegenden Experiment gefundenen Daten eine Mischung der oft konträren Befunde darstellt. So zeigt sich bei den meisten Personen das typische 6-Chromaklassen-aufsteigend

zu 6-Chromaklassen-absteigend Antwortmuster bei der Tritonusaufgabe. Andererseits zeigen sich hohe Schwankungen in Bezug auf die Sicherheit mit der ein Toonpaar bewertet wird. Die Befunde lassen sich allgemein gut in die von z.B. Deutsch (1986) oder Giangrande et al. (2003) gefundenen Daten einordnen. Auch verwundert es nicht, dass einzelne Personen nicht oder nur im Ansatz in der Lage sind, zwischen unterschiedlichen Tonhöhen zu unterscheiden.

In Bezug auf die Richtung der Antwortmuster auf eine präferierte höchste Tonklasse hin zeigt sich über alle Probanden hinweg eine Tonhöhenpräferenz zugunsten der Chromaklassen D/Dis. Zwar erreicht diese Ausprägung eine statistische Bedeutsamkeit, bei der Auswertung der höchsten Tonklassen (PPC) nach der Modellanpassung zeigt sich jedoch eine hohe Varianz, so dass mehrere Personen eindeutig andere Chromaklassen als höher bewerten. Geht man von den berichteten Verteilungen aus, so sollten die PPCs von Personen einer Sprachgemeinschaft im Bereich von ± 2 Halbtönen finden lassen (Deutsch, 1987). Die vorliegenden Daten streuen wesentlich weiter, so dass die Medien-Hypothese hier nicht bestätigt werden kann. Die verallgemeinerte Aussage, deutsche Hörer weisen die höchste Tonklasse im Bereich Cis/D auf, erscheint aufgrund der hohen Varianz nicht haltbar.

Reihenfolgeeffekte lassen sich zwar finden, diese scheinen jedoch keinem systematischen Muster zu folgen. Hier könnten Folgeexperimente z.B. mit verzögerter Bearbeitungszeit oder intermetierend dargebotenen Reizen (z.B. weißes Rauschen zur Löschung des Primingeffektes) zu weiterer Klärung beitragen. Weiterhin könnte an die Befunde von Repp und Thompson (2009) angeknüpft werden, welche das Tritonusparadox als stabilen Effekt darstellen, und die Resistenz gegenüber Reihenfolgeeffekten vermuten lassen.

6.4 Vollständiger Paarvergleich zum Einfluss der Teiltonanzahl

Mit dem folgenden Experiment soll überprüft werden, wie sich die Anzahl der Komponenten bei oktav-komplexen Tönen auf die Tonhöhenurteile auswirkt. Im speziellen Fokus dieses vollständigen Paarvergleiches steht die Frage, welchen Einfluss die Anzahl der Komponenten für die bei oktav-komplexen Tönen übliche Zirkularität in den Urteilen ausübt.

6.4.1 Fragestellung

Im Gegensatz zu der Vielzahl von Arbeiten zum Tritonusparadox liegen zur Frage nach dem Einfluss der Teiltonanzahl nur wenige Studien vor. In seinen Untersuchungen variiert Pollack (1978) u.a. die Anzahl der Teiltöne und erhebt die relative Tonhöhe einzelner Intervalle. Er findet heraus, dass um den Effekt der unendlichen Tonhöhe zu erzielen, eine Mindestanzahl von 3 Teiltönen je oktav-komplexen Ton vorhanden sein muss. Die Verallgemeinerbarkeit dieser Befunde bleibt jedoch diskutabel, da z.B. nur ausgewählte Intervalle präsentiert wurden und die Form der Hülle von der von Shepard (1964) oder Deutsch (1986) verwendeten Hülle abweicht (vgl. Kapitel 5).

Geht man von der Annahme aus, dass die Dimension des Chromas—im Gegensatz zur Dimension der periodischen Tonhöhe—stärker ausgeprägt ist, je mehr Teiltöne gleichen Chromas vorhanden sind, so lässt sich folgende Hypothese ableiten. Mit abnehmender Anzahl von Teiltönen

- wird eine steigende Unsicherheit der Urteile erwartet
- wird ein Übergang von zirkulären zu linearen Urteilen erwartet
- werden zwischen den Bedingungen keine Unterschiede, welche auf die Chromaklasse zurückzuführen sind, erwartet.

Dabei sollte das Modell des Tonklassenkreises seine Gültigkeit verlieren und die Urteile den Regeln harmonisch-komplexer Töne folgen.

Nimmt man ausgehend von Terhardt (1972b) an, dass die Teiltöne eines oktav-komplexen Tones sich als einzelne mental repräsentierte Tonhöhen wiederfinden und dass bei einem Vergleich zweier oktav-komplexer Töne ein Mechanismus wirkt, welcher korrespondierende Tonhöhen miteinander vergleicht, so kann weiterhin angenommen werden, dass die Anzahl der korrespondierenden Tonhöhen sich auf das relative Tonhöhenurteil (auf- oder absteigend)

auswirkt. So sollte bei maximaler Anzahl von Teiltönen (10 im hörbaren Frequenzbereich) auch die meisten Einzel-Tonhöhen-Vergleiche stattfinden. Dies jedoch nur unter der Voraussetzung, dass keine eindeutige fundamentale Frequenz extrahiert werden kann.

Es muss eine hohe Ambivalenz hinsichtlich der Fundamentalen bestehen. Dies ist bei oktav-komplexen Tönen der Fall. Das Frequenzspektrum eines oktav-komplexen Tones mit nur zwei Teiltönen ist identisch mit dem Frequenzspektrum eines harmonisch-komplexen Tones. Beide bestehen aus einem Grundton und einem Oberton, die Fundamentale ist klar definiert. Mit dem Zufügen weiterer Komponenten nach der jeweiligen Regelmäßigkeit differieren die Spektren beider Klänge.

Die genannten Hypothesen sollen auf zwei Arten überprüft werden, (1) auf der Grundlage der erhobenen 'Roh'-Daten und (2) anhand der an diesen Daten angepassten Modellparameter des P4-Modells. Folgende Parametervariation wird auf der Grundlage der genannten Hypothesen erwartet.

- *Amplitude*: Der Parameter Amplitude sollte sich in Bezug auf die Anzahl der Teiltöne nicht ändern. Wenn eine Änderung zu finden ist, so sollte diese mit einer Änderung des Parameters *Slope* einhergehen und durch eine Konfundierung beider Parameter erklärt werden können.
- *Phase*: Für die Phase bzw. die präferierte Tonklasse wird eine individuell verschiedene, über die Bedingungen (Anzahl der Teiltöne) hinweg jedoch konstante Ausprägung erwartet.
- *Slope*: Für die Steigung der Modellfunktion wird erwartet, dass der Parameter *Slope* mit der Anzahl der Teiltöne abnimmt und somit und die Sicherheit der Urteile zunimmt.
- *Bias*: Für den Parameter *Bias* werden intraindividuelle Variationen erwartet. Diese sollten über die Bedingungen hinweg Konstanz aufweisen. Ein Einfluss der Teiltonanzahl auf die Entscheidungspräferenzen (Auf- bzw. Absteigend-Antwort-Bias) ist nicht zu erwarten.

6.4.2 Methode

Versuchspersonen

An dem Experiment nahmen zehn Versuchspersonen im Alter von durchschnittlich 24 Jahren (SD=4.1) teil. Die 8 weiblichen und 2 männlichen Hörer

waren alle Studenten oder Mitarbeiter der Martin-Luther-Universität Halle-Wittenberg. Alle Teilnehmer nahmen freiwillig an den fünf ca. 50 Minuten dauernden Sitzungen teil. Als Aufwandsentschädigung erhielten sieben Hörer für die Teilnahme an den zwei Vorexperimenten und den fünf Sitzungen fünf Versuchspersonenstunden gutgeschrieben.

Aufgabe der Versuchsperson

Die Aufgabe der Versuchspersonen war es, Tonpaare hinsichtlich der Tonhöhenänderung als auf- oder absteigend zu beurteilen und ihre Entscheidung durch das Drücken einer Taste zu markieren.

Versuchsplan

Die Variation der Teiltonanzahl erfolgte in fünf experimentellen Bedingungen (2,4,6,8,10 Teiltöne je oktav-komplexem Ton). Jede Bedingung umfasste einen vollständigen Paarvergleich, in dem alle Paarkombinationen eines aus 12 Tonklassen bestehenden Sets, mit Ausnahme der Vergleiche eines Tones mit sich selbst, präsentiert wurden. Die fünf Bedingungen wurden geblockt und in zufälliger Reihenfolge dargeboten.

Stimuli

Die verwendeten oktav-komplexen Töne wurden nach der in Abschnitt 5 berichteten Funktion erstellt. Zur Vermeidung von Störgeräuschen durch abruptes Ein- bzw. Ausklingen wurden alle Reize mit 10 ms langen linearen Ein- bzw. Ausschwingphasen generiert. Inklusive dieser Phasen betrug die Darbietungsdauer eines Tones 400 ms.

Die Variation der Teiltonanzahl unter der Hülle ist konfundiert mit der Hüllform und im weitesten Sinne auch mit der Position der Hülle entlang der Frequenzachse. Um diese Konfundierung möglichst gering zu halten, wurde das Zentrum der Hüllkurve für alle Bedingungen konstant gehalten. Hierbei wurde die Anzahl der Teiltöne pro oktav-komplexem Ton in fünf Bedingungen so variiert, dass das Zentrum der Hüllkurve über alle Bedingungen konstant positioniert war. Die Anzahl der Teiltöne und die korrespondierenden Frequenzen (f_{\min}) sind in Tabelle 6.4 aufgelistet. Die maximale Amplitude der einzelnen Teiltöne war über die Bedingungen hinweg konstant.

Tabelle 6.4

Teiltonanzahl und kleinste Frequenz unter der Hülle f_{\min} der fünf experimentellen Bedingungen.

Anzahl der Teiltöne	Bedingungen				
	2	4	6	8	10
Frequenz f_{\min} in Hz	261.6	130.8	65.4	32.7	16.4

Apparatur

Die Durchführung des Experimentes erfolgte unter den in Abschnitt 6.1 auf Seite 73 dargestellten technischen und räumlichen Bedingungen.

Durchführung

Vor Beginn eines Experimentes wurde die Hörfähigkeit jeder Versuchsperson mit zwei audiometrischen Vorexperimenten (siehe Abschnitt 6.1) erfasst. Nach der Präsentation der Instruktion auf dem Bildschirm sowie kurzer mündlicher Instruktion durch den Versuchsleiter wurde das Experiment gestartet.

Ein Trial setzte sich zusammen aus dem Reizpaar (400ms Ton1, 20ms Pause, 400ms Ton2) gefolgt von der Antwort der Versuchsperson und einer Pause von zwei Sekunden zum nachfolgenden Trial. Zur Beurteilung eines Reizpaares wurde kein Zeitlimit festgelegt.

Die Präsentation der Reize erfolgte in zwei Teilen, einem mit vorab definierter Abfolge und einem adaptiven Teil. Im ersten Teil wurde jedes der insgesamt 132 Paare (12 Tonklassen \times 11 Intervalle) in sieben Blocks dargeboten, insgesamt 660 Trials. Dabei wurde die Reihenfolge zwischen den Hörern randomisiert. Dies erfolgte unter der Bedingung, dass innerhalb der 132 Trials gleiche Intervalle geblockt präsentiert wurden. Die Reihenfolge der Intervalle wurde ebenfalls randomisiert. Die geblockte Darbietung der 11 Intervalle erfolgte mit dem Ziel, die durch Intervallwechsel verursachten Reihenfolgeeffekte zu mindern und somit die Stabilität der Urteile zu erhöhen.

Nach Beendigung dieses ersten Teiles wechselte die Reizauswahl in einen adaptiven Modus. Hierbei wurde die bereits in Abschnitt 6.1 dargestellte adaptive Stufen-Methode (ML-PEST) verwendet. Für jeden der zwölf Starttöne (erster Ton eines Paares) wurde die Distanz im Uhrzeigersinn zum zweiten Ton als unabhängige Reizdimension betrachtet. Die Stufen (staircases) aus denen die Reize für den folgenden Trial auf der Grundlage der bisherigen Antworten gezogen wurden, waren die elf Distanzen im Uhrzeigersinn.

Nach jedem Trial wurde eine psychometrische Funktion mit zwei Parametern (vgl. Abschnitt 6.1 Funktion 6.6) an die bis zu diesem Zeitpunkt erhobenen Antworten angepasst. Dabei wurden auch die Antworten aus dem ersten Teil des Experimentes mit einbezogen. Während i.d.R. kleine Distanzen eher als aufsteigend wahrgenommen wurden, wurden größere Distanzen eher absteigend beurteilt. Da die Reize zufällig aus einer Gauß-Verteilung um den geschätzten Lageparameter (α) gezogen wurden, führte dies zur Auswahl von ambivalent beurteilten Intervallen im Bereich des Tritonus.

Hier wird auch das Ziel der zusätzlichen, adaptiven Reizauswahl deutlich. Mit Blick auf die anschließende Modellanpassung gewährleistet die adaptive Methodik eine zusätzliche Datenerhebung auf Reizstufen mit hohem Informationsgehalt wobei eine Mindestanzahl an Präsentationen durch den vollständigen Paarvergleich sichergestellt wird.

Der adaptive Teil des Experimentes wurde beendet, wenn das vorher definierte Abbruchkriterium erreicht wurde. Dies konnte durch das Erreichen der maximalen Trialsanzahl oder durch das Unterschreiten eines Grenzwertes für die Varianz des geschätzten Lageparameters erfolgen (vgl. 6.1).

Die Zahl der adaptiven Trials war für jeden der 12 Startton-Bedingungen auf minimal 20 und maximal 40 Trials festgelegt. Somit konnte eine Sitzung um minimal 240 und maximal 480 Trials verlängert werden. Nach Erreichen der Mindestanzahl an Trials wurde eine Startton-Bedingung beendet, falls die Varianz der letzten 5 Trials einen Wert von 0.5 unterschreitet. Die Normalverteilung um den geschätzten Lageparameter, aus der zufällig der Vergleichsreiz für den kommenden Trial gezogen wurde, hatte eine konstante Standardabweichung von 3.0 Halbtonschritten.

Über beide Teile hinweg wurde nach jeweils 350 Trials eine Pause angeboten, welche durch das Drücken einer beliebigen Taste beendet werden konnte.

6.4.3 Ergebnisse

Ergebnisse der adaptiven Methodik. Ein Hörer bearbeitet durchschnittlich 4670 Trials. In jeder der fünf Sitzungen wurden somit im Mittel 933 Tonpaare beurteilt. Die in jeder Sitzung geleistete durchschnittliche Anzahl der zusätzlichen adaptiven Trials betrug 278. Im Verlauf der Parameteranpassung, bei der für jeden der 12 Starttöne eine psychometrische Funktion über die Distanz im Uhrzeigersinn geschätzt wurde, wurden für jede dieser 12 Dimensionen durchschnittlich 23 zusätzliche Tonpaare dargeboten. Die Gesamtzahl der dargebotenen adaptiven Trials je Bedingung ist in Tabelle 6.5 dargestellt. In Bezug

Tabelle 6.5

Die Anzahl aller adaptiv dargebotenen Trials je Bedingung, die relative Häufigkeit der Antwort 'aufsteigend' und die über alle Tonpaare gemittelten Varianzen der Abweichung von den erwarteten Antworten sensu Shepard, abgetragen für alle Bedingungen.

Anzahl der Teiltöne	Bedingungen				
	2	4	6	8	10
N Trials	9408	9370	9333	9246	9336
N(aufsteigend)/N	0.48	0.48	0.48	0.49	0.50
Varianz	0.216	0.193	0.185	0.183	0.174

auf die zusätzlich dargebotenen adaptiven Trials lassen sich keine signifikanten Unterschiede zwischen den Bedingungen finden ($\chi^2 = 1.54, df = 4, p = 0.81$). Hinsichtlich der abgegebenen Urteile (aufsteigend vs absteigend) gab es ebenfalls keine signifikanten Differenzen zwischen den Bedingungen ($\chi^2 = 15.0, df = 12, p = 0.24$).

In Abbildung 6.9 sind alle Bedingungen die relativen Häufigkeiten je Startton und Distanz im Uhrzeigersinn abgetragen. In Abbildung 6.9 unten rechts sind die über alle Bedingungen gemittelten Antworten dargestellt. Wie auf den ersten Blick zu erkennen ist, weichen die Antwortmuster über die Bedingungen hinweg systematisch voneinander ab. Je geringer die Anzahl der Teiltöne wird, je stärker tendiert die Antwortverteilung von der üblichen „Wasserfall“-Form zu Mustern mit „winkelhalbierender“ Struktur. Deuten diese Veränderungen bereits auf einen Wechsel von zirkulären Urteilen zu direktionalen Urteilen auf der Grundlage der fundamentalen Frequenz hin?

Zur Überprüfung der Fragestellung, ob die Sicherheit der Urteile mit der Anzahl der Teiltöne einhergeht, wurden die Abweichungen der Antworten von einem theoretischen Antwortmuster auf der Grundlage der Shepardschen Theorie berechnet. Tonintervalle kleiner als 6 Halbtonschritte sollten demnach stets aufsteigend und Intervalle größer als 6 Halbtonschritte sollten stets absteigend bewertet werden. Paare im Tritonusintervall wurden von der Berechnung der Sicherheit der Urteile ausgeschlossen. Die über die Tonpaare gemittelte Varianz der Abweichungen sind in Tabelle 6.5 abgetragen. Es zeigt sich, dass die Abweichung vom erwarteten Antwortmustern nach Shepard mit abnehmender Teiltonzahl zunimmt.

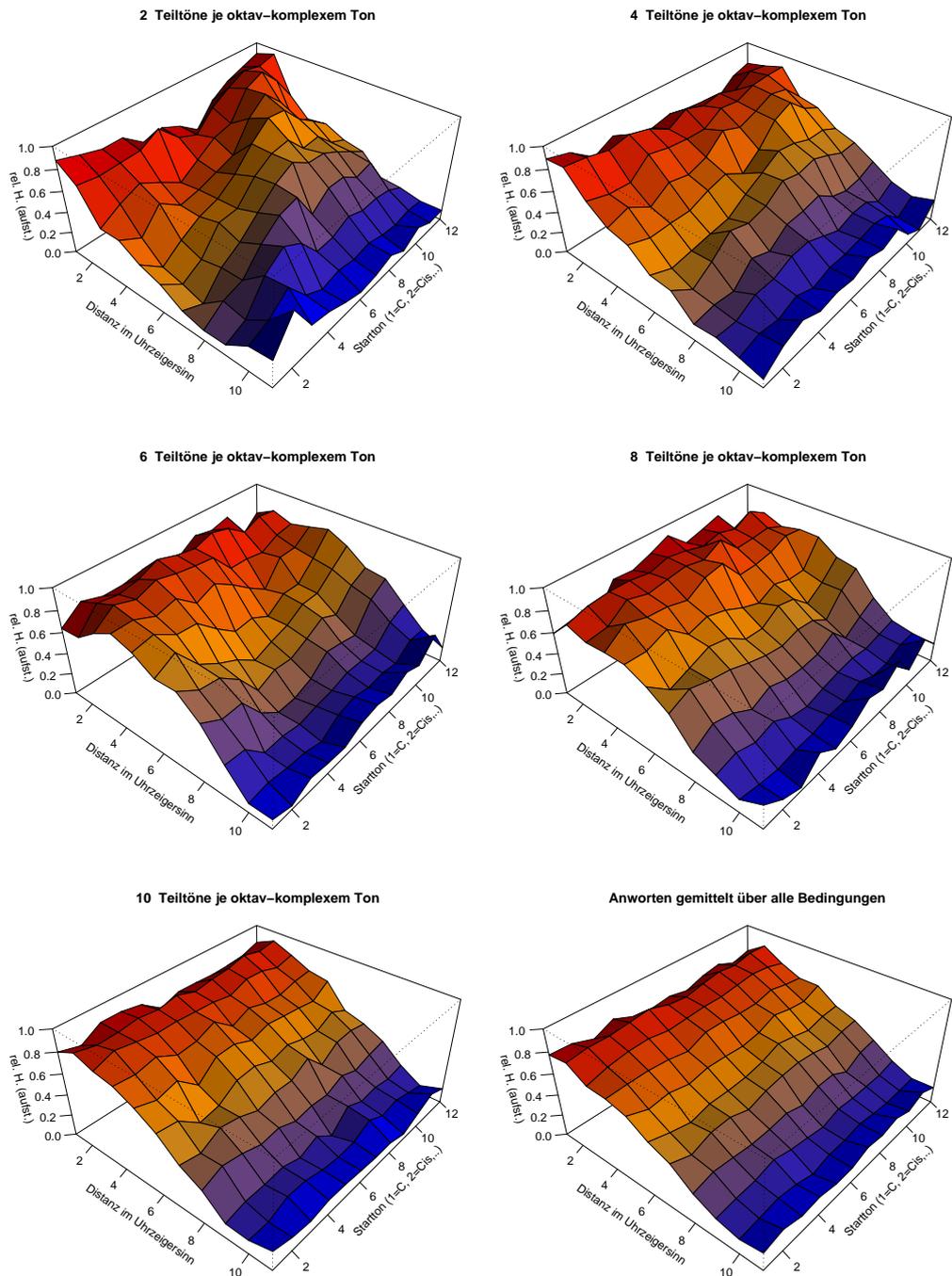


Abbildung 6.9. Die relativen Häufigkeiten der Antwort 'Tonhöhe ist aufsteigend' je Startton und Distanz im Uhrzeigersinn abgetragen für alle Bedingungen (2,4,6,8 oder 10 Oktaven werden von der Hülle überspannt) und dem Mittelwert aller Bedingungen.

Die zur Überprüfung der statistische Bedeutsamkeit der Unterschiede zwischen den Bedingungen wurde ein Bartlett-Test auf Homogenität der Varianzen durchgeführt. Der Bartlett-Test ist die multifaktorielle Variante des F-Tests und wird zur Überprüfung der Varianzgleichheit von mehr als zwei Gruppen eingesetzt (Bartlett, 1937). Es zeigt sich ein hoch signifikanten Einfluss der Teiltonanzahl auf die Stabilität der Urteile ($K^2 = 69.66, df = 4, p < 0.001$).

Parameterschätzung

Für jede Person wurde getrennt nach Bedingungen das P4-Modell mittels Kleinst-Quadrat-Schätzung an die Daten angepasst. Die geschätzten Parameter sowie die χ^2 -Anpassungen sind zusammen mit den Abbildungen der Anwohnhäufigkeiten und der Modellschätzungen in Anhang 6.9 aufgelistet. Nimmt man einen Annahmebereich von $p > 0.2$ an, so wurde bei 26 von insgesamt 50 Datensätzen eine sehr gute bis gute Anpassung erreicht. Dabei zeigt sich, dass die Spanne der Anpassungsstatistik viele extreme Werte annimmt. Nach Augenschein werden, bis auf Einzelfälle, die wesentlichen Lageparameter, z.B. Ausprägung der peak-pitch-class (Parameter Amplitude) oder die Sicherheit der Urteile (Parameter Steigung) hinreichend erfasst. Die empirischen und theoretischen Häufigkeiten für alle Probanden und alle Bedingungen sind im Anhang 6.9 dargestellt.

Um zu überprüfen, welchen Einfluss die Anzahl der Teiltöne auf die Tonhöhenurteile ausübt, wurden für jeden Parameter das arithmetische Mittel getrennt nach Bedingungen berechnet (siehe Abb. 6.10). Da der Parameter Phase Werte auf einer zirkulären Dimension annimmt, wurde entsprechend das zirkuläre arithmetische Mittel berechnet⁴.

Der Parameter *Amplitude* kann als Abweichung der Urteile von einer perfekten theoriekonformen Beurteilung sensu Shepard interpretiert werden. Im Gegensatz zur bereits erwähnten Sicherheit der Urteile ist hier die Abweichung von einem für alle Tonklassen identischen Antwortmuster oder bildlich ausgedrückt von einem „idealisierten, homogenen Wasserfall“ gemeint. Wäre die Amplitude Null, so würde das Modell vorhersagen, dass kein Tritonusparadox auftritt. Alle Urteile bei Tritonuspaaren würden in der Hälfte der Fälle aufsteigend bzw. absteigend beurteilt werden. Der Parameter Amplitude kann unter Einbeziehung der Parameter Steigung und Bias somit auch als *Effekt des Tritonusparadoxes* bezeichnet werden. Die Ausprägung der Amplitude wird in

⁴Die statistischen Methoden für zirkuläre Daten sind im Paket `circular` der freien Statistik-Software R enthalten.

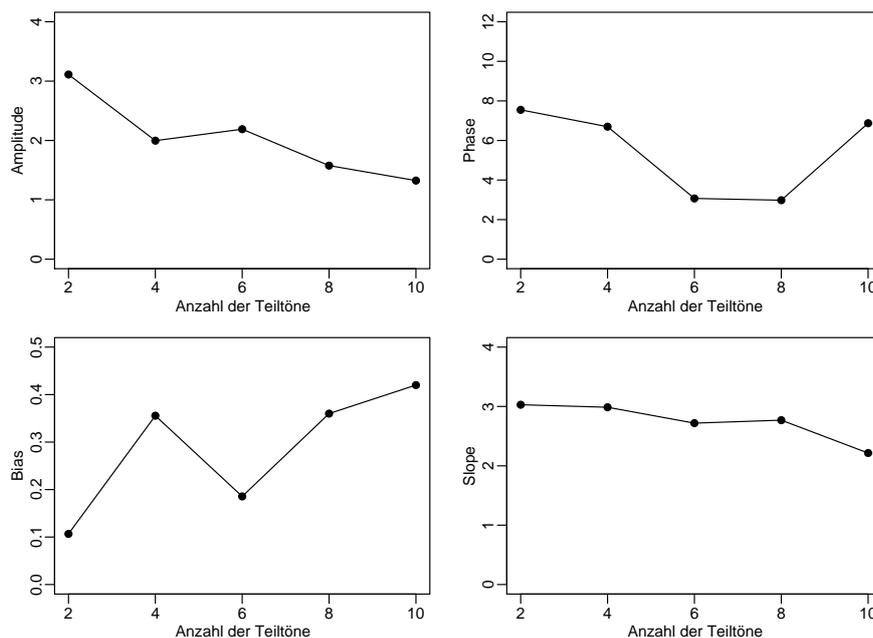


Abbildung 6.10. Die Ausprägung der über alle Personen gemittelten, geschätzten Parameter abgetragen für jede der fünf Bedingungen.

Halbtonschritten angegeben.

In Abbildung 6.10 zeigt sich ein deutliche Verringerung der Amplitude über die fünf Bedingungen hinweg. Um die statistische Bedeutsamkeit beurteilen zu können, wurde eine Varianzanalyse mit Messwiederholung gerechnet. Es kann gezeigt werden, dass die Anzahl der Teiltöne einen hoch signifikanten Effekt auf den geschätzten Modellparameter Amplitude ausübt ($F_{4,36} = 8.34, p < 0.001$). Je mehr Teiltöne ein oktav-komplexer Ton beinhaltet, desto geringer ist der Effekt der Tonklasse (Chroma) in Bezug auf die Abweichung von einem einheitlichen Antwortmuster. Berücksichtigt man wiederum die Parameter Steigung und Bias, so kann gesagt werden, dass bei zunehmender Teiltonanzahl der Effekt des Tritonusparadoxes abnimmt.

Der Parameter *Phase* gibt Auskunft darüber, welche Tonklassen bezüglich der Tonhöhe präferiert werden bzw. bei welchen Tonklassen der durch den Parameter Amplitude bezifferte Effekt des Tritonusparadoxes auftritt. Er korrespondiert mit dem von Deutsch (1986) vorgeschlagenen Maß der subjektiv höchsten Tonklasse beim Tritonusparadox (peak-pitch-class) und kann in dieses überführt werden. Beim Vergleich der Mittelwerte (vgl. Abb. 6.10) wird deutlich, dass über die Bedingungen hinweg keine Konstanz oder Monotonie zu finden ist, sondern die Mittelwerte signifikant voneinander abweichen. Dies wurde mittels einer ANOVA für zirkuläre Daten überprüft ($F_{4,36} = 6.42, p < 0.001$).

Tabelle 6.6

Varianz, Rayleigh-Test-Statistik und Annahmewahrscheinlichkeit p für den Parameter Phase für alle fünf experimentellen Bedingungen.

Anzahl der Teiltöne	2	4	6	8	10
Varianz (zirkulär)	0.50	0.64	0.30	0.83	0.77
Rayleigh Statistik	0.49	0.36	0.69	0.17	0.22
p	0.12	0.31	0.009	0.80	0.64

Interessant hierbei ist, dass zwischen den Bedingungen mit 2,4 und 10 Teiltönen und den Bedingungen mit 6 und 8 Teiltönen eine Differenz von etwa 4 Halbtonschritten besteht.

Um die subjektiv höchsten Tonklassen P_{ppc} zu ermitteln, wurde unter Anwendung der Transformationsfunktion 6.10 auf die geschätzten Phasenwerte das Maximum der Modellfunktion bei Distanz 6 berechnet. Der so ermittelte Wert gibt an, welche Tonklasse am höchsten wahrgenommen wird.

In Abbildung 6.11 sind die Verteilungen der individuellen Höchsttonklassen getrennt nach Bedingungen sowie in Kombination dargestellt. Die signifikanten Unterschiede zwischen den Bedingungen deuten auf eine eher polymodale Verteilung der präferierten Höchsttonklassen hin. Klare Häufungen im Sinne der Zuordnung der P_{ppc} -Werte zu einer Kreishälfte finden sich bei den Bedingungen 2, 6 und bis auf eine Person auch bei der Bedingung 4. In den übrigen Bedingungen 8 und 10 sowie in der Kombination aller Bedingungen sind keine Häufungen erkennbar. Um dies zu überprüfen, wurde ein Rayleigh-Test auf zirkuläre Homogenität durchgeführt. Dieser Test prüft die Gleichverteilung zirkulärer Daten gegen die Alternativhypothese einer unimodalen Verteilung. Wie aus Abbildung 6.11 und Tabelle 6.6 zu entnehmen ist, kann nur bei Bedingung 6 von einer eindeutigen unimodalen Verteilung ausgegangen werden. Hierbei zeigt sich jedoch eine breite Streuung der subjektiv höchsten Tonklassen im Bereich von Dis bis G.

Der Parameter *Slope* korrespondiert mit der Steigung der Modellfunktion entlang der Distanz-Achse und kann als Maß für die Reliabilität der Urteile interpretiert werden. Wie aus Abbildung 6.10 entnommen werden kann, nimmt der Parameter Slope stetig ab, was bedeutet, dass die Steigung der Modellfunktion und somit auch die Sicherheit der Urteile mit der Anzahl der Teiltöne zunimmt. Überprüft man den Einfluss der Teiltonanzahl auf die Stabilität der Urteile mittels einer ANOVA mit Messwiederholung, so erweist sich dieser

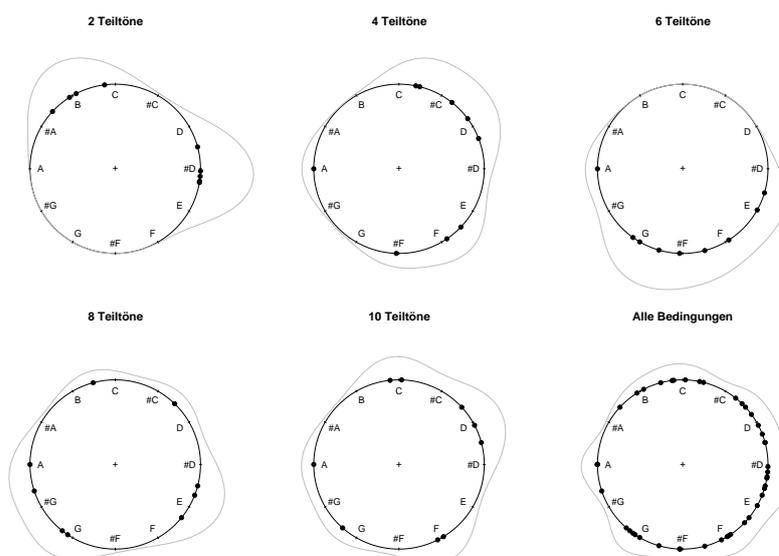


Abbildung 6.11. Die individuell höchsten Tonklassen (peak-pitch-classes, ppc) auf der Grundlage der geschätzten Phasenparameter aller Personen abgetragen auf dem Tonklassenkreis für die fünf Teiltonbedingungen und der Summe aller Bedingungen. Die grauen Außenlinien sind theoretische Dichtefunktionen auf der Grundlage der empirischen ppc-Werte. Sie veranschaulichen die Auftretenswahrscheinlichkeit der Phasenwerte unter Voraussetzung einer von-Mises-Verteilung (vergleichbar mit einer Normalverteilung auf einer zirkulären Dimension). Als Konzentrationsparameter der von-Mises-Verteilung wurde hier der Wert 10 gewählt.

Trend als nicht bedeutsam ($F_{4,36} = 0.78, p = 0.54$).

Der Parameter *Bias* gibt Auskunft über die Antwortverzerrung hinsichtlich einer Aufsteigend- oder Absteigend-Antwort-Tendenz und wird in Halbtönen angegeben. Beurteilt ein Hörer mehr Tonpaare aufsteigend als absteigend, so äußert sich das in einem positiven Biaswert. Werden vermehrt Tonpaare als absteigend beurteilt, so führt dies zu einem negativen Bias.

Die geschätzten Urteilstendenzen fallen insgesamt gering aus und übersteigen im Mittel einen halben Halbtönenschritt nicht. Die interindividuell stark variierenden Werte erreichen ein Maximum bei 2.37 Halbtönenschritten. Ein bedeutsamer Einfluss der Teiltonanzahl kann nicht nachgewiesen werden ($F_{4,36} = 0.46, p = 0.76$).

Zirkuläre vs Grundton-Urteile. An dieser Stelle soll überprüft werden, ob die Anzahl der Teiltöne einen Einfluss auf die Zirkularität von Tonhöhenurteilen ausübt. Gemäß der geschilderten Befunde wird erwartet, dass die Zirkularität mit abnehmender Teiltonanzahl in ein transitives Fundamental-Frequenz-

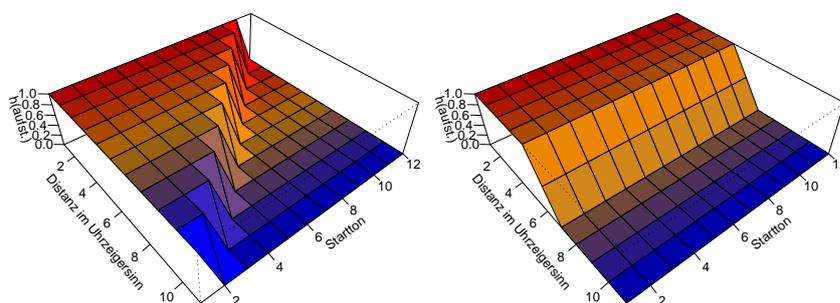


Abbildung 6.12. Darstellung der erwarteten Häufigkeiten der Antwort „Tonhöhe ist aufsteigend“ bei perfektem Grundton-Urteil, \mathbf{F} -Matrix (links) und bei perfektem Chroma-Urteil, \mathbf{C} -Matrix (rechts).

Urteil übergeht.

Zur Überprüfung dieser Hypothese wird eine individuelle Maßzahl, der *Fundamental-Chroma-Index (FCI)*, vorgeschlagen. Diese berechnet sich aus den Urteilen einer Person bei einem vollständigen Paarvergleich und gibt Auskunft darüber, ob ein Paarvergleich eher zirkulären oder transitiven Urteilen folgt.

Dazu wurden die relativen Häufigkeiten der Urteile „Tonpaar ist aufsteigend“ in die bekannte 11×12 Distanz-Startton-Matrix \mathbf{H} überführt. Weiterhin wurden zwei theoretische Matrizen aufgestellt (vgl. Abb. 6.12), eine mit den erwarteten Wahrscheinlichkeiten bei einem perfekten Grundton-Urteil (\mathbf{F}) und eine mit den erwarteten Wahrscheinlichkeiten bei einem perfekten Chroma-Urteil auf der Basis des Tonklassenkreises (\mathbf{C}).

Die Wahrscheinlichkeiten der Matrix der Grundton-Urteile \mathbf{F} entsprechen einer oberen Dreiecksmatrix und sind 1 (=aufsteigend) für alle Zellen f_{ij} für die gilt $i + j < 12$, sonst 0. Sie gleichen dem theoretischen Antwortmuster bei einem Paarvergleich von zwölf aufeinander folgenden natürlichen Klängen oder Sinustönen.

Die Wahrscheinlichkeiten in der Matrix der Chroma-Urteile \mathbf{C} sind ausschließlich durch die Distanz im Uhrzeigersinn definiert und somit für alle (Start-)Tonklassen identisch. Sie sind 1 für alle c_{ij} für die gilt $i < 5$, 0,5 für alle $i = 6$ und sonst 0. Das theoretische Antwortmuster in dieser Matrix entspricht Tonhöhenurteilen auf der Basis des Tonklassenkreises. Kurze Distanzen im Uhrzeigersinn zwischen zwei Tönen werden aufsteigend wahrgenommen,

längere Distanzen absteigend. Das Tritonusintervall mit der Distanz 6 erhält die Ratewahrscheinlichkeit 0.5. Der FC-Index wird wie folgt berechnet

$$FCI = \frac{\sum |\mathbf{H} - \mathbf{C}| - |\mathbf{H} - \mathbf{F}|}{36}. \quad (6.11)$$

Der Index wird aus der Differenz der Abweichungen von den jeweiligen theoretischen Wahrscheinlichkeiten gebildet. Er kann Werte von -1 bis 1 annehmen, wobei ein Wert von -1 für perfekte Grundton-Urteile steht und ein Wert von 1 für perfekte Chroma-Urteile. Ein Wert von 0 entspricht einer mittleren Abweichung und signalisiert demzufolge ein mehrdeutiges Antwortverhalten.

Der Teiler 36 ergibt sich, da beide Matrizen \mathbf{F} und \mathbf{C} aus jeweils 132 Zellen (11×12) bestehen, wovon ein Großteil der Zellen mit einem identischen Wert belegt sind. Die Teilmenge der Zellen mit verschiedenen Werten umfasst 42 Zellen. Nur diese 42 verschiedenen Zellen sind für die Berechnung relevant, da die Differenz von identischen Werten stets 0 ergibt. Da die 12 Tritonusintervalle der \mathbf{C} -Matrix mit Wahrscheinlichkeiten von 0.5 belegt sind, kann maximal eine Abweichung von insgesamt $12 \times 0.5 = 6$ errechnet werden. Somit erhält man die gewünschte relative Häufigkeit der Abweichungen der 132 Zellen bei einer Division der Summe der Abweichungen durch $42 - 6 = 36$.

Der vorgeschlagene FC-Index ist ein hypothesengeleiteter Wert, dessen Extreme zwar von den theoretischen Verteilungen festgelegt werden, empirisch sind jedoch viele andere Antwortmuster denkbar, welche mittlere Ausprägungen zur Folge hätten.

Die FC-Indizes für alle Personen getrennt nach Bedingungen sowie den Personen- und Bedingungsmittelwerten sind in Tabelle 6.7 abgetragen. Trotz der weiten Streuung der FC-Indizes (0.002 bis 0.94) liegen alle Werte im positiven Bereich, was auf eine Beurteilung der Tonhöhenänderung auf der Basis des Tonklassenkreises schließen lässt. Über die Mittelwerte der Bedingungen hinweg lässt sich, entgegen den Erwartungen, keine Steigung im FC-Index erkennen. Eine einfaktorielle Varianzanalyse mit Messwiederholung mit dem Faktor Teiltonanzahl ergab keinen signifikanten Unterschied zwischen den Bedingungen und somit keinen Hinweis auf den Einfluss der Anzahl der Teiltöne auf den FC-Index ($F_{4,36} = 2.03, p = 0.11$).

Der FC-Index kann aufgrund der genannten Überschneidung der beiden theoretischen Matrizen diese nicht gänzlich unabhängig voneinander erfassen. Daher wurden im Nachgang die Abweichungen von der \mathbf{F} -Matrix und die Abweichung von der \mathbf{C} -Matrix gesondert untersucht. In Abbildung 6.13 sind die Mittelwerte dargestellt. Es zeigt sich, wie auch schon am FC-Index, dass die Abweichungen von der idealen Chroma-Antwort ($M=0.24$) geringer sind als

Tabelle 6.7
Die FC-Indizes für alle Personen abgetragen für die fünf Bedingungen sowie deren Mittelwert.

VP	Bedingungen					Mittel
	2	4	6	8	10	
AW	0.187	0.772	0.649	0.441	0.717	0.45
JM	0.390	0.416	0.024	0.298	0.451	0.30
JO	0.764	0.748	0.722	0.623	0.581	0.72
MA	0.519	0.603	0.574	0.679	0.648	0.57
MM	0.840	0.566	0.019	0.247	0.431	0.58
SB	0.901	0.862	0.640	0.780	0.823	0.82
SK	0.002	0.529	0.343	0.221	0.145	0.22
SS	0.067	0.354	0.474	0.084	0.446	0.20
TM	0.948	0.535	0.010	0.069	0.178	0.50
TR	0.500	0.459	0.085	0.468	0.345	0.40
Mittel	0.512	0.584	0.345	0.391	0.477	0.47

die Abweichung von einem idealen Fundamental-Urteil ($M=0.37$). Weiterhin wird deutlich, dass die Abweichungen bei einer geringen Anzahl von Teiltönen (2, 4) höher ist als bei den übrigen Bedingungen. Allerdings gilt dies für die C-Abweichungen genauso wie für die F-Abweichungen. Die erwartete Gegenläufigkeit beider Parameter zeigt sich nicht. Im Gegenteil besteht ein hoher positiver Zusammenhang ($r = 0.89$). Untersucht man den Effekt der Anzahl der Teiltöne mit einer Varianzanalyse mit Messwiederholung, so finden sich in beiden Fällen signifikante Unterschiede (C-Abweichung: $F_{4,36} = 8.90, p < 0.001$; F-Abweichung: $F_{4,36} = 11.90, p < 0.001$). Mit sinkender Zahl der Komponenten in einem oktav-komplexen Ton steigt die Differenz zu den erwarteten Antwortmustern (vgl. Abb. 6.13).

6.4.4 Diskussion

Im vollständigen Paarvergleich von oktav-komplexen Tönen wurde der Einfluss der Anzahl der Komponenten auf die Beurteilung der Tonhöhe von Tonpaaren untersucht. Die Hypothese, dass sich mit zunehmender Anzahl der Teiltöne die Urteile vom Fundamental-Urteil hin zum Chroma-Urteil ändern, konnte nicht bestätigt werden. Dennoch zeigt sich, dass Hörer die dargebotenen Tonpaare eher nach dem Muster für zirkuläre Urteile als auf der Basis der fundamentalen Frequenz einschätzen.

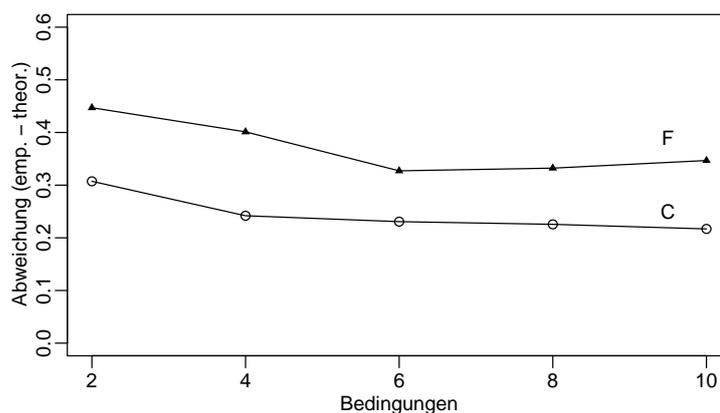


Abbildung 6.13. Die standardisierte Summe der Abweichungen der empirischen Antworten von den erwarteten Antworten gemäß der reinen Chroma-Urteile (C) und der reinen Fundamental-Urteile (F).

In Bezug auf die Parameter des P4-Modells bestätigte sich die Hypothese, dass mit zunehmender Anzahl von Teiltönen auch die Sicherheit der Urteile zunimmt, was sich im Abfall des Parameters Steigung (slope) zeigt.

Weiterhin konnte gezeigt werden, dass mit zunehmender Anzahl der Teiltöne die Zirkularität, gemessen an der Abweichung zur erwarteten Verteilung, zunimmt. Die von Pollack (1978) postulierte Mindestanzahl von drei Komponenten kann somit, unter Einwand der hohen individuellen Variationen, bedingt bestätigt werden.

Die im Anhang dargestellten individuellen Antwortmuster verdeutlichen die hohe Variation. So reichen die Urteile von einem eher zufälligen Antwortmuster (z.B. JM) über ein deutliches Chroma-Urteil gemäß der Shepardschen Theorie (z.B. MA) bis hin zu Urteilen auf der Grundlage der fundamentalen Frequenz (z.B. SK, AW). Andere, teilweise zu den erwarteten Mustern konträre Urteile lassen sich ebenso finden. Hier kann auch der Grund für die nur bedingte Aussagekraft des FC-Index vermutet werden.

Erstaunlich erscheint auch, dass die Urteile bei Tönen mit nur zwei Komponenten nicht stärker den Tonhöhenurteilen von harmonisch komplexen Tönen gemäß der F-Matrix ähneln. Das Antwortmuster bei Bedingung 2 (vgl. Abb. 6.9 oben links) lässt zwar einen Urteilswechsel entlang der Diagonalen der Startton-Distanz-Matrix (obere linke zur unteren rechten Dreiecksmatrix) erkennen, zeigt jedoch auch einen „Einbruch“ der Antwortsicherheit um den Bereich der Distanz 4/5 bei Startton Dis/E.

Im Falle des Tritonusintervalls zeigte sich, weder über die Personen noch über die Bedingungen ein einheitliches Muster. Geht man von der Medienhy-

pothese (vgl. Abschnitt 5.1.4) sensu Deutsch (1987) aus, so sollte die Anzahl der Teiltöne, zumindest in den höheren Ausprägungen (6,8,10), keinen Effekt auf die Wahrnehmung der peak-pitch-class ausüben. Eine unimodale Konzentration der P_{ppc} -Werte zeigt sich nur bei der Bedingung mit 6 Komponenten. Das Zentrum der Häufung liegt hier bei Tonklasse Fis. Dieses Ergebnis steht im Widerspruch zu den Befunden aus Experiment 3 zur Tritonusaufgabe. Dort lag die Häufung bei D/Dis.

Ein weiterer Kritikpunkt ist, dass die maximale Lautstärke der Einzelkomponenten zwar konstant gehalten wurde, da jedoch zwischen den Bedingungen die Anzahl der Komponenten variiert wurde, wurde auch die Lautstärke variiert. Diese Konfundierung kann einen weiteren Einflussfaktor auf die Tonhöhenbeurteilung darstellen. In einem Prüfexperiment könnte die Änderung der Lautstärke angeglichen werden und an die im Vorexperiment erhobene individuelle Absolutschwelle gekoppelt werden.

6.5 Vollständiger Paarvergleich zum Einfluss der Hüllenposition und -form

Mit diesem Experiment soll überprüft werden, welche Mechanismen bei der Beurteilung der Tonhöhe von oktav-komplexen Tönen wirken. Vor diesem Hintergrund wird eine alternative Hypothese zur Erklärung des sog. Tritonusparadoxons vorgeschlagen, überprüft und diskutiert. Weiterhin lassen sich Aussagen darüber treffen, welchen Einfluss die Form und die Position der Hüllkurve auf die Wahrnehmung von Tonhöhenänderungen bei Shepardreizen ausüben.

6.5.1 Fragestellung

Zur vorliegenden Fragestellung wurden zahlreiche Studien vor allem von Diana Deutsch und ihren Kollegen (vgl. Abschnitt 5.1.4) durchgeführt. Diese zeigen, dass die Position der Hülle keinen Einfluss auf die Tonhöhenurteile ausübt, d.h. Tonpaare werden unter verschiedenen positionierten Hüllen konsistent als auf- bzw. absteigend beurteilt.

Als Ursache postuliert Deutsch (1986) einen mental repräsentierten Tonklassen- oder Chromakreis (mental pitch template). Eine Person hat demnach eine feste, höchste Tonklasse bzw. eine ordinale Tonklassen-Struktur. Der mentale Tonklassenkreis bezieht sich nur auf das Chroma eines Tones und ist unabhängig von der Frequenz. Folglich sollte die Positionsänderung einer sonst konstanten Hüllkurve keinen Einfluss auf die Tonhöhenänderung ausüben.

Die empirischen Befunde hierzu sind widersprüchlich. So konnte z.B. Repp (1997) einen klaren Einfluss der Hüllenposition aufzeigen. Auch in eigenen Untersuchungen zur Tonhöhenwahrnehmung konnte ein Einfluss der Hüllenposition auf die Wahrnehmung von Tonhöhenänderungen nachgewiesen werden (Krüger & Lukas, 2002).

Eine alternative Hypothese zum Tritonusparadox könnte wie folgt formuliert werden. Das Tritonusparadox ist ein Effekt, der aufgrund eines Konfliktes zwischen zwei Hör- bzw. Urteils-Strategien entsteht, nämlich zwischen dem Fundamental-Urteil auf der einen und dem Distanz-Urteil auf der anderen Seite.

Es ist hinlänglich bekannt, dass kleine bzw. große Intervalle im Uhrzeigersinn stabil als auf- bzw. absteigend bewertet werden. Da hier die Distanz auf dem Chromakreis das Kriterium bildet, werden diese Urteile hier als Distanz-Urteile bezeichnet. Mit zunehmender Annäherung an das Tritonusintervall ändern sich die Urteile. Sie bleiben zwar weiterhin stabil, werden nun jedoch auch

von anderen Faktoren, wie dem Chroma oder der kleinsten bzw. fundamentalen Frequenz, beeinflusst.

Eine Variation der *Position* der Hülle sollte sich auf Urteile auf der Grundlage der fundamentalen Frequenz auswirken nicht aber auf Urteile auf der Grundlage einer individuell präferierten höchsten Chromaklasse. Sollte die fundamentale Frequenz einen Einfluss haben, so sollte sich die Verschiebung der Hüllkurve in einer Änderung der präferierten höchsten Tonklasse äußern. Weiterhin sollten alle Personen die gleichen oder ähnliche Tonklasse präferieren.

Die *Form* der Hülle—im Sinne der Wegnahme der Hüllkurve—sollte sich ebenfalls nur auf Urteile auf der Basis der Fundamentalen Frequenz auswirken, da die sonst durch die Hülle gefilterten Grundfrequenzen in diesem Falle deutlich wahrnehmbar sind.

Nimmt man an, dass es zwei Mechanismen beim Vergleich der Tonhöhe von komplexen Tönen gibt, so lässt sich die genannte Hypothese wie folgt untermauern: Je näher die Paare dem Tritonusintervall kommen, je identischer werden die Abstände zwischen benachbarten Teiltönen zweier Tonklassen. Die Distanz als alleiniges Kriterium reicht hier nun nicht mehr aus, da nicht sicher entschieden werden kann, welche Distanz die Kürzere ist und welche Teiltöne im Sinne der Gestaltgesetze zusammengehören. An diesem Punkt greift die „normale“ Urteilsgrundlage, die fundamentale Frequenz.

6.5.2 Methode

Versuchspersonen

An dem Experiment nahmen zehn Versuchspersonen im Alter von 22 bis 40 Jahren ($M=28.2$, $SD=6.0$) teil. Die 3 weiblichen und 7 männlichen Hörer waren alle Studenten oder Mitarbeiter der Martin-Luther-Universität Halle-Wittenberg. Alle Teilnehmer nahmen freiwillig an den vier ca. 50 Minuten dauernden Sitzungen teil. Als Aufwandsentschädigung erhielten fünf Hörer für die Teilnahme an den zwei Vorexperimenten und den vier Sitzungen vier Versuchspersonenstunden gutgeschrieben.

Aufgabe der Versuchsperson

Die Aufgabe der Versuchsperson war, zu entscheiden, ob der zweite Ton eines dargebotenen Tonpaares höher oder tiefer war als der erste Ton, die Tonhöhe also ansteigt oder abfällt. Das Urteil der Versuchsperson wurde durch das Drücken einer Pfeiltaste (hoch=aufsteigend, runter=absteigend) erhoben.

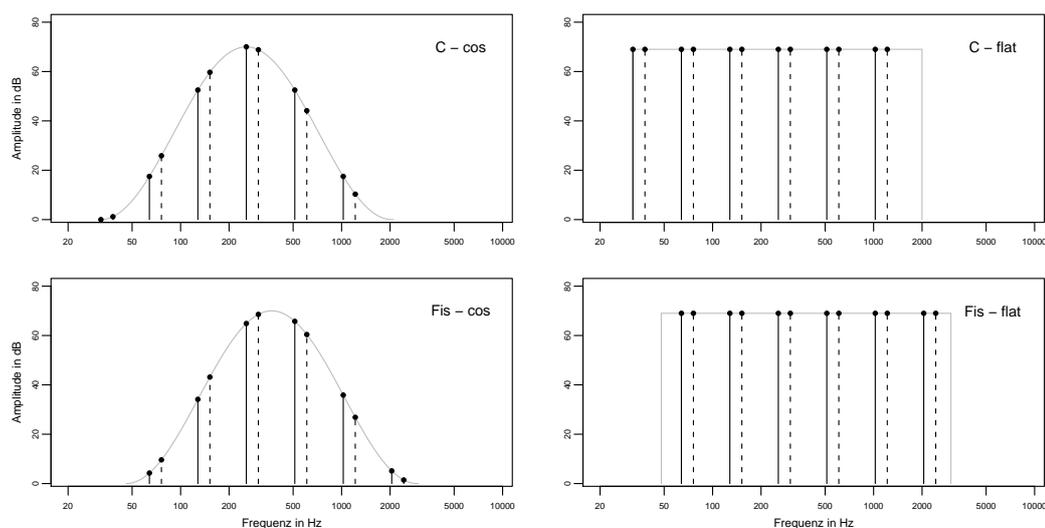


Abbildung 6.14. Die theoretischen Spektren der Tonklassen C (durchgezogene Linien) und Dis (gestrichelte Linien) abgetragen für die vier experimentellen Bedingungen (2 Hüllenformen und 2 Hüllenpositionen). Die jeweiligen Hüllkurven, welche hier gut die Funktion eines Band-Pass-Filters verdeutlichen, sind grau eingezeichnet.

Versuchsplan

Zur Überprüfung der Hypothese wurden vier experimentelle Bedingungen realisiert. Diese bestanden aus zwei Faktoren (Hülle, Position) mit jeweils zwei Stufen. Im Faktor Hülle wurde die Form der Hüllkurve so variiert, dass die Teiltonlautstärke entweder durch die übliche Glockenform oder durch eine gleichmäßige, flache Amplitude definiert wurde. Im Faktor Position wurde die Lage der Hüllkurve durch den Parameter f_{\min} mit den Ausprägungen 32.7 Hz (C) oder 46.2 Hz (Fis) variiert. Die Kombination der beiden Faktoren führte zu den vier in Abb. 6.14 dargestellten Bedingungen. In jeder Bedingung wurde ein vollständiger Paarvergleich der 12 oktav-komplexen Tönen durchgeführt.

Stimuli

Das verwendete Tonklassenset bestand aus den üblichen 12 oktav-komplexen Tönen, welche nach der in Abschnitt 5 dargestellten Funktion konstruiert wurden. Jede Tonklasse bestand aus 6 Teiltönen. Die glockenförmige Hüllkurve entsprach der herkömmlichen, in Abschnitt 5 beschriebenen Cosinusfunktion. Für die Bedingung der flachen Hülle wurden alle Teiltonamplituden auf einen konstanten Wert gesetzt. Dieser entsprach der maximalen Amplitude unter der glockenförmigen Hüllkurve.

Apparatur

Die Durchführung des Experimentes erfolgte unter den in Abschnitt 6.1 auf Seite 73 dargestellten technischen und räumlichen Bedingungen.

Durchführung

Vor Beginn des Experimentes wurde die Hörfähigkeit der Teilnehmer mit zwei audiometrischen Vortests überprüft (siehe Abschnitt 6.1 auf Seite 73). Nach der schriftlichen Darbietung der Instruktion sowie anschließender mündlicher Rückfrage durch den Versuchsleiter wurde das Experiment gestartet.

Die Durchführung dieses Experimentes ist identisch mit der des in Abschnitt 6.4 beschriebenen Bedingungen Experimentes zum Einfluss der Teiltonanzahl. So wurden auch hier die Reizpaare erst nach einer vorher festgelegten Reihenfolge und dann adaptiv präsentiert.

6.5.3 Ergebnisse

Einen Überblick über die gewonnenen Daten liefert die Abbildung 6.15. Hier sind die relativen Häufigkeiten der Antwort 'Tonhöhe ist aufsteigend' für die vier Bedingungen (2 Hüllenformen und 2 Hüllenpositionen) getrennt nach Startton und Distanz im Uhrzeigersinn abgetragen. Alle individuellen Tonhöhenurteile getrennt nach Bedingungen sind im Anhang 6.9 dargestellt.

Im Vergleich zeigt sich, dass die Wegnahme der glockenförmigen Hüllkurve zu deutlich anderen Antwortmustern führt. Im Gegensatz zu den sonst bei oktav-komplexen Tönen üblichen Antwortmustern, mit sicheren Urteilen bei großen und kleinen Distanzen, zeigen sich auf den beiden rechten Grafiken der Abbildung 6.15 Tonhöhenurteile, welche denen von Sinuston- oder harmonisch-komplexen Tonpaaren ähneln. Charakteristisch ist hierbei der stufenartige Wechsel von auf- und ab-Urteilen entlang der Diagonalen der Startton-Distanz-Matrix (vgl. die Wahrscheinlichkeit für erwartete Tonhöhenurteile auf der Basis des Grundtons in Abb. 6.12). Die tiefsten Frequenzen sind hier C (32.7 Hz) bzw. Fis (46.2 Hz). Paare, welche mit diesen Tönen beginnen, werden konsistent als aufsteigend bewertet.

Interessant ist auch, dass die Sicherheit der Urteile zu den Ecken der Matrix (Distanz/Startton, 1/1 und 11/12) hin abnimmt und dieses Muster nicht bei den üblichen Paarvergleichen mit oktav-komplexen Tönen zu finden ist. In beiden Bedingungen mit Hüllkurve (linke Abbildungen) wird deutlich, dass die Tonpaare relativ unabhängig vom jeweiligen Startton beantwortet werden.

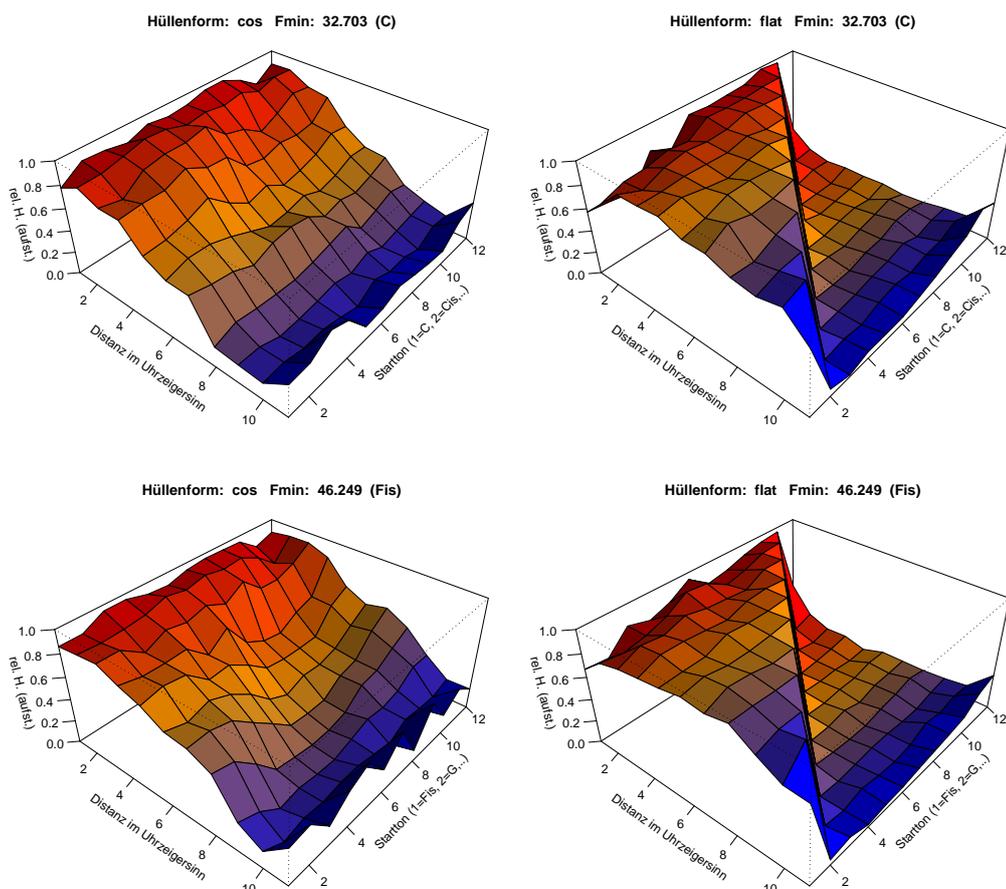


Abbildung 6.15. Die relativen Häufigkeiten der Antwort 'Tonpaar ist aufsteigend' je Startton und Distanz im Uhrzeigersinn abgetragen für die vier Bedingungen (2 Hüllenformen und 2 Hüllenpositionen). Die Ziffern 1 bis 12 an der Starttonachse entsprechen den Tonklassen, ausgehend von der kleinsten Frequenz f_{min} , so dass der Startton 1 der Chromaklasse C bei $f_{\text{min}}=32.7$ (oben) und Fis bei $f_{\text{min}}=46.2$ (unten) entspricht.

Allein die Distanz im Uhrzeigersinn zwischen den Teiltönen der zu vergleichenden Töne entscheidet, ob ein Paar auf- oder absteigend gehört wird. Im mittleren Distanzbereich um das Tritonusintervall lassen sich jedoch auch in dieser Bedingung variierende Antworthäufigkeiten und somit ein Einfluss des Starttons (Tritonusparadox) aufzeigen.

Um zu ermitteln, in welchem Ausmaß ein Hörer welche Urteilsstrategie (Distanz- und Fundamentalurteile) wählt, wurde der in Abschnitt 6.4.3 eingeführte FC-Index berechnet. Der FC-Index gibt Auskunft über die Ähnlichkeit der empirischen Häufigkeitsverteilung zu zwei theoretischen Antwortverteilungen und somit einen Hinweis auf die zugrundeliegende Hör- bzw. Urteilsstrategie. Der FCI wird auf der Basis der Daten eines vollständigen Paarvergleichs,

Tabelle 6.8
Die FC-Indizes für alle Personen, abgetragen für die vier Bedingungen sowie deren Mittelwert.

VP	Bedingungen				Mittel
	C - env	Fis - env	C - flat	Fis - flat	
AP	0.52	0.33	-0.90	-0.87	-0.23
JM	0.25	0.20	-0.96	-0.94	-0.37
MA	0.45	0.75	-0.50	-1.00	-0.07
MD	0.57	0.57	-0.97	-0.98	-0.20
MK	-0.03	0.03	-0.97	-0.93	-0.47
MO	0.16	-0.18	-0.85	-0.84	-0.42
RZ	0.30	0.91	-0.85	-0.96	-0.15
SB	0.82	0.30	-0.40	-0.72	-0.01
SK	0.34	-0.14	-0.88	-0.76	-0.36
SO	0.64	0.15	-0.83	-0.70	-0.19
Mittel	0.40	0.29	-0.81	-0.87	-0.25

der Startton-Distanz-Matrix, berechnet und nimmt Werte im Intervall $[-1,1]$ an. Wobei ein FCI von -1 einem perfekten Fundamentalurteil entspricht, wie es z.B. bei Sinustonpaaren zu finden ist. Ein FCI von 1 entspricht einem perfekten Distanzurteil.

In der Tabelle 6.8 sind die FC-Indizes aller Versuchspersonen getrennt nach Bedingungen aufgelistet. Auch hier zeigen sich deutlich die Unterschiede zwischen den Bedingungen des Faktors Hüllenform. Bei vorhandener Hüllkurve (Bedingung: env) liegt der FC-Index mit Ausnahme von drei Hörern (MK, MO, SK) durchweg im positiven Bereich ($M=0.35$, $SD=0.30$). Bei Präsentation der Reize mit identischer Amplitude (Bedingung: flat) zeigt sich, dass alle Kennwerte bei einem Mittel von -0.84 ($SD=0.16$) deutlich im negativen Bereich liegen.

Mittelt man die FC-Indizes beider Hüllenformen, ergibt sich ein FCI von -0.25. Es werden somit insgesamt mehr Fundamentalurteile als Distanzurteile abgegeben.

Parameterschätzung

Wie bereits im vorangehenden Abschnitt 6.4 beschrieben, wurde auch bei diesem Experiment das P4-Modell an die Daten angepasst, um anhand der geschätzten Parameter die einzelnen Einflussfaktoren genauer beschreiben zu

können. Hierzu wurde mittels Kleinst-Quadrat-Schätzung für jede Versuchsperson und jede Bedingung die beste Anpassung an das Modell ermittelt.

Nimmt man eine Irrtumswahrscheinlichkeit von $p < 0.2$ als Kriterium für eine hinreichende Passung an, so wurden von den 40 Datensätzen (10 Personen $\times 2 \times 2$ Bedingungen) bei nur 6 eine erfolgreiche Anpassung erzielt. Die 6 Schätzungen entfielen auf Datensätze der Bedingungen mit Hüllkurve, kein Datensatz mit flacher Hülle konnte hinreichend angepasst werden.

Zur weiteren statistischen Auswertung wurde nur der Parameter Phase verwendet. Wie in Abschnitt 6.4 dargestellt eignet sich dieser als Lokationsschätzer für die am höchsten wahrgenommene Tonklasse (peak-pitch-class). Auch ist dieser unabhängig von den übrigen drei Parametern.

Im Folgenden soll untersucht werden, ob der Einfluss der Position der Hülle in der Bedingung cos einen Einfluss auf die Tonhöhenwahrnehmung bei oktavkomplexen Tönen ausübt. Von Interesse sind hierbei die Antworten auf die Tonpaare im Tritonusintervall und speziell die dort am höchsten bewertete Tonklasse (peak-pitch-class).

Die individuell am höchsten wahrgenommenen Tonklassen wurden für die beiden Positionen $f_{\min} = 32.7$ (C) und $f_{\min} = 46.2$ (Fis) aus den geschätzten Phasenwerten des P4-Modells berechnet. Dafür wurden die ermittelten Werte mittels der Funktion 6.10 transformiert. Die so erhaltenen Werte (P_{ppc}) im Intervall $[0,12[$ sind individuelle Schätzungen für die am höchsten wahrgenommene Tonklasse.

Da die P_{ppc} -Werte der Hüllen bei C und bei Fis um eine halbe Oktave versetzt sind (vgl. Abb. 6.15) und somit ein Wert von 0 jeweils mit der Tonklasse bei f_{\min} korrespondiert, wurden die Werte so transformiert, dass C=0, Cis=1,...,B=11 entspricht.

Für die bei C positionierte Hülle erhält man einen zirkulären Mittelwert \bar{P}_{ppc} von 7.83, was der Tonklasse Gis entspricht. Der mittlere zirkuläre \bar{P}_{ppc} bei der bei Fis positionierten Hülle liegt bei 0.07 und entspricht somit der Tonklasse C. In Abbildung 6.16 sind die Ergebnisse der Modellierung der höchsten Tonklasse den empirischen relativen Häufigkeiten des Paarvergleichs von Tritonusintervallen gegenübergestellt. Die Abbildung zeigt auch, dass die individuell höchsten Tonklassen eine Streuung über jeweils etwa die Hälfte des Tonklassenkreises aufweisen, was sich in einer zirkulären Varianz von 1.43 Halbtonschritten äußert. Der kleinste Abstand zwischen den mittleren Höchsttonklassen der C- und der Fis-Bedingung misst 4.3 Halbtonschritte. Um zu überprüfen, ob die Unterschiede zwischen den Mittelwerten der geschätzten Höchsttonklassen der beiden Bedingungen statistisch bedeutsam sind, wurde eine ANOVA für

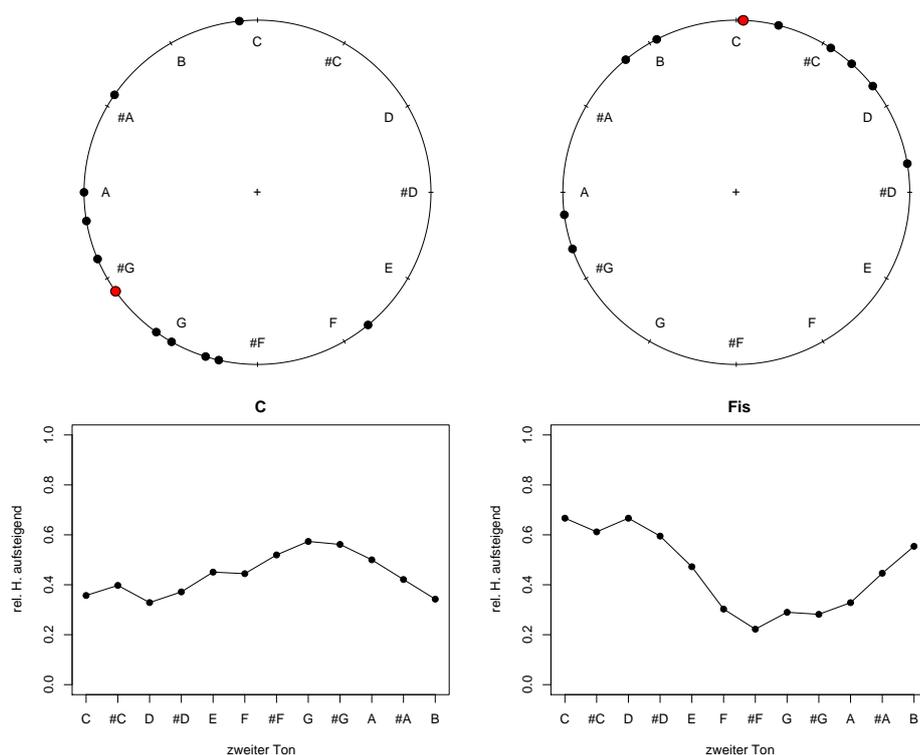


Abbildung 6.16. Einfluss der Hüllenposition auf die Tonhöhenwahrnehmung beim Tritonusintervall für die Bedingung C (links) und Fis (rechts). Oben: Die auf dem Tonklassenkreis abgetragenen individuellen P4-Modell-Schätzungen der am höchsten wahrgenommenen Tonklasse P_{ppc} aller Versuchspersonen und dem Gruppenmittelwert. Unten: Die relativen Häufigkeiten der Antwort „aufsteigend“ abgetragen für den zweiten Ton aller Tonpaare im Tritonusintervall. Deutlich zu erkennen ist, dass der Mittelwert der Parameterschätzung gut mit der maximalen Urteilshäufigkeit beim Paarvergleich korrespondiert.

zirkuläre Daten gerechnet. Es zeigt sich, dass die Position der Hülle einen substantiellen Effekt auf die Wahrnehmung der Tonhöhe bei Paarvergleichen im Tritonusintervall ausübt ($F_{1,19} = 22.84, p < 0.001$).

Die Antwortmuster der Bedingung mit flacher Hülle erscheinen für eine Analyse hinsichtlich des Tritonusparadoxons bzw. der peak-pitch-class ungeeignet, da Tonhöhenurteile fast ausschließlich auf der Basis der tiefsten Tonklasse (C bzw. Fis) gefällt werden. Die höchste Tonklasse ist hier immer die mit der größten fundamentalen Frequenz.

6.5.4 Diskussion

In der vorgestellten Untersuchung wurde in einem vollständigen Paarvergleichsexperiment die relative Tonhöhe von Paaren oktav-komplexer Töne erhoben. Dabei wurde die Position und die Form der Hüllkurve in jeweils zwei Bedingungen variiert. Im Gegensatz zu den Befunden von Pollack (1978) zeigt sich, dass der Hüllkurve von oktav-komplexen Tönen eine essentielle Bedeutung zukommt. Wird diese bei der Berechnung der Töne weggelassen, so dass identische Teiltonamplituden resultieren, so zerfällt der von Shepard (1964) demonstrierte Einfluss des Chromas vollständig. Die Tonhöhe eines Tones bemisst sich, wie sonst bei harmonisch-komplexen Klängen üblich, an der Frequenz der Fundamentalen. Die Töne lassen sich bzgl. der Tonhöhe in eine eindeutige Rangreihe bringen, die Zirkularität in den Urteilen entfällt. Noch stärker als bei der Variation der Anzahl der Teiltöne (vgl. Experiment 6.4) wird hier ein Wechsel der Urteilsstrategie deutlich.

Dies stellt die Ergebnisse einschlägiger Studien in Frage. So verwenden z.B. Parncutt und Tan (2005) keine Hülle bei der Konstruktion oktav-komplexer Töne und argumentieren, dass die anschließende Verstärkung und der Frequenzverlauf der individuellen Hörkurve diesen Effekt erzielen würden.

In Bezug zur subjektiv höchsten Tonklasse beim Tritonusparadox konnte gezeigt werden, dass die Position der Hülle einen substantiellen Effekt ausübt. Die subjektiv höchste Tonklasse verschiebt sich um 4.3 Halbtöne, wenn die Hülle um 6 Halbtöne verschoben wird. Dies, als auch die hohe interindividuelle Variation, spricht gegen die von Deutsch (1986) postulierte Medien-Hypothese zur einheitlichen Tonklassenpräferenz einer Sprachgemeinschaft. Eine Mittelung der Ergebnisse unter verschiedenen Hüllen, wie es in einschlägigen Studien (Deutsch, 1986; Giangrande, 1998) durchgeführt wurde, erscheint angesichts der aktuellen Befunde nicht zulässig. Im Konsens mit den Ergebnissen stehen z.B. die Befunde von Repp (1997) und Terhardt, Stoll, Schermbach und Parncutt (1986). Eine mögliche Erklärung dieser Ergebnisse liefert jedoch auch die Hypothese zum Einfluss der spektralen Dominanzregion auf das Tritonusparadox nicht (Terhardt, 1989, vgl. 5.1.4).

Dadurch, dass die Urteile bei flacher Hülle den Urteilen auf der Basis der fundamentalen Frequenz gleichen, erscheint die Aussage, dass bei oktav-komplexen Tönen nicht eindeutig ein Grundton bestimmbar ist fragwürdig. Aufgrund der fehlenden Vielfachen der Fundamentalen, mit Ausnahme der Oktavintervalle, lässt sich z.B. durch die Theorie der virtuellen Tonhöhe (subharmonische Koinzidenzdetektion) kein eindeutiger Grundton ermitteln. Fer-

ner nimmt dadurch der Einfluss der spektralen Dominanzregion zu. Die Ergebnisse legen hingegen nahe, dass oktav-komplexe Töne ohne Hülle anhand der fundamentalen Frequenz bzw. einer äquivalenten Größe, in ihrer Tonhöhe bewertet werden können. Diese Tatsache stellt die bisherigen Überlegungen bzw. Grundlagen der Zirkularität der Urteile in Frage. Um sicherzustellen, dass keine Artefakte oder spektrale Hinweisreize, z.B. durch Störfrequenzen übersteuerter Signale, entstanden sind, wurden alle oktav-komplexen Töne mittels Fourieranalyse nochmals überprüft. Im FFT-Spektrum zeigten sich jeweils nur die erwarteten sechs Teiltöne je oktav-komplexem Ton. Zur Klärung dieser Fragestellung würden sich weitere Experimente anbieten, z.B. ein Paarvergleich mit Variation der fundamentalen Frequenz von einzelnen Tönen eines Sets.

Als Grund für die mäßige Anpassung des P4-Modells an die Daten könne u.a. die beiden Urteilsstrategien angeführt werden. Das P4-Modell wurde auf der Grundlage empirischer Befunde entwickelt und ist nicht bzw. nur bedingt für Fundamental-Urteile, wie sie hier bei flacher Hülle auftreten, geeignet. Dies spricht einerseits gegen die Generalisierbarkeit des P4-Modells. Andererseits kann dies auch als Hinweis auf zwei unterschiedliche Hör- bzw. Urteilsstrategien gedeutet werden.

Wie auch beim vorangehenden Paarvergleich lässt sich die Konfundierung von Lautstärke mit Hüllenform, wenn auch phänomenologisch der Unterschied der Gesamtlautstärken kaum merklich erscheint, nicht ausräumen. Auch hier wären weitere Experimente zur Kontrolle dieser Faktoren nötig.

6.6 Dichotische Darbietung von oktav-komplexen Tönen

Werden auf beiden Ohren gleichzeitig unterschiedliche Reize dargeboten, so spricht man von dichotischer Darbietung von Tönen. Im Gegensatz dazu steht die diotische Darbietung, bei der die Stimuli für beide Ohren identisch sind. Werden, wie im vorliegenden Fall, oktav-komplexe Töne so aufgeteilt, dass bestimmte Komponenten nur auf dem einen Ohr, andere nur auf dem anderen Ohr dargeboten werden, so ergeben sich eine Reihe interessanter Fragestellungen.

Mit einem Paarvergleichsexperiment soll geklärt werden, inwieweit die Distanz zwischen den Teiltönen einen Einfluss auf die Stabilität von Tönhöhenurteilen ausübt. Weiterhin ermöglicht die Versuchsanordnung zu überprüfen, ob der Tonhöhenvergleich von oktav-komplexen Tönen zentralen oder peripheren Mechanismen unterliegt. In Abschnitt 4.5 auf Seite 35 wurde bereits auf einige Theorien zur binauralen Wahrnehmung eingegangen.

6.6.1 Fragestellung

In seiner Theorie geht Shepard (1964) davon aus, dass die Distanz zwischen zwei Tonklassen auf dem Tonklassenkreis darüber entscheidet, welches Tonpaar als höher bzw. tiefer empfunden wird. Ist die Distanz vom ersten zum zweiten Ton im Uhrzeigersinn kürzer als gegen den Uhrzeigersinn, wird das Tonpaar als aufsteigend empfunden. Der zweite Ton dieses Paares wird folglich als höher empfunden (vgl. Abschnitt 5).

Konstruiert man nun, ausgehend von diesem Distanzkriterium, dichotische Reize, welche monaural z.B. aufsteigend empfunden werden, binaural hingegen absteigen und vice versa, so ließe sich entscheiden, ob zentrale oder periphere Mechanismen beteiligt sind. Nicht alle Kombinationen von Teiltönen des ersten und zweiten oktav-komplexen Tones erfüllen diese Bedingung. In Abbildung 6.17 sind Reizpaare dargestellt, welche monaural absteigend und binaural als aufsteigend empfunden werden sollten.

Folgende Hypothesen sollen überprüft werden.

1. Bei Gültigkeit des Distanzkriteriums sollte sich beim Wechsel von der diotischen zur dichotischen Reizpräsentation auch das Tonhöhenurteil umkehren.

2. Die Sicherheit der Urteile sollte von der Distanz zwischen den Teiltönen der zu vergleichenden Tonklassen abhängen.
3. Behält das Distanzkriterium des kürzesten Weges auch bei Intervallen über die Oktavgrenze hinaus seine Gültigkeit, so kann überprüft werden, welche bzw. wie viele Mechanismen bei der Bildung der Tonhöhe beteiligt sind, ein zentraler oder zwei periphere?

Letztere Hypothese lässt sich wie folgt erläutern. Fügt man die beiden Teilspektren des linken und des rechten Ohres zusammen, so erhält man wieder das Spektrum des Originaltonpaares. Da jedes Teilspektrum für sich betrachtet ein zum Originalton umgekehrtes Tonhöhenurteil auslösen sollte, lässt sich überprüfen, an welchem Punkt der Informationsverarbeitung die Tonhöhe aus dem Reizspektrum gebildet wird. Ein zentraler Mechanismus sollte, da er auf einem vollständigen Spektrum operiert, ein zum Originaltonpaar identisches Urteil liefern. Wird die Tonhöhe jedoch auf einer früheren Stufe des Signalweges gebildet, welche vor der Fusion der Teilspektren operiert (peripher), so sollte ein zum Originaltonpaar komplementäres Urteil zu erwarten sein.

Einschlägige Untersuchungen zur binauralen Wahrnehmung von oktavkomplexen Tönen lassen sich kaum finden. Berthoud, Jarchow und Long (2004) führten ein Experiment mit dichotischen Shepardreizen durch, dabei teilten diese die Komponenten eines Shepardtones so auf die beiden Ohren auf, dass z.B. die oberen fünf Oktaven dem einen Ohr und die unteren fünf Oktaven dem anderen Ohr dargeboten wurden. Ziel des Experimentes war die Überprüfung der von Deutsch (1974) postulierten Suppressionshypothese bei der Oktaven-Illusion. Die Autoren fanden individuell unterschiedliche Antworten, welche sowohl für die Fusion als auch für die Suppression sprachen. Zusammenfassend wird berichtet, dass bei Intervallen der Distanzen 1, 2 und 3 eher eine Fusion stattfindet, hingegen bei der Distanz 4 eine Suppression. Warum neuronale Mechanismen vom dargebotenen Intervall abhängen, ist nicht nachvollziehbar. Auch erscheint—wie im Folgenden gezeigt wird—die einfache Aufteilung (gerade-rechts/ungerade-links) der Komponenten nicht hinreichend, um die Fragestellung zu beantworten.

Regeln zur Aufteilung der Komponenten

Im Folgenden soll gezeigt werden, welche Bedingungen zu einer Umkehr der relativen Tonhöhe führen. Die Formalisierung beruht auf einer Arbeit von Lukas (2004), die Fragestellung leitet sich aus Vorexperimenten ab (Krüger, 2006; Krüger & Lukas, 2005).

Gegeben sei ein Paar oktav-komplexer Töne T_1 und T_2 , wobei T_1 die zuerst dargebotene Tonklasse sei. Die n Teiltöne jedes Tones seien mit $t_{11}, t_{12}, \dots, t_{1n}$ bzw. $t_{21}, t_{22}, \dots, t_{2n}$ bezeichnet. Wie müssen nun die Komponenten der beiden oktav-komplexen Töne T_1 und T_2 auf das linke und rechte Ohr aufgeteilt werden, damit sich die kürzeste Distanz zwischen den Komponenten t_1 und t_2 umkehrt und—der Theorie folgend—somit auch die relative Tonhöhe von aufsteigend zu absteigend wechselt und umgekehrt?

Man betrachte dazu die Distanz zwischen den Komponenten der Tonklassen T_1 und T_2 in Richtung der aufsteigenden Frequenz oder anders ausgedrückt der Distanz im Uhrzeigersinn⁵. Die Distanz im Uhrzeigersinn in Halbtonschritten von t_1 zur nächst höheren Komponente t_2 der Klasse T_2 sei mit d bezeichnet. Die umgekehrte Distanz von t_2 nach t_1 sei mit d' bezeichnet, da eine Oktave 12 Halbtöne umfasst gilt $d' = 12 - d$. Beachtet man nur die Variation gerade und ungerade, so ergeben sich folgende zwei Möglichkeiten.

1. Die geraden Teiltöne der beiden Tonklassen T_1 und T_2 werden stets einer Seite (rechts oder links) zugeordnet und die ungeraden Teiltöne der anderen Seite.
2. Die geraden Teiltöne der Tonklasse T_1 werden stets der anderen Seite zugeordnet, als die geraden Teiltöne der Tonklasse T_2 . Mit den ungeraden Teiltönen wird genauso verfahren, wobei die geraden und ungeraden Komponenten einer Tonklasse niemals der gleichen Seite zugeordnet werden.

Alle möglichen Kombinationen sind in Tabelle 6.9 auf der nächsten Seite zusammengefasst. Bei einer Darstellung aller $2n - 1$ Distanzen zwischen den Teiltönen zweier vollständiger oktav-komplexer Töne erhält man folgende Reihe

$$| d | d' | d | d' | d | d' | \dots | d | d' | .$$

Bei einer Aufteilung nach *Kombination A* (siehe Tab. 6.9 auf der nächsten Seite) erhält man in jedem einzelnen Ohr die Distanzen

$$| d | d' + d + d' | d | d' + d + d' | \dots | d | d' + d + d' | .$$

⁵Die Verwendung des Begriffes 'Uhrzeigersinn' scheint hier irreführend und bedarf einer Erklärung, da eine Richtung auf einem Kreis impliziert wird, welcher nicht gegeben ist. Oktav-komplexe Töne lassen sich auf dem sog. Tonklassenkreis darstellen (vgl. Kapitel 5 auf Seite 40). Die hier gewählte Spektraldarstellung entspricht für den Bereich einer Oktave der eines Kreises. Der Begriff 'Distanz im Uhrzeigersinn' ist somit äquivalent der Distanz in Richtung aufsteigender Frequenz.

Tabelle 6.9

Mögliche Kombinationen für die Aufteilung der Teiltöne

Kombination	Variation	Seite		Teiltöne t_1	Teiltöne t_2
A	A1	R	→	ungerade	ungerade
		L	→	gerade	gerade
	A2	R	→	gerade	gerade
		L	→	ungerade	ungerade
B	B1	R	→	ungerade	gerade
		L	→	ungerade	gerade
	B2	R	→	gerade	ungerade
		L	→	gerade	ungerade

Geht man nun davon aus, dass auch bei diesen „halben“ Spektren die Richtung der Tonhöhenänderung nur von d abhängt, so sollte sich im Fall $d < d'$ die Richtung nicht ändern, da auch $d < d' + d + d'$ gilt. Die kürzeste Distanz bleibt im einzelnen Ohr (monaural) als auch im kombinierten Reiz (binaural) erhalten. Ist $d > d'$ so gilt jedoch immer noch $d < d' + d + d'$. Die kürzere Distanz kehrt sich um und die Tonhöhenänderung sollte in diesem Fall die Richtung ändern (vgl. Abb. 6.17 d).

Wird das Spektrum des Originaltones nach der *Kombination B* aufgeteilt, erhält man je Ohr die folgenden Distanzen

$$| d + d' + d | d' | d + d' + d | d' | \dots | d + d' + d | d' | .$$

Ist $d < d'$ gilt $d' < d + d' + d$ (vgl. Abb. 6.17 e). Im Gegensatz zum Originalspektrum ist hier die kürzere Distanz in Uhrzeigerrichtung vom zweiten zum ersten Ton. Die Kombination B liefert im Falle $d < d'$ die gesuchte Umkehr der Tonhöhenänderung. Für den Fall $d > d'$ ändert die Kombination B nicht die kürzere Distanz und folglich auch nicht die Tonhöhenänderung.

Die Regeln zur Aufteilung der spektralen Komponenten der beiden Tonklassen sind somit fast geklärt. Wenn $d < d'$, dann wähle Kombination B und wenn $d > d'$ dann wähle Kombination A.

Nun ist noch zu berücksichtigen, dass bei einem vollständigen Paarvergleich alle möglichen Reizkombinationen Verwendung finden und somit auch Fälle, bei denen die kleinste Frequenz der zweiten Tonklasse kleiner ist als die kleinste Frequenz der ersten Tonklasse ($t_{21} < t_{11}$). Für die Regel der Zuordnung stellt

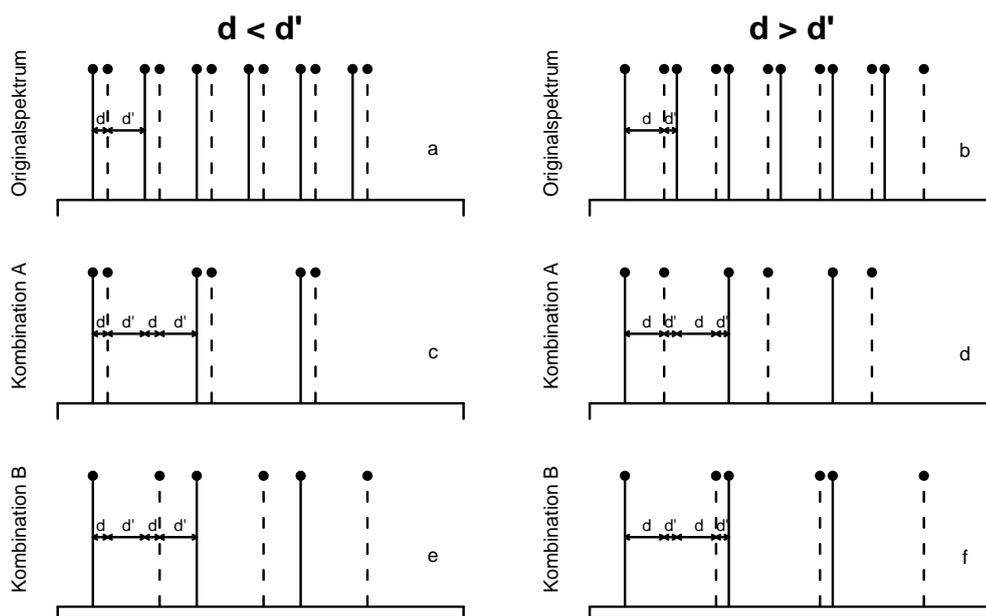


Abbildung 6.17. Schematische Darstellung der Aufteilung der Teiltöne eines Tonpaares (zweiter Ton gestrichelt) für ein Ohr. Oben: Vollständige Spektren eines aufsteigenden (a) und eines absteigenden Tonpaares (b). Mitte: Die nach Kombination A (gleiche Position gleiche Seite) aufgeteilten Spektren. Unten: Die nach Kombination B (gleiche Position ungleiche Seite) aufgeteilten Spektren. Ein Wechsel der kürzesten Distanz zwischen den Teiltönen der ersten und zweiten Tonklasse wird von Abb. a zu e sowie von Abb. b zu d erreicht.

dies kein Problem. Da jedoch d als Distanz vom ersten zum zweiten Ton *im Uhrzeigersinn* bzw. in Richtung aufsteigender Frequenz definiert wurde, erhält man nun als erste Distanz den Abstand von t_{11} nach t_{22} , da ja t_{21} ausgelassen wurde. Was heißt das für die obige Regel? Wenn $t_{21} < t_{11}$ gilt, vertauschen sich d und d' , folglich muss auch Kombination A und Kombination B vertauscht werden. Die Regel kehrt sich in diesem Fall um (siehe Tab. 6.10 auf der nächsten Seite).

Bisher wurden nur die Distanzen $d < 6$ und $d > 6$ betrachtet. Welche Aufteilung soll jedoch im Fall des Tritonus, also bei $d = d' = 6$ gewählt werden? Wie in Abschnitt 5.1.4 auf Seite 46 bereits erläutert wurde, erhöht sich die Varianz der Urteile im Paarvergleich mit zunehmender Nähe zum Tritonusintervall. Abhängig vom Startton werden manche Paare als aufsteigend andere als absteigend wahrgenommen. Mit der Aufteilung der Teiltöne auf beide Ohren sollte das Tritonusparadoxon zugunsten sicherer Urteile verschwinden, da es je Ohr nun wieder eine kürzeste Distanz gibt. Gilt $t_{11} < t_{21}$, so erhält man

Tabelle 6.10

Regeln zur Aufteilung nach den Kombinationen in Tab. 6.9

Distanz von T_1 nach T_2	kleinste Frequenz	Bedingung in zur Aufteilung von T_1 und T_2
$d < 6$	t_{11}	$T_1 \rightarrow A, T_2 \rightarrow B$ oder $T_1 \rightarrow B, T_2 \rightarrow A$
$d < 6$	t_{21}	$T_1 \rightarrow A, T_2 \rightarrow A$ oder $T_1 \rightarrow B, T_2 \rightarrow B$
$d > 6$	t_{11}	$T_1 \rightarrow A, T_2 \rightarrow A$ oder $T_1 \rightarrow B, T_2 \rightarrow B$
$d > 6$	t_{21}	$T_1 \rightarrow A, T_2 \rightarrow B$ oder $T_1 \rightarrow B, T_2 \rightarrow A$

bei Verwendung der Kombination A stets $d < d' + d + d'$ und somit aufsteigende Urteile je Ohr. Aufteilung B führt folglich zu absteigenden Urteilen. Gilt hingegen $t_{21} < t_{11}$, so führt Kombination A zu absteigenden und Kombination B zu aufsteigenden Urteilen.

6.6.2 Methode

Versuchspersonen

Am Paarvergleichsexperiment nahmen acht Hörer im Alter von 20 bis 53 Jahren ($M = 28.7$, $SD = 11.2$) an der halbstündigen Untersuchung teil. Die vier weiblichen und vier männlichen Versuchsteilnehmer waren, bis auf eine Studentin der Germanistik, alle Studierende oder Mitarbeiter des Institutes für Psychologie Halle. Alle Versuchspersonen nahmen freiwillig am Experiment teil. Drei Versuchsteilnehmer erhielten für ihre Teilnahme eine halbe Versuchspersonenstunde gutgeschrieben. Durch zwei audiometrische Voruntersuchungen (siehe Abschnitt 6.1) wurde die Hörleistung aller Versuchsteilnehmer als normal eingestuft.

Aufgabe der Versuchsperson

Die Aufgabe der Versuchsperson war, zu entscheiden, ob der zweite Ton eines vorher dargebotenen Tonpaares höher oder tiefer war als der erste Ton, die Tonhöhe also ansteigt oder abfällt. Das Urteil der Versuchsperson (aufsteigend oder absteigend) wurde durch Tastendruck erhoben, wobei die Tastaturbelegung kongruent mit den zu beurteilenden Empfindungen war, d.h. bei aufsteigend empfundenem Perzept sollte die Pfeiltaste-Hoch gedrückt werden und bei absteigendem Perzept die Pfeiltaste-Runter.

Versuchsplan

Zur Überprüfung der oben genannten Hypothese wurden drei Sets oktavkomplexer Töne gebildet, *original*, *split* und *control*. Die Bedingungen sind in Abbildung 6.18 auf der nächsten Seite am Beispieltonpaar C-Dis verdeutlicht. Die Töne im Set *original* entsprachen den Reizen aus dem Experiment von Shepard (1964). Dabei wurden beiden Ohren identische Reize, bestehend aus allen Teiltönen, dargeboten (diotisch). Ausgehend von diesem Set wurden im Set *split* die Teiltöne nach der obigen Regel so auf beide Ohren verteilt, dass sich—betrachtet man nur ein Ohr—die Distanz zwischen den Komponenten des ersten und des zweiten Tones umkehrt (dichotisch). Der Shepardschen Theorie folgend, sollte sich mit einer Umkehr der Distanz auch das Tonhöhenurteil umkehren. Um kontrollieren zu können, ob die in der Bedingung *split* verwendeten Reize auch zu der erwarteten Umkehr im Tonhöhenurteil führen, wurde die Bedingung *control* eingeführt. Hierbei wurden die Reize auf einer Seite der *split* Bedingung gleichzeitig beiden Ohren dargeboten (diotisch).

Die drei Bedingungen wurden geblockt mit der festen Reihenfolge *original*, *control*, *split* präsentiert. Die Trials innerhalb eines Blocks waren vollständig randomisiert. Die Randomisierung erfolgte zu Beginn eines jeden Experimentes, so dass die Trialabfolge auch zwischen Versuchspersonen variiert wurde. Die Anzahl der Trials je Block setzte sich wie folgt zusammen:

- *original*: 12 Tonklassenpaare \times 2 Richtungen (auf/ab) \times 3 Wiederholungen = 72 Trials
- *split*: 12 Tonklassenpaare \times 2 Richtungen (auf/ab) \times 2 Variationen \times 3 Wiederholungen = 144 Trials
- *control*: 12 Tonklassenpaare \times 2 Richtungen (auf/ab) \times 2 Variationen \times 3 Wiederholungen = 144 Trials

Stimuli

Die Reize der *original*-Bedingung bestanden aus jeweils 10-Teiltönen je Ohr. Da die Teiltöne der Bedingung *split* und *control* auf die beiden Ohren aufgeteilt wurden, bestanden diese Töne aus 5 Teiltönen je Ohr. Alle Teiltönen begannen in Sinusphase. Die zur Gewichtung der Teiltonamplituden verwendete Cosinus-Hülle (vergleiche Funktion 5.1 auf Seite 42) war bei $f_{min} = 16.4$ Hz (C) fixiert. Die Reize der Bedingung *original* wurden mit einer Lautstärke von 65 dBA dargeboten. Für alle Bedingungen wurde dieselbe Hüllkurve verwendet, d.h. die

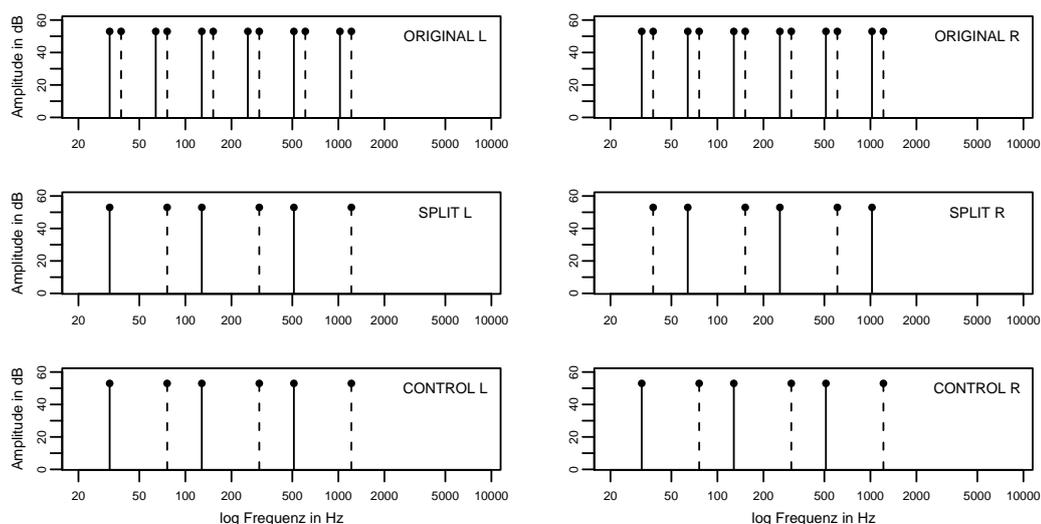


Abbildung 6.18. Die Teiltonspektren der drei experimentellen Bedingungen dargestellt am Tonpaar C (durchgehende Linie) - Dis (gestrichelte Linie). In der Bedingung *original* sind auf beiden Ohren alle Teiltöne vorhanden. In der Bedingung *split* sind die Teiltöne so aufgeteilt, dass sich die Distanz zwischen den Teiltönen je Ohr umkehrt. Das Teiltonspektrum der Bedingung *control* entspricht stets einer Seite der Bedingung *split*. Aus Gründen der Übersichtlichkeit wurde bei der Darstellung auf die Hüllenfunktion verzichtet.

Lautstärke der einzelnen Bedingungen wurde nicht zusätzlich korrigiert. Zur Vermeidung von Störgeräuschen durch abruptes Ein- bzw. Ausklingen wurden alle Reize mit 10 ms langen linearen Ein- bzw. Ausschwingphasen generiert. Inklusive dieser Phasen betrug die Darbietungsdauer eines Tones 400 ms.

Dargeboten wurden ausschließlich Tonklassenpaare im Intervall eines Ganztones (Distanz 2: C-D, Cis-Dis, . . . , Ais-C, B-Cis; Distanz 10: C-Ais, Dis-B, . . . , Ais-Gis, B-A).

Apparatur

Die Darbietung der Reize erfolgte unter den in Abschnitt 6.1 auf Seite 73 dargestellten technischen und räumlichen Bedingungen.

Durchführung

Zur Überprüfung der Hörfähigkeit begann jede Versuchsperson mit zwei audiometrischen Vorexperimenten (siehe Abschnitt 6.1 auf Seite 73). Nach der Erhebung von Alter, Geschlecht und Musikalität begann das eigentliche Experiment

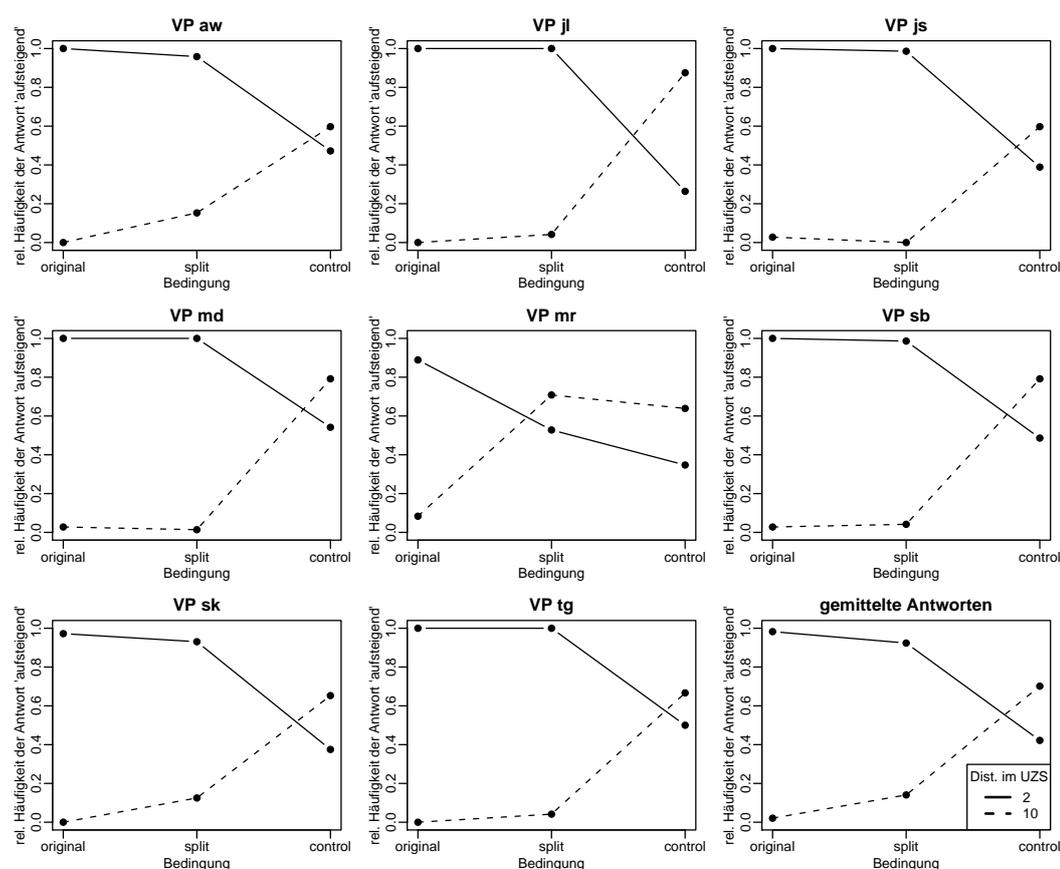


Abbildung 6.19. Die relative Häufigkeit der Antwort 'aufsteigend' aller Versuchsteilnehmer abgetragen je Bedingung für die Distanzen 2 und 10. Die Zusammenfassung (unten, rechts) zeigt die gemittelten Antworten aller Teilnehmer.

mit der Präsentation der Instruktion auf dem Bildschirm. Nach anschließender kurzer mündlicher Instruktion durch den Versuchsleiter wurde das Experiment gestartet. Nach jeweils 180 Trials wurde eine Pause eingefügt, welche die Versuchsteilnehmer nach eigenem Ermessen durch Tastendruck beenden konnten.

Ein Trial setzte sich zusammen aus dem Reizpaar (400ms Ton1, 20ms Pause, 400ms Ton2), gefolgt von der Antwort der Versuchsperson und einer Pause von 2 Sekunden zum nachfolgenden Trial. Zur Beurteilung eines Reizpaares wurde kein Zeitlimit festgelegt.

6.6.3 Ergebnisse

Zur Auswertung wurden die Urteile in den drei Bedingungen getrennt für die beiden Distanzen gemittelt. Die Tonhöhenurteile aller Teilnehmer sowie die Gruppenmittelwerte sind in Abbildung 6.19 zusammengefasst.

Tabelle 6.11

Kennwerte des Chi-Quadrat-Tests für Tonhöhenurteil (auf-/absteigend) und Distanz (2/10) für die drei Bedingungen gemittelt über alle Versuchsteilnehmer.

Bedingung	N	Häufigkeit 'aufsteigend'		χ^2	df	p
		Distanz 2	Distanz 10			
<i>original</i>	288	283	6	529	1	< 0.001
<i>split</i>	576	532	81	706	1	< 0.001
<i>control</i>	576	243	404	90	1	< 0.001

Wie aus den gemittelten Antworten (Abb. 6.19 unten rechts) gut zu erkennen ist, gibt es systematische Unterschiede zwischen den Bedingungen. Die Tonpaare der Bedingung *original* wurden am zuverlässigsten als aufsteigend (Distanz 2) und absteigend (Distanz 10) beurteilt. Mit 5 bzw. 6 Fehlern bei Distanz 2 und 10 wurde diese Bedingung mit jeweils rund 98 Prozent hypotesenkonform beantwortet.

Ebenfalls im Einklang mit der Hypothese, jedoch weniger zuverlässig, wurde die Bedingung *split* bewertet. Mit 92 Prozent der Antwort 'aufsteigend' bei Distanz 2 und 86 Prozent der Antwort 'absteigend' bei Distanz 10 zeigt sich, dass bei der Beurteilung der dichotischen Reize der *split* Bedingung ebenfalls das Distanzkriterium greift.

Im Vergleich zu den beiden genannten Bedingungen, zeigt die Bedingung *control* ein umgekehrtes Antwortmuster. Die Paare der Distanz 2 wurden mit 58 Prozent als 'absteigend' bewertet und die 10er Distanzen hypotesenkonform mit 70 Prozent als 'aufsteigend'.

Zur Beantwortung der Frage, ob die gefundenen Unterschiede statistisch bedeutsam sind, wurde für jede Bedingung ein Chi-Quadrat-Test zum Einfluss der Distanz auf die Tonhöhenurteile berechnet. Die in Tabelle 6.11 dargestellten Ergebnisse zeigen, dass in allen Bedingungen ein hoch signifikanter Einfluss der Distanz zu finden ist. Tonintervalle mit kurzer Distanz im Uhrzeigersinn werden als aufsteigend, lange Distanzen als absteigend wahrgenommen. Ein umgekehrtes Muster findet sich bei der Bedingung *control*. Daraus folgt, dass die Reize der Bedingung *split* bei monauraler Darbietung tatsächlich zu einer Umkehr der Tonhöhenänderung im Vergleich zur Bedingung *original* führen. Weiterhin folgt daraus, dass das von Shepard (1964) vorgeschlagene Distanzkriterium auch bei Intervallen größer als eine Oktave greift. Deutlich wird dies

Tabelle 6.12

Kennwerte des Chi-Quadrat-Tests für Tonhöhenurteil (auf-/absteigend) und Distanz (2/10) für die drei Bedingungen.

Versuchspersonen Initialien	χ_1^2 Werte je Bedingung		
	<i>original</i>	<i>split</i>	<i>control</i>
A.W.	68.6***	91.4***	1.8
J.L.	68.6***	128.7***	52.4***
J.S.	64.3***	136.1***	5.4*
M.D.	64.3***	136.1***	9.0**
M.R.	43.6***	4.2*	11.1***
S.B.	64.3***	124.8***	13.3***
S.K.	64.3***	90.5***	10.0***
T.G.	87.0***	128.7***	3.5

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

besonders bei der Bedingung *control*. Dabei zeigt sich ein Anstieg der Varianz der Urteile über die Bedingungen hinweg ($original < split < control$).

Vergleicht man die in Abbildung 6.19 auf Seite 135 dargestellten individuellen Ergebnisse, so zeigt sich ein einheitliches Urteilmuster. Sieben von acht Versuchsteilnehmer beantworteten die präsentierten Tonpaare hypothesenkongform. Auffällig ist die abweichende Reaktion der Teilnehmerin M.R., welche die erwartete Umkehr im Antwortmuster zusätzlich zur Bedingung *control* auch bei der Bedingung *split* zeigt.

Die in Tabelle 6.12 dargestellten individuellen Kennwerte des Chi-Quadrat-Tests deuten jedoch darauf hin, dass dieser Unterschied vor dem Hintergrund der übrigen Antworten vernachlässigbar erscheint.

Um die Frage zu klären, welchen Einfluss das Chroma auf die Beurteilung der Tonhöhe hat, wurde die relative Häufigkeit der Antwort 'aufsteigend' getrennt für die 12 Tonklassen ausgewertet (siehe Abb. 6.20 auf der nächsten Seite). Mittels eines Chi-Quadrat-Tests wurde überprüft, ob die Antworten über die 12 Tonklassen hinweg einer Gleichverteilung folgen.

Die in Tabelle 6.13 abgetragenen Kennwerte zeigen deutliche Unterschiede zwischen den sechs Bedingungen. Erhöhte Abweichungen lassen sich speziell bei den ersten beiden Tonpaaren der Distanz 10 (C-Ais und Cis-B, vgl. Abb.

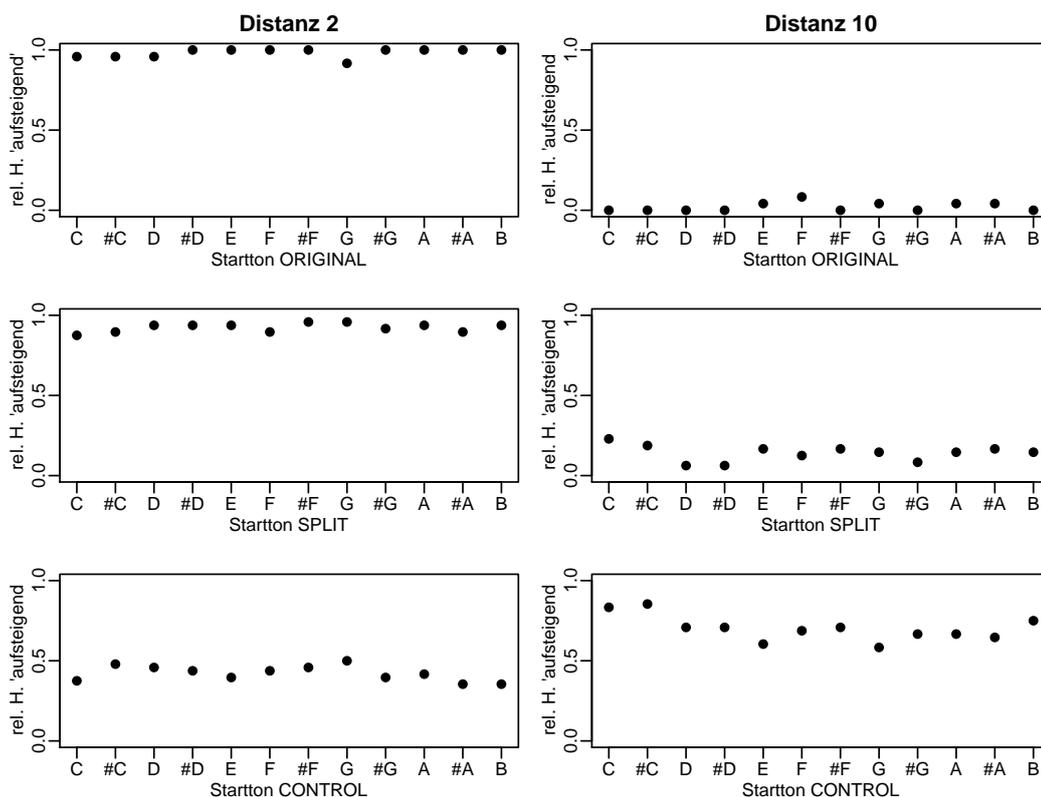


Abbildung 6.20. Die relative Häufigkeit der Antwort 'aufsteigend' aller Versuchsteilnehmer abgetragen je Startton für alle Distanzen und Bedingungen.

6.20) finden. Nimmt man als Ablehnungsbereich eine Wahrscheinlichkeit von $p < 0.2$ an, so zeigt sich ein Einfluss der Tonklasse nur bei Distanz 10 der Bedingung *control*.

6.6.4 Diskussion

In vorliegenden Experiment wurde die relative Tonhöhe von oktav-komplexen Tönen im Intervall eines Ganztones im Paarvergleich erhoben. Dabei wurden die Komponenten jedes Tones in drei Bedingungen so variiert, dass sie vollständig (*original*), aufgeteilt (*split*) und aufgeteilt diotisch (*control*) präsentiert wurden.

Es konnte gezeigt werden, dass das Kriterium der Distanz auch über Oktavgrenzen hinweg Gültigkeit behält. Dies auch, wenn jede zweite Komponente eines Tones entfernt wird, jedoch mit dem Unterschied, dass sich das Tonhöhenurteil nun umkehrt. Hierbei sollte bedacht werden, dass die gezeigte Umkehr des Tonhöhenurteils den oben geschilderten Regeln zur Aufteilung

Tabelle 6.13

Kennwerte des Chi-Quadrat-Tests zum Einfluss der Tonklasse auf die Tonhöhenurteile. Abgetragen sind die Werte für alle Bedingungen getrennt nach Distanzen.

Bedingung	N	χ_{11}^2	p
<i>original</i>			
Distanz 2	288	12.0	0.36
Distanz 10	288	10.2	0.51
<i>split</i>			
Distanz 2	576	5.5	0.90
Distanz 10	576	11.1	0.43
<i>control</i>			
Distanz 2	576	4.9	0.93
Distanz 10	576	16.6	0.12

der Komponenten geschuldet ist. Würden andere Regeln gewählt werden, so würde vermutlich das Distanzkriterium weiterhin zutreffen, eine Umkehr des Tonhöhenurteils würde jedoch nicht zwangsläufig erfolgen.

Die Ergebnisse des Experimentes lassen eine weitere entscheidende Schlussfolgerung zu, nämlich, dass die Wahrnehmung von Tonhöhenänderungen auf zentralen Mechanismen beruhen muss, welche die Informationen aus beiden Ohren zu einem Perzept integrieren. Dies beruht auf der Tatsache, dass zwischen der Bedingung *original* und der Bedingung *split* keine Umkehr im Urteil erfolgte. Betrachtet man nur ein Ohr, so ergibt sich eine Umkehr der Distanzen zwischen den genannten Bedingungen und folglich eine Umkehr des Urteils. Da sich jedoch in der Bedingung *split* die kontralateral dargebotenen Komponenten zu einem vollständigen Ton ergänzen, und die Urteile auch so ausfallen wie in der Bedingung *original*, kann das Tonhöhenurteil nur an einem Punkt der Informationsverarbeitung gebildet werden, an dem ein aus beiden Ohren integriertes Perzept vorliegt. Bei einer Durchführung des Experimentes ohne die Bedingung *control* wäre diese Aussage nicht haltbar, da hier belegt werden konnte, dass das Distanzkriterium auch bei unvollständigen oktav-komplexen Tönen gilt.

Diesem Aspekt der 'Ausdünnung' durch Wegnahme von Teiltönen kann

auch die verminderte Varianz in den Urteilen der Bedingung *control* in Rechnung gestellt werden. Eine mögliche Erklärung könnte wie folgt aussehen. Im vorliegenden Experiment wurden nur Paare im Abstand eines Ganztones verglichen. Da bei der Aufteilung der Komponenten aus kleinen Distanzen eher große Abstände erwachsen und bei großen Distanzen (z.B. Tritonus) nach der Aufteilung eher kleinere Distanzen (vgl. Abb. 6.18), könnte in einem Folgeexperiment überprüft werden, ob die höhere Varianz durch die hier gewählte kleine Distanz 2 bedingt wird. Bei Verwendung größerer Distanzen könnten reliablere Antworten in der Bedingung *control* zu Lasten der Bedingungen *original* erwartet werden.

6.7 Dichotische Darbietung II: Erweiterung um das Tritonusintervall

6.7.1 Fragestellung

Da im vorangehenden Experiment nicht eindeutig geklärt werden konnte, ob die Antworten in der Bedingung *control* stabil dem erwarteten Distanzkriterium folgen, soll eine Erweiterung des ersten Experimentes um die Distanzen 4,8 und 6 zur Klärung beitragen.

Bei Shepard-Tönen aus der Bedingung *original* verringert sich die Sicherheit der Antworten mit der Annäherung an die Distanz 6. Da benachbarte Teiltöne der Bedingung *control* ein Intervall von zwei Oktaven umfassen, tritt hier der umgekehrte Effekt auf. Mit Annäherung an die Distanz 6 im Originaltonpaar verringert sich die, nun zwei Oktaven umfassende, Distanz der Bedingung *control*. Aufgrund der Regel zur Aufteilung der Teiltöne auf beide Ohren erreicht die kürzeste Distanz zwischen Teiltönen t_1 und t_2 hier ein Minimum, folglich sollte diese Distanz mit der höchsten Genauigkeit beantwortet werden. Die gleiche Verringerung tritt natürlich auch monaural bei der Bedingung *split* auf.

6.7.2 Methode

Da dieses Experiment eine Erweiterung des ersten Experimentes darstellt, sind auch die oben beschriebenen methodischen Aspekte wie z.B. Apparatur, Reizdarbietung und Durchführung mit diesem identisch. Im Folgenden werden daher nur die Unterschiede zum ersten Experiment beschrieben.

Versuchspersonen

Für dieses Folgeexperiment konnten die gleichen acht Hörer wie beim ersten Experiment zur dichotischen Wahrnehmung (Abschnitt 6.6) gewonnen werden. Wiederum erhielten drei Versuchsteilnehmer für ihre Teilnahme Versuchspersonenstunden gutgeschrieben. Die übrigen Hörer nahmen unentgeltlich daran teil.

Stimuli und Versuchsplan

Wie auch im vorangehenden Experiment wurden wieder drei Bedingungen getestet, *original*, *split* und *control*. Der wesentliche Unterschied zum ersten Ex-

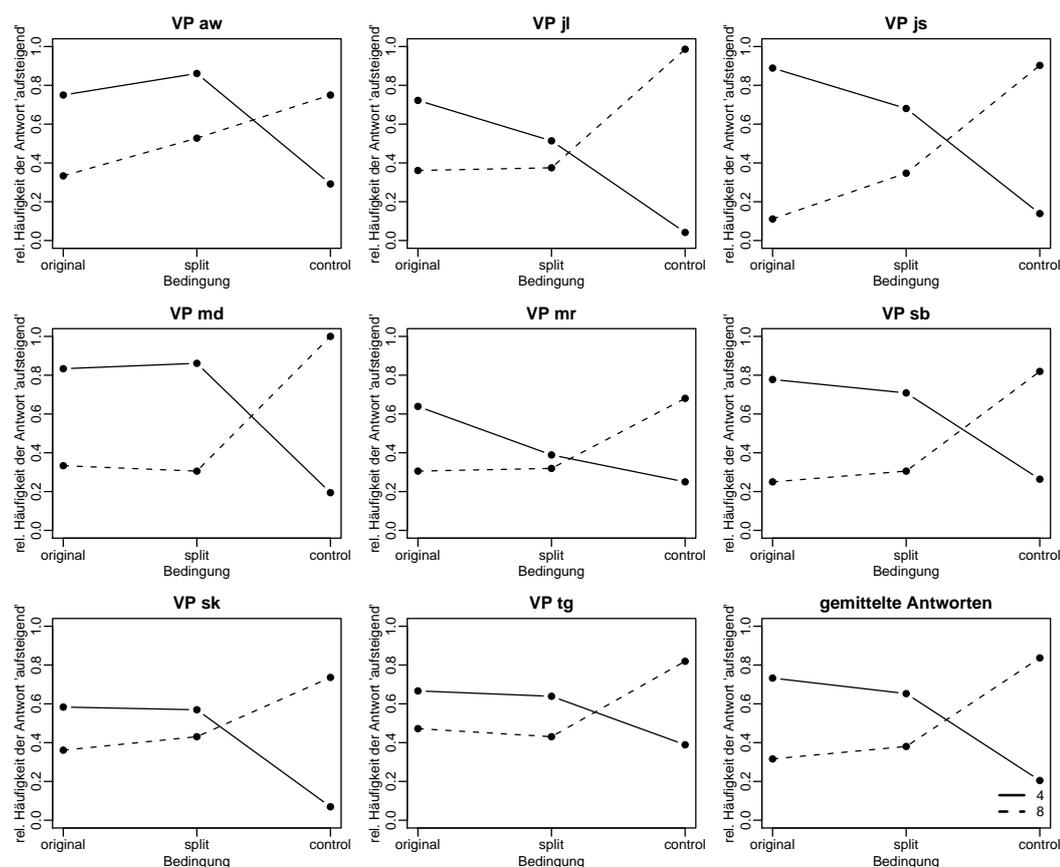


Abbildung 6.21. Die relative Häufigkeit der Antwort 'aufsteigend' aller Versuchsteilnehmer, abgetragen je Bedingung für die Distanzen 4 und 8. Die Zusammenfassung (unten rechts) zeigt die gemittelten Antworten aller Teilnehmer.

periment lag in der Wahl der präsentierten Intervalle je Bedingung. In diesem Folgeexperiment wurden Tonpaare mit den Distanzen 4 und 8 als auch Tonpaare im Intervall von 6 Halbtonschritten präsentiert.

Die Reihenfolge der Präsentation war wiederum innerhalb jedes Bedingungsblocks vollständig randomisiert, d.h. alle Distanzen (4,6,8) wurden gemeinsam innerhalb eines Blockes (*original*, *control*, *split*) dargeboten.

6.7.3 Ergebnisse

Distanz 4 und 8

In Abbildung 6.21 sind die Mittelwerte der Antwort 'aufsteigend' der jeweiligen Bedingungen abgetragen. Die Tonpaare der Bedingung *original* werden bei Distanz 4 in 74 Prozent der Fälle als aufsteigend bewertet und bei Distanz 8 in 28

Tabelle 6.14

Kennwerte des Chi-Quadrat-Tests für Tonhöhenurteil (auf-/absteigend) und Distanz (4/8) für die drei Bedingungen.

Versuchspersonen Initialien	χ^2_1 Werte je Bedingung		
	<i>original</i>	<i>split</i>	<i>control</i>
A.W.	10.9***	17.3***	28.5***
J.L.	8.1**	2.3	124.7***
J.S.	40.5***	14.7***	81.1***
M.D.	16.5***	43.5***	93.8***
M.R.	6.7**	0.5	25.1***
S.B.	18.0***	21.8***	42.5***
S.K.	2.3	2.7	63.7***
T.G.	2.0	5.5*	26.1***

* = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$

Prozent der Fälle als aufsteigend bewertet. Hypothesenkonform wird somit bei den Distanzen 4 und 8 nach den Prinzipien der Nähe zwischen den Teiltönen geantwortet. In der Bedingung *split* werden 67 Prozent aller Paare bei Distanz 4 und 37 Prozent der Paare bei Distanz 8 als aufsteigend wahrgenommen. Da nun vier statt vorher zwei Halbtonschritte zwischen benachbarten Teiltönen liegen, werden die Bedingungen *original* und *split* nun weniger reliabel beurteilt. Für die Bedingung *control* trifft dies nicht zu. Hier verringert sich die Distanz zwischen den Teiltönen im Zwei-Oktav-Intervall, was sich in stabileren Urteilen widerspiegelt. Alle dargebotenen Tonpaare werden in 80 Prozent der Fälle als aufsteigend (Distanz 4) bzw. absteigend (Distanz 8) bewertet. Die Antworten in Abhängigkeit von der Distanz 4 bzw. 8 in den drei Bedingungen *original*, *split* und *control* sind mit χ^2 -Werten von 90.8, 79.6 und 322.7 alle statistisch hoch signifikant voneinander verschieden ($df = 1, p < 0.001$). Urteile bei Distanzen 4 werden somit überzufällig häufig anders bewertet als Urteile im Intervall von 8 Halbtonschritten im Uhrzeigersinn.

Bei allen Versuchsteilnehmern zeigt sich eine Umkehr der empfundenen relativen Tonhöhe in der Bedingung *control* im Vergleich zu den anderen Bedingungen. Die individuellen Differenz zwischen den Distanzen 4 und 8 sind in Tabelle 6.14 anhand der χ^2 -Werte dargestellt. Deutlich hierbei wird die Streu-

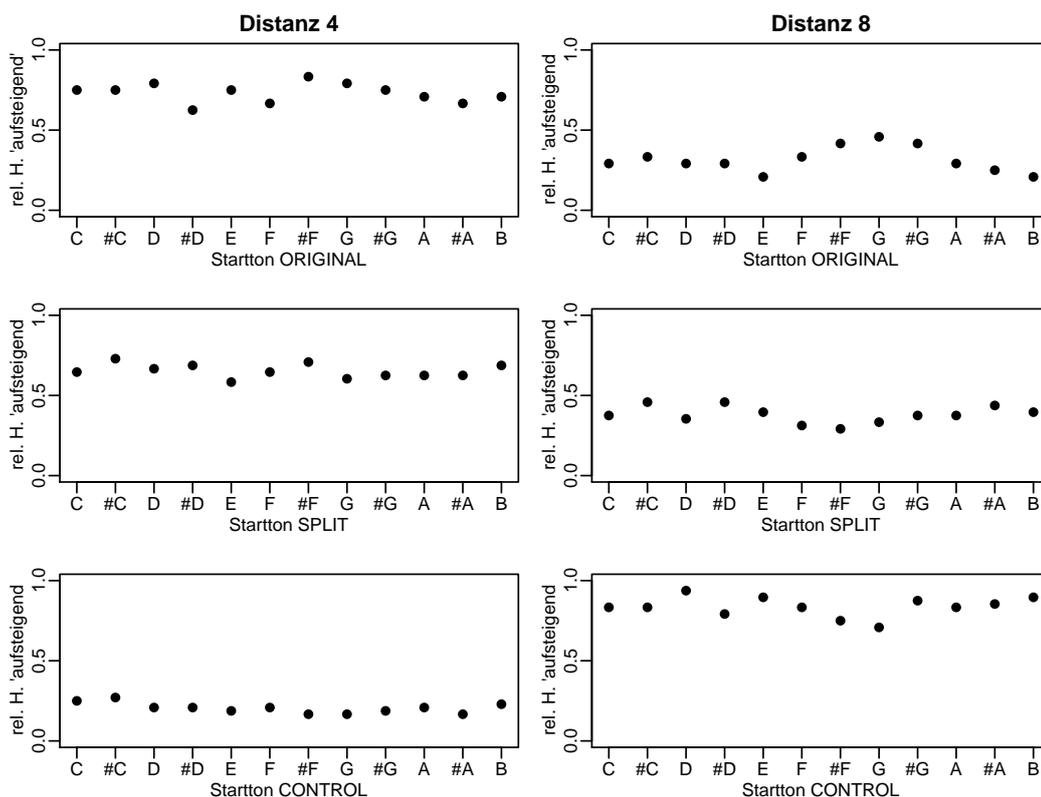


Abbildung 6.22. Die relativen Häufigkeiten der Antwort 'aufsteigend', abgetragen für den ersten Ton eines Paares, Distanz 4 und 8 und die drei Bedingungen.

ung zwischen den Versuchsteilnehmern.

Der Einfluss des Starttones auf die Beurteilung der Tonhöhe ist in Abbildung 6.22 dargestellt. Im Vergleich zu den Urteilen bei Distanz 2 und 10 findet sich hier eine deutlich erhöhte Varianz der Urteile in Abhängigkeit vom ersten Ton eines Paares. Bei Distanz 4 zeigt sich in allen Bedingungen ein Trend zum Urteil 'aufsteigend' im Bereich des Starttones Cis. Bei Distanz 8 ist dieser Trend angedeutet. Auffällig ist hier ein Absinken der Aufwärtsurteile im Bereich der Tonklassen E und F. Zur Überprüfung des Einflusses der Tonklasse auf die Tonhöhenwahrnehmung wurde ein Chi-Quadrat-Test auf Gleichverteilung gerechnet. Die in Tabelle 6.15 dargestellten Kennwerte weisen auf einen geringen aber nicht systematischen Einfluss der Tonklasse auf die Tonhöhenwahrnehmung hin. Auffällig ist eine erhöhter Einfluss bei allen drei Bedingungen der Distanz 8 Intervalle.

Tabelle 6.15

Kennwerte des Chi-Quadrat-Tests zum Einfluss der Tonklasse auf die Tonhöhenurteile. Abgetragen sind die Werte für alle Bedingungen getrennt nach Distanzen.

Bedingung	N	χ_{11}^2	p
<i>original</i>			
Distanz 4	216	7.8	0.72
Distanz 8	216	11.2	0.43
<i>split</i>			
Distanz 4	432	12.1	0.36
Distanz 8	432	13.7	0.25
<i>control</i>			
Distanz 4	432	6.3	0.86
Distanz 8	432	15.6	0.16

Distanz 6

Zur Auswertung der Antworten auf Tonpaare im Tritonusintervall wurden, wie auch bei den Distanzen 2,4,8, und 10, die Daten aller Versuchspersonen für alle drei Bedingungen gemittelt. In Abbildung 6.23 links zeigt sich, dass alle drei Bedingungen mit einer relativen Häufigkeit von etwa 0.5 als aufsteigend beurteilt wurden. Stellt man wie in Abbildung 6.23 rechts die Antworten in Abhängigkeit von der Starttonklasse dar, wird deutlich, dass diese Mittelung nicht gerechtfertigt ist.

In der zu Beginn dieses Kapitels beschriebenen Regel zur Aufteilung der Komponenten auf beide Ohren wird neben der Distanz zwischen den Teiltönen auch die Komponente mit der kleinsten Frequenz berücksichtigt. Da die Distanz hier stets 6 ist, ändert sich die Aufteilung nur, wenn sich die Tonklasse mit der kleinsten Frequenz ändert. Dies ist immer dann der Fall, wenn die Starttonklasse zwischen F und Fis wechselt und die zweite Tonklasse sich somit von B zu C ändert. Da die Komponente mit der kleinsten Frequenz der Tonklasse C kleiner ist als die kleinste Frequenz der Tonklasse B, ändert sich hier die Regel (AB,BA,AA,BB) innerhalb einer Distanz. Um die Ergebnisse korrekt zu interpretieren, wurden die Daten für Distanz 6 separat für die jeweiligen Kombinationen gemittelt (Abb. 6.23 rechts).

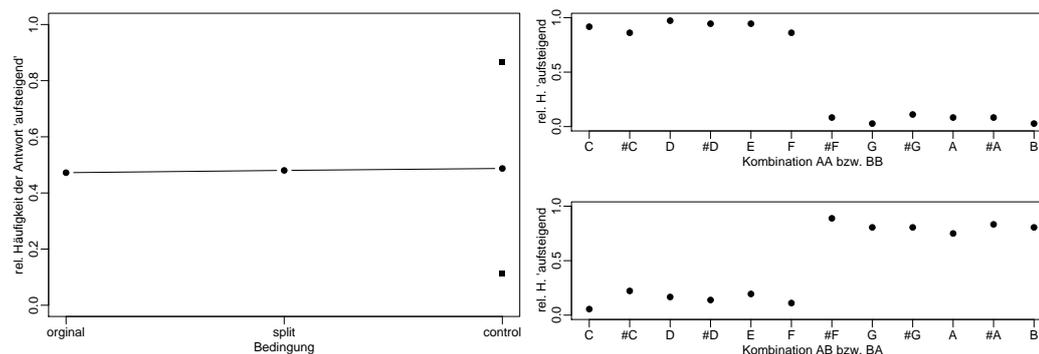


Abbildung 6.23. Links: Die relativen Häufigkeiten der Antwort 'aufsteigend' bei Distanz 6, abgetragen für den ersten Ton eines Paares. Die beiden Rechtecke über und unter dem Mittel in der Bedingung *control* sind die Mittelwerte der nach Kombinationen getrennten Antworten. Rechts: Die Unterschiede der Urteile in Abhängigkeit vom Startton bei der Bedingung *control* beruhen auf der Regel zur Aufteilung der Komponenten auf beide Ohren.

So dargestellt zeigt sich, dass die Tonhöhenurteile in der Bedingung *control* bei Distanz 6 im Vergleich zu den übrigen Bedingungen die geringste Varianz aufweisen. Die vormals ambivalenten Tonpaare im Tritonusintervall werden nach der Aufteilung der Komponenten mit einer relativen Häufigkeit von 0,87 bzw. 0,11 als aufsteigend bewertet.

Eine getrennte Auswertung der Ergebnisse in Abhängigkeit vom Startton für die Bedingung *original* und *split* war nicht notwendig, da sich hier keine Abweichungen in Bezug auf die Regeln der Aufteilung der Komponenten zeigten.

Zusammenfassung aller Distanzen

Fasst man die Befunde aus den beiden Experimenten zusammen, so zeigt sich ein konsistentes Ergebnismuster. In Abbildung 6.24 sind die relativen Häufigkeiten der Antwort 'Tonhöhe ist aufsteigend' für die Distanzen 2, 4, 6, 8 und 10 für die drei Bedingungen abgetragen. Die Tonhöhenurteile in den Bedingungen *original* und *split* können, bis auf eine leicht erhöhte Varianz in der Bedingung *split*, als identisch angesehen werden. Umgekehrtes gilt für die Bedingung *control*. Hier nimmt die Sicherheit der Urteile mit Annäherung an die Distanz 6 zu. Deutlich wird auch die Umkehr der Urteile bei der Bedingung *control* im Vergleich zu den anderen beiden Bedingungen.

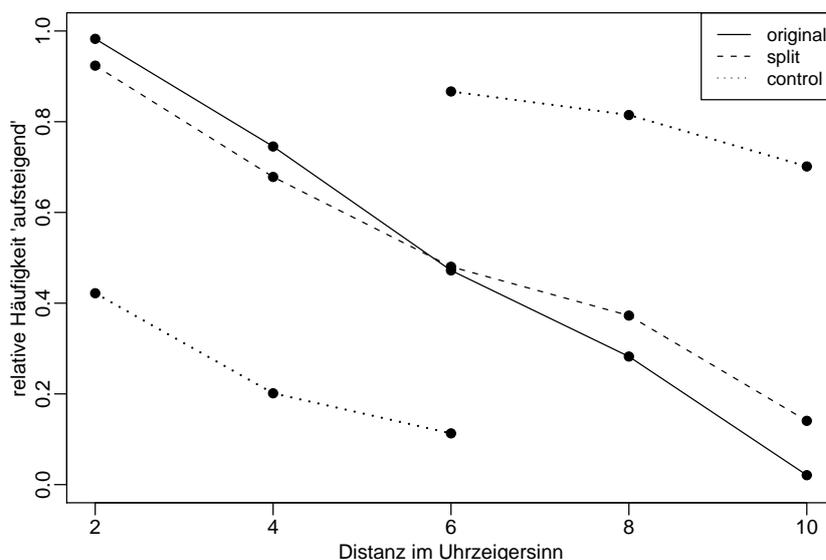


Abbildung 6.24. Die Zusammenfassung zeigt die gemittelten Antworten aller Teilnehmer für die drei Bedingungen. Die beiden Datenpunkte bei Distanz 6 in der Bedingung *control* sind auf die Teilung der Daten bei Startton < 6 bzw. ≥ 6 zurückzuführen.

6.7.4 Diskussion

Aufbauend auf den Befunden des vorhergehenden Experimentes wurden im vorliegenden Paarvergleich zusätzlich Urteile bei den Distanzen 4, 6 und 8 erhoben. Es konnte gezeigt werden, dass die über die drei Bedingungen gefundene erhöhte Varianz (*original* $<$ *split* $<$ *control*) als Effekt der Distanz und der Teiltonanzahl interpretiert werden kann. Je mehr Teiltöne (zentral) vorhanden sind und je geringer der Abstand zwischen den Komponenten des ersten und des zweiten Tones, desto sicherer fallen die Urteile aus.

Weiter wurde deutlich, dass die Wahrnehmung von oktav-komplexen Tönen auch bei Zwei-Oktav-Intervallen dem Gestaltgesetz der Nähe folgt. Die sonst entscheidende Korrelate für Tonhöhenurteile, die fundamentale Frequenz, ist weiterhin zur Erklärung der Tonhöhenurteile bei oktav-komplexen Tönen nicht notwendig. Dies gilt hier insbesondere auch für die 'ausgedünnten' oktav-komplexen Töne der Bedingung *control*. Gleichwohl bleibt an dieser Stelle offen, wie die Varianz—oder anders ausgedrückt: der Verlust an Urteilsstabilität—zu erklären ist, welcher zunimmt, je weniger Komponenten verfügbar sind und je größer die Distanz zwischen den zu vergleichenden Komponenten ist.

Ferner kann argumentiert werden, dass der Entscheidungsprozess, welcher

zur Bildung des Tonhöhenurteils notwendig ist, auf einer Tonhöhenrepräsentation operiert, welche Informationen aus beiden Ohren fusioniert. Die Befunde aus den beiden Experimenten zum dichotischen Hören sprechen somit für die in Abschnitt 4.5 dargestellte Theorie der Fusion von Teilreizen (Arehart & Burns, 1999; Demany & Semal, 1988). Geht man von der Suppressionshypothese (u.a. Deutsch & Roll, 1976) aus, so sollte die Bedingung *split* wie die Bedingung *control* bewertet werden, da, wenn sich nur für ein Ohr entschieden werden würde, sich das Urteil umkehren müsste. Einwände könnten hier sein, dass die lateralen Reize nicht genug konkurrieren oder dass der Entscheidungsprozess im Paarvergleich eine übergeordnete Rolle spielt.

Hinsichtlich die Frage des Ortes der Verarbeitung von Tonhöhe finden sich, übereinstimmend mit den Ergebnissen von Houtsma und Goldstein (1971), eindeutige Belege für einen zentralen Mechanismus. Hierbei kann jedoch der Einfluss der peripheren Teilsysteme nicht ausgeklammert werden.

Im Konsens mit Pollack (1978) entsteht Zirkularität auch dann, wenn nicht alle Oktavkomponenten eines oktav-komplexen Tones vorhanden sind. Die von ihm formulierte Voraussetzung, es müssen hinreichend Zwei-Komponenten-Vergleiche innerhalb einer halben Oktave vorhanden sein, kann erweitert werden. Vergleiche finden auch bei Intervallen größer als eine halbe Oktave statt. Zur Frage, welche Anzahl dabei als hinreichend bezeichnet werden kann (Pollack schlägt $N=3$ vor), kann an dieser Stelle keine Aussage getroffen werden.

6.8 Zusammenfassung und Diskussion der experimentellen Befunde

Im folgenden Abschnitt sollen die Querverbindungen zwischen den Experimenten gezogen werden. So sollen die weitgehend isoliert behandelten experimentellen Befunde inhaltlich zusammengeführt und vor dem Hintergrund zentraler Fragestellungen diskutiert werden. Auch soll an dieser Stelle Raum für neue Fragestellungen und Anregungen für weitere Experimente gegeben werden.

Wurde die Shepardsche Theorie bestätigt?

In den vorgestellten Experimenten wurden Paare von oktav-komplexen Tönen dargeboten und von den Versuchspersonen hinsichtlich der relativen Tonhöhe als aufsteigend oder absteigend beurteilt. Die Theorie von Shepard, nach der die Tonhöhenänderung von der Distanz zweier oktav-komplexer Töne auf dem Tonklassenkreis abhängt, konnte bestätigt werden. Allerdings zeigt sich speziell in den Experimenten 6.4 und 6.5 zum vollständigen Paarvergleich, dass es große individuelle Unterschiede gibt und diese sich nicht nur auf die Zuverlässigkeit der Tonhöhenurteile auswirken, sondern auch auf das Ausmaß des Einflusses bestimmter Tonklassen auf die Urteile.

Wie von Shepard (1964) formuliert, hängt die relative Tonhöhe primär von der Distanz auf dem Tonklassenkreis ab, der Einfluss der Tonklasse lässt sich jedoch nicht nur auf die Tritonusdistanz reduzieren, sondern erfasst in Einzelfällen, trotz sonst stabiler Urteile sogar die Distanzen 2 bis 10.

Am Einfluss des Parameters Amplitude (*alpha*) des P4-Modells (vgl. Abschnitt 5.2) konnte dieser Sachverhalt gut illustriert und modelliert werden. Eine Amplitude mit dem Wert 0 würde exakt zu der Shepardschen Theorie passen. Unabhängig vom Einfluss einzelner Tonklassen würde hierbei die relative Tonhöhe allein von der Distanz abhängen. Anhand der an die Daten angepassten Parameter zeigte sich jedoch eine mittlere Amplitude von etwa 2 Halbtonschritten. Dies bedeutet, dass die Schwelle vom Auf- zum Abwärts-Urteil im Mittel einen Ganzton um die Tritonusdistanz schwankt. In Einzelfällen wurden Amplitudenwerte von 3 und mehr Halbtonschritten gefunden. Genau genommen ist somit das Tritonusparadox gar kein Paradox des Tritonusintervalls, sondern reicht weit darüber hinaus.

Die oben gestellte Frage muss somit mit einem „ja, aber“ beantwortet werden. Aufgrund der Ergebnisse aus den vollständigen Paarvergleichen konnte verdeutlicht werden, dass die Shepardsche Theorie im Allgemeinen gilt, aber

der Einfluss der Tonklasse mit Annäherung an die Tritonusdistanz zunimmt und gleichzeitig der Einfluss der Distanz abnimmt.

Welche Einflussfaktoren gibt es?

In den durchgeführten Experimenten konnte das Phänomen der Zirkularität der Tonhöhenurteile, welches oktav-komplexen Tönen eigen ist, wiederholt bestätigt werden (z.B. Distanz 1 wird immer als aufsteigend bewertet). Es wurde jedoch auch gezeigt, wann dieses Phänomen zerfällt und in die 'normale' Tonhöhenbeurteilung übergeht. Dies wurde besonders beim Experiment zum Einfluss der Hüllform deutlich. Bei Wegnahme der gewichtenden Hüllkurve zeigt sich, dass die Urteile wie z.B. bei Sinustönen oder harmonisch-komplexen Tönen gemäß der fundamentalen Frequenz beurteilt werden. Dies unterstreicht einmal mehr den Einfluss der Hüllkurve. Die exakte Form der Hülle scheint dabei keinen bedeutsamen Einfluss zu haben, solange das Ein- und Ausschleichen der Teiltöne begünstigt wird. Deutsch (1986) stellt eine halbkreisförmige Hülle dar, Pollack (1978) verwendet eine Dreiecksform und Teranishi (2001) benutzt eine trapezförmige Hülle.

Wird z.B. davon ausgegangen, dass eine wesentliche Bedeutung der Hülle das Bilden der fundamentalen Frequenz erschweren bzw. verhindern soll, so kommt der tieferfrequenten Hüllenseite deutlich mehr Bedeutung zu. Anders ausgedrückt, auf die 'obere', höherfrequente Hälfte der Hülle könnte verzichtet werden. Hier würde sich ein weiterführendes Experiment anbieten, um diese Fragestellung zu überprüfen. Denkbar wäre auch, ein Maskierungsexperiment, bei dem die niederfrequenten Komponenten oder andere Teiltöne maskiert werden und somit nicht in einen Vergleichsprozess einfließen könnten.

Ein weiterer wesentlicher Einfluss zeigte sich in Bezug auf die Anzahl der Teiltöne eines oktav-komplexen Tones. Im Paarvergleichsexperiment 6.4 konnte verdeutlicht werden, dass mit zunehmender Anzahl von Teiltönen auch die Sicherheit der Urteile zunimmt. Speziell konnte hierbei auch der Wechsel vom zirkulären Urteil zum Fundamentalurteil aufgezeigt werden.

Diese Beobachtung unterstützt auch die Aussage von Pollack (1978), wonach es mindestens drei korrespondierende Teiltonpaare benötigt, um zirkuläre Urteile zu erhalten. Ebenfalls sprechen die Befunde aus dem Experiment zur dichotischen Darbietung für diese Hypothese. Hier wurde gezeigt, dass trotz der Wegnahme von Teiltönen zirkuläre Urteile möglich waren, die Sicherheit jedoch dabei stark abnahm. Nicht die Anzahl der Teiltöne ist allein entscheidend, sondern die Möglichkeit des direkten Vergleiches von Komponenten.

Speziell im zweiten Experiment zur dichotischen Darbietung (Distanz 4 und 8) wurde dieser Sachverhalt deutlich. Die Bedingung *control* wurde—mit nur halb so vielen Komponenten—sicherer bewertet als die Bedingung *original* mit dem (zentral) vollen Teiltonspektrum. Der direkte Vergleich oder eine möglichst geringe Distanz zwischen zwei zu vergleichenden Komponenten ist entscheidender als die bloße Anzahl der Teiltöne.

Auch hier wäre ein zusätzliches Experiment denkbar. Beim Vergleich von zwei oktav-komplexen Tönen könnte ein Paar von Teiltönen getauscht werden, so dass bei sonst z.B. stets 'aufsteigenden' Paaren, z.B. bei Distanz 2, ein 'absteigendes' Komponentenpaar enthalten ist. Variiert man nun die Position dieses Paares, sollte sich eine Aussage über den jeweiligen Einfluss ableiten lassen. Interessante Ergebnisse dürften sich hier bei der Position der kleinsten (fundamentalen) Frequenz oder bei dem zentralen Paar mit der größten Amplitude finden lassen.

In einem weiteren Schritt wäre auch eine Kombination mehrerer oppositioneller Paare denkbar. Auch könnten anhand dieses Experimentes (z.B. durch Verschiebung des Sets) Rückschlüsse auf den Einfluss bestimmter Frequenzen, wie z.B. der von Terhardt (1972b) eingeführten spektralen Dominanzregion, gezogen werden.

Gibt es eine subjektiv höchste Tonklasse?

Im Experiment 6.3 zur Tritonusaufgabe wurden Tonpaare im Abstand von sechs Halbtönen dargeboten. Hier konnte zwar gezeigt werden, dass die höchste Tonklasse für deutschsprachige Hörer zwischen Cis und D liegt, die Ergebnisse jedoch aufgrund der hohen Varianz nicht generalisierbar sind.

Tritonuspaare wurden auch in den Experimenten zum vollständigen Paarvergleich und in der Bedingung *original* der Experimente zur dichotischen Darbietung präsentiert. Im Paarvergleichsexperiment (Abschnitt 6.4) wurde die Teiltonanzahl in 5 Stufen variiert. Nur in der Bedingung mit 6 Teiltönen zeigt sich eine signifikant unimodale Verteilung. Die allgemein am höchsten bewertete Tonklasse liegt hier bei Fis. Die übrigen Bedingungen weisen eine breite Streuung der individuellen Tonklassen auf.

Auch in der Bedingung *original* des Experimentes in Abschnitt 6.7 wurden Tritonuspaare bei oktav-komplexen Tönen mit 10 Teiltönen dargeboten. Hier wurde die Tonklasse E als höchste Tonklasse bewertet. Es zeigt sich jedoch auch hierbei kein eindeutiges Ergebnis. Die individuellen Tonklassen variieren so stark, dass im Mittel keine Tonklasse mehr als 10% von einer relativen

Häufigkeit von 0.5 abweicht. Somit kann auch hier nicht von *der* höchsten Tonklasse gesprochen werden.

Bezieht man diese Befunde auf die von Deutsch (1987) postulierte Medien-Hypothese, wonach alle Hörer einer Sprachgemeinschaft die gleiche höchste Tonklasse aufweisen, so können verschiedene Schlussfolgerungen abgeleitet werden. Bei Gültigkeit der Medien-Hypothese könnte vermutet werden, dass die Versuchspersonen aus so unterschiedlichen Gebieten Deutschlands kommen, dass die sprachliche Gemeinsamkeit durch regionaltypische Dialekte verzerrt wird. Ein Beleg für den Einfluss des Dialektes wurde von Chalikia und Vaid (1999b) vorgelegt.

Dies erscheint jedoch eher unwahrscheinlich. Auch wenn der Sozialisationshintergrund der einzelnen Versuchspersonen nicht erhoben wurde, kann aufgrund der Studienplatzwahl zur Wohnortnähe davon ausgegangen werden, dass die meisten Teilnehmer Studenten aus Sachsen-Anhalt waren. Umgekehrt kann die Studie von Chalikia und Vaid (1999b) auch als Beleg für die Instabilität der subjektiv höchsten Tonklasse innerhalb einer Sprachgemeinschaft herangezogen werden.

Gegen die Sprachabhängigkeit der höchsten Tonklasse spricht weiterhin der Einfluss verschiedener Reizparameter. So konnte im zweiten Experiment zum vollständigen Paarvergleich (Abschnitt 6.5) gezeigt werden, dass bei einer Verschiebung der Hülle um eine halbe Oktave sich auch die höchste Tonklasse von Gis auf C verschiebt. Gleiche Befunde wurden u. a. von Repp (1997) bei einer Verschiebung des Zentrums der Hülle von A nach Dis berichtet.

Fast man die bisherigen Ergebnisse zusammen, so zeigt sich, dass Tritonuspaare von den meisten Probanden reliabel als auf- oder absteigend bewertet werden. Die Hypothese der Sprachabhängigkeit der höchsten Tonklasse kann jedoch nicht verifiziert werden. Welche Tonklasse von einer Person als die subjektiv höchste empfunden wird, scheint nicht sprachgebunden, sondern beruht vermutlich stärker auf individuell unterschiedlichen Wahrnehmungleistungen.

Ein Beleg für interindividuelle Unterschiede bei der Tonhöhenwahrnehmung oktav-komplexer Töne wird von Krüger und Lukas (2002) berichtet. Bei Paarvergleichsexperimenten wurden den Hörern zwei Sets von oktav-komplexen Tönen präsentiert. Dabei waren die Hüllen beider Sets um eine halbe Oktave versetzt. Es konnte gezeigt werden, dass sich die höchste Tonklasse einiger Personen ebenfalls um eine halbe Oktave verschob, während andere Hörer stabil eine höchste Tonklasse präferierten.

Als Erklärung für die individuellen Unterschiede wurden verschiedene Hypothesen postuliert. So gehen z.B. Terhardt und Grubert (1988) davon aus,

dass die Ausprägung der spektralen Dominanzregion (vgl. Abschnitt 5.1.4) einen Einfluss auf die Tritonusaufgabe hat. Die Tatsache, dass Hörer bestimmte Frequenzregionen stärker gewichten als andere, erscheint allein aufgrund der individuellen Physiologie plausibel.

In die gleiche Reihe könnte z.B. die Ausprägung der Hörschwelle eines Probanden gestellt werden. Geht man von diesem Erklärungsansatz aus, so könnte zwar die Variation zwischen den Hörern dadurch erklärt werden, jedoch stellt sich die Frage, ob auch der Wechsel der höchsten Tonklasse bei Verschiebung der Hülle, welcher bei einigen Hörern auftritt und bei anderen nicht, damit erklärt werden kann.

Dass die Unterschiede in den allgemeinen Hörparametern die Phänomene beim Tritonusparadox erklären können, erscheint prinzipiell möglich, kann jedoch an dieser Stelle nicht abschließend beurteilt werden. Zur Beantwortung dieser Frage müsste vielmehr vorab ein Mechanismus benannt werden, welcher erklärt, wie die eine oder andere Spektralkomponente gewichtet wird und wie sich dies auf den Vergleich zweier oktav-komplexer Töne auswirkt. Bereits jetzt kann vermutet werden, dass die hypothetische Gewichtungsfunktion zwischen den Probanden extrem variieren müsste, um die Ergebnisse bei der Tritonusaufgabe beschreiben zu können.

Ein weiterer Ansatz, welcher die individuellen Variationen erklären könnte, bezieht sich auf die Unterteilung der Probanden in analytische und synthetische Hörer (vgl. Abschnitt 4.2.2). Im Gegensatz zu den bisher diskutierten individuellen Parametern (Hörschwelle, Dominanzregion, Sprachabhängigkeit i.S. einer Präferenztonhöhe) wird hier von einer unbewussten Wahrnehmungsleistung ausgegangen, welche die Aufmerksamkeit eines Hörers entweder auf spektrale (analytische Wahrnehmung) oder virtuelle (synthetische Wahrnehmung) Tonhöhekomponenten lenkt. Parncutt und Tan (2005) kombinieren diesen Ansatz mit einer individuell ausgeprägten spektralen Dominanzfunktion, um die Antworten in der Tritonusaufgabe zu erklären.

Allein das Wissen, ob ein Hörer eher analytisch oder synthetisch wahrnimmt, lässt kaum Rückschlüsse auf die Tonhöhenpräferenz bei oktav-komplexen Tönen zu. Hinweise auf eine klare Trennung der Probanden in zwei Gruppen lassen sich zwar im Paarvergleichsexperiment 6.4 in den Bedingungen mit 2 und 4 Teiltönen finden, allerdings bleibt 1.) unklar, ob es sich bei den Gruppen auch um die synthetischen bzw. die analytischen Hörer handelt und 2.) lassen sich diese Gruppierungen nicht in den übrigen Bedingungen wiederfinden.

Um die individuelle Hörpräferenz für das Grundtonhören oder das Ober-

tonhören zu erheben, könnte ein Vorexperiment durchgeführt werden. Dazu könnte die relative Tonhöhe eines Paares von Zwei-Komponenten-Tönen beurteilt werden. Wird z.B. ein Ton, bestehend aus einer 750 Hz und einer 1000 Hz Komponente, mit einem Ton, bestehend aus einer 800 Hz und einer 1000 Hz Komponente, verglichen, so könnte das Urteil Auskunft über den Hörertyp geben. Wird das Paar als aufsteigend beurteilt, so handelt es sich um einen analytischen Hörer, da hier die Teiltöne mit 750 Hz und 800 Hz verglichen werden. Wird das Paar als absteigend beurteilt, so kann von einem synthetischen Hörer ausgegangen werden, da hier die virtuellen, fundamentalen Tonhöhen der Töne 250 Hz und 200 Hz verglichen werden.

In experimentellen Untersuchungen konnten die Hörer zwar der einen oder anderen Gruppe zugeordnet werden (Smooenburg, 1970; Houtsma & Fleuren, 1991), deutlich wurde jedoch auch, dass die Unterteilung eher als Trend anzusehen ist und es nicht *den* analytischen oder *den* synthetischen Hörer gibt. Hier bleibt offen, ob es einen Zusammenhang zwischen Hörertyp und Tritonusparadox gibt. Potentiell wäre dieser Zugang jedoch in der Lage zu erklären, warum einige Probanden bei einer Verschiebung der Hüllkurve bei ihrem Urteil bleiben, während andere das Tonhöhenurteil ändern.

Zur Beantwortung der Frage, ob es Korrelate zur subjektiv höchsten Tonklasse gibt, konnten in Bezug auf die Variablen Alter, Geschlecht, Händigkeit, Musikalität, Hörschwelle und Diskriminationsleistung keine Hinweise gefunden werden. Die Liste dieser Variablen stellt einen Ausschnitt an verfügbaren Korrelaten dar, welche für eine Verbindung mit dem Tritonusparadox in Frage kommen könnten und ließe sich problemlos erweitern. Zu erwähnen ist hierbei, dass es sich lediglich um eine post-hoc-Testung handelt, welche allein auf der Verfügbarkeit dieser Daten in den dargestellten Untersuchungen fußt.

Als Fazit zum Tritonusparadox kann festgehalten werden, dass es zwar eine individuell stabil ausgeprägte höchste Tonklasse gibt, diese jedoch abhängig ist von bestimmten Reiz- und Personenparametern. Reizparameter sind z.B. die Position der Hülle, die Anzahl der Teiltöne oder die Form der Hülle. Die relevanten Personenparameter bleiben hypothetisch. Kandidaten sind die Sprach- bzw. Hörprägung als Oberbegriff verschiedener auditiver Parameter oder die Hörgewohnheit i.S. des analytischen oder synthetischen Hörers. Weiterhin ist der Grad der Ausprägung der subjektiv höchsten Tonklasse abhängig von der Fähigkeit einer Person, Tonhöhe zu diskriminieren.

Warum gibt es zwei Hörstrategien?

Im mehreren Experimenten wurde die Hypothese der zwei Hör- bzw. Urteilsstrategien (Distanz- bzw. Fundamental-Urteile) aufgestellt und überprüft. Die Phänomene auf der Basis oktav-komplexer Töne funktionieren nur, weil neben dem sonst üblichen Grundton-Mechanismus ein weiterer Mechanismus besteht, welcher auf einen Vergleich der Abstände zwischen den Teiltönen aufbaut. Man könnte noch weiter gehen und behaupten, dass dieser Mechanismus der basälere ist, denn er kann nur greifen, bevor eine Reduktion der Teilfrequenzen auf eine dem Grundton bzw. der Periode äquivalente Repräsentation von (virtueller) Tonhöhe geschieht. Hinweise auf einen derartigen Mechanismus lassen sich aus den Experimenten zum dichotischen Hören ableiten. Hier wurde auf die zentrale Lokalisation der Tonhöhemechanismen hingewiesen. Aufgrund der hohen Konnektivität, Redundanz und Parallelität in der Informationsverarbeitung des menschlichen Gehirns, erscheint es wahrscheinlicher, von zwei getrennten Subsystemen auszugehen.

An dieser Stelle könnte eine Vielzahl weiterer Frage formuliert werden; z.B. lassen sich Urteile aufgrund der Distanz auch bei anderen Reizen finden oder anwenden, hat die fundamentale oder kleinste Frequenz von oktav-komplexen Tönen einen Einfluss auf die Tonhöhe, evtl. sogar im Bereich des Tritonus?

In der vorliegenden Dissertation konnten zahlreiche Fragestellungen zur Wahrnehmung der Tonhöhe bei oktav-komplexen Tönen beantwortet und neue Fragestellungen eröffnet werden. Die diskutierten Befunde weisen auf alternative Erklärungsansätze hin und legen somit den Grundstein für eine Reihe weiterführender, experimenteller Untersuchungen.

6.9 Ausblick

Unser Wahrnehmungssystem verfügt über eine Vielzahl von Mechanismen, um die uns umgebende Umwelt zu strukturieren. Dies geschieht bereits auf sehr frühen Stufen der Verarbeitung und zieht sich hin bis zum eigentlichen Erkennen, Interpretieren und Bewerten des Wahrgenommenen. Über die unterschiedlichen Modi der fünf Sinne hinweg treten die gleichen Mechanismen auf. Im Vergleich des Visuellen mit dem Auditiven lässt sich zeigen, dass z.B. der auditive Kontureffekt dem bei den Mach-Bändern bekannten Kontrastverstärkung gleichkommt. Nacheffekte lassen sich für beide Systeme nachweisen. Das visuelle Nachbild findet z.B. in der Zwickerschen-Folgetonhöhe ihr Pendant. Die Fülle der in den Gestalt-Gesetzen beschriebenen Organisationsprinzipien behalten ihre Gültigkeit auch bei auditiven Reizen.

In den beschriebenen Experimenten konnten einige dieser Prinzipien demonstriert werden. So konnte gezeigt werden, dass die Tonhöhe bei oktavkomplexen Tönen sich analog zum visuellen Phänomen der Scheinbewegung verhält. Dabei tritt an die Stelle der Bewegung die Tonhöhenänderung.

Speziell von den oktavkomplexen Tönen wird behauptet, dass die sonst üblichen Prinzipien (Urteile gem. der fundamentalen Frequenz) ihre Berechtigung verlieren und sich statt dessen zirkuläre Urteile einstellen. In den vollständigen Paarvergleichsexperimenten konnte gezeigt werden, dass der Übergang von der einen zur anderen Urteilsstrategie im Wesentlichen abhängig ist von der Hüllkurve. Ein geringfügiger Einfluss lässt sich auch auf die Anzahl der Teiltöne zurückführen.

In Bezug auf eine Erklärung zum oftmals untersuchten Tritonusparadoxon wird deutlich, dass das scheinbar paradoxe Antwortverhalten nicht erst beim Tritonusintervall einsetzt, sondern in Abhängigkeit von der Fähigkeit zur Tonhöhendiskrimination bereits weit davor. Über die dabei involvierten Einflussfaktoren lässt sich nach wie vor streiten. Zwar kann die Sprachhypothese aufgrund zahlreicher Studien als substantiell angesehen werden, jedoch lassen sich aus den vorliegenden Ergebnissen keine Hinweise für eine Sprachabhängigkeit der Urteile finden. Es konnte ein signifikanter Hülleneffekt nachgewiesen werden. Hier bleibt weiter unklar, aufgrund welcher Basis die individuellen Unterschiede entstehen.

Ausgehend von den vorwiegend psychophysischen Befunden sollten bei zukünftigen Überlegungen weitere kognitionspsychologische Themenfelder wie Aufmerksamkeit, Gedächtnis oder Entscheidungsfindung stärker in die bisherigen Theoriebildungen integriert werden. In Ansätzen lassen sich diese Aspekte

bereits finden, z.B. die Aufmerksamkeitslenkung beim analytischen oder holistischen Hören (vgl. Abschnitt 4.2.2) oder der Einfluss des impliziten Gedächtnisses beim absoluten Gehör (vgl. Abschnitt 3.1.5). Letztlich erscheint es lohnend, den gesamten Prozess von Wahrnehmung, Repräsentation, Interpretation, Bewertung und Reaktion zu beachten.

Literaturverzeichnis

- Arehart, K. H. & Burns, E. M. (1999). A comparison of monotonic and dichotic complex-tone pitch perception in listeners with hearing loss. *Journal of the Acoustic Society of America*, 106, 993–997.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society Series A*, 160, 268–282.
- Bernstein, J. G. & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *Journal of the Acoustic Society of America*, 113(6), 3323–3334.
- Berthoud, S., Jarchow, T. & Long, D. (2004). The perception of dichotic shepard tone intervals - fusion or suppression? In H. H. Bülhoff, H. A. M. und R. Ulrich & F. A. Wichmann (Hrsg.), *Proceedings of the 7th Tübingen Perception Conference (TWK 2004)*. Kirchentellinsfurt: Knirsch.
- Bilsen, F. (1977). Pitch of noise signals: Evidence for a 'central spectrum'. *Journal of the Acoustic Society of America*, 91, 150–161.
- Bilsen, F. A. (1973). On the influence of the number and phase of harmonics on the perceptibility of pitch of complex signals. *Acoustica*, 28, 60–65.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Science, University of Amsterdam*, 17, 97–110.
- Braun, M. (2001). Speech mirrors norm-tones: Absolute pitch as a normal but precognitive trait. *Acoustics Research Letters Online*, 2(3), 85–90.
- Burns, E. M. (1981). Circularity in relative pitch judgments for inharmonic complex tones: The shepard demonstration revisited, again. *Perception & Psychophysics*, 30, 467–472.
- Burns, E. M. & Ward, W. D. (1982). Intervals, scales, and tuning. In D. Deutsch (Hrsg.), *The psychology of music* (S. 241–269). New York: Academic Press.
- Cariani, P. (2001). Temporal codes, timing nets, and music perception. *Music Perception*, 30(2), 526–566.

- Carlyon, R. P., Demany, L. & Deeks, J. (2001). Temporal pitch perception and the binaural system. *Journal of the Acoustic Society of America*, *109*(2), 686–700.
- Carlyon, R. P. & Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms. *Journal of the Acoustic Society of America*, *95*(6), 3541–3554.
- Chalikia, M. H. & Leinfelt, F. (2000). Listeners in Sweden perceive the tritone stimuli in a manner different from that of Americans and similar to that of British listeners. *Journal of the Acoustic Society of America*, *108*, 2572.
- Chalikia, M. H., Miller, K. J. & Vaid, J. (2001). *The tritone paradox is perceived differently by Koreans and Americans*. Paper presented at the 101st Annual Convention of the American Psychology Association, San Francisco, CA.
- Chalikia, M. H. & Vaid, J. (1999a). Do bilinguals perceive the tritone paradox as monolinguals do? *Journal of the Acoustic Society of America*, *106*(4), 2207–2207.
- Chalikia, M. H. & Vaid, J. (1999b). Perception of the tritone paradox by listeners in Texas. *Journal of the Acoustic Society of America*, *106*, 2572.
- Chambers, C. D., Mattingley, J. B. & Moss, S. A. (2002). The octave illusion revisited: Suppression or fusion between ears? *Journal of Experimental Psychology: Human Perception and Performance*, *28*(6), 1288–1302.
- Cohen, M. A., Grossberg, S. & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of the Acoustic Society of America*, *98*(2), 862–879.
- Cramer, E. & Huggins, W. H. (1958). Creation of pitch through binaural interaction. *Journal of the Acoustic Society of America*, *30*, 413–417.
- Cuddy, L. L. (1968). Practice effects in the absolute judgements of pitch. *Journal of the Acoustic Society of America*, *34*, 1069–1076.
- Dawe, L. A., Platt, J. R. & Welsh, E. (1998). Spectral-motion aftereffects and the tritone paradox among canadian subjects. *Perception & Psychophysics*, *60*(2), 209–220.
- de Cheveigné, A. (1998a). Cancellation model of pitch perception. *Journal of the Acoustic Society of Jpn.*, *103*, 1261–1271.
- de Cheveigné, A. (1998b). *Time domain processing in the auditory system*. Paper presented at the International Conference on Neural Information

- Processing (ICONIP), Kitakyushu, Japan.
- de Cheveigné, A. (2004). Pitch perception models. In C. Plack & A. Oxenham (Hrsg.), *Pitch*. New York: Springer.
- de Cheveigné, A. & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustic Society of America*, *111*(4), 1917–1930.
- Demany, L. & Semal, C. (1988). Dichotic fusion of two tones one octave apart: Evidence for internal octave template. *Journal of the Acoustic Society of America*, *83*, 687–695.
- Deutsch, D. (1973). Octave generalization of specific interference effects in memory for tonal pitch. *Perception & Psychophysics*, *13*, 271–275.
- Deutsch, D. (1974). An auditory illusion. *Nature*, *251*, 307–309.
- Deutsch, D. (1986). A musical paradox. *Music Perception*, *3*, 275–280.
- Deutsch, D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics*, *41*, 563–575.
- Deutsch, D. (1988). The semitone paradox. *Music Perception*, *6*(2), 115–132.
- Deutsch, D. (1991). The tritone paradox: An influence of language on music perception. *Music Perception*, *8*(4), 335–347.
- Deutsch, D. (2007). Mothers and their offspring perceive the tritone paradox in closely similar ways. *Archives of Acoustics*, *32*(3), 3–14.
- Deutsch, D. & Dolson, M. (2004). Speech patterns heard early in life influence later perception of the tritone paradox. *Music Perception*, *21*(3), 357–372.
- Deutsch, D., North, T. & Ray, L. (1990). The tritone paradox: Correlate with the listener's vocal range for speech. *Music Perception*, *7*, 371–384.
- Deutsch, D. & Roll, P. L. (1976). Separate What and Where decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(1), 23–29.
- Dougherty, J. M. & Garner, W. R. (1948). Pitch characteristics of short tones. II. Pitch as a function of tonal duration. *Journal of Experimental Psychology*, *38*(4), 478–494.
- Duifhuis, H., Willems, L. F. & Sluyter, R. J. (1982). Measuring of pitch in speech. An implementation of Goldstein's theory of pitch perception. *Journal of the Acoustic Society of America*, *71*, 1568–1580.
- Fastl, H. & Stoll, G. (1979). Scaling of pitch strength. *Hearing Research*, *1*, 293–301.
- Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge: University Press.

- Fletcher, H. & Munson, W. A. (1937). Relation between loudness and masking. *Journal of the Acoustic Society of America*, 9, 1–10.
- Giangrande, J. (1998). The tritone paradox: Effects of pitch class and position of the spectral envelope. *Music Perception*, 15(3), 253–264.
- Giangrande, J., Tuller, B. & Kelso, J. A. S. (2003). Perceptual dynamics of circular pitch. *Music Perception*, 20(3), 241–262.
- Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustic Society of America*, 54, 1496–1516.
- Goodman, A. (1965). Reference zero levels for pure tone audiometer. *American Speech-Language-Hearing Association (ASHA)*, 7, 262–263.
- Grimault, N., Micheyl, C., Carlyon, R. P. & Collet, L. (2002). Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. *Perception & Psychophysics*, 64(2), 189–197.
- Groen, J. J. (1964). Super- and subliminal binaural beats. *Acta Otolaryngol*, 57, 224–231.
- Hall, D. E. (1991). *Musical acoustics*. Belmont: Brooks & Cole.
- Hall, J. L. (1981). Hybrid adaptive procedures for estimation of psychometric functions. *Journal of the Acoustic Society of America*, 69, 1763–1769.
- Hall, J. W., Haggard, M. & Fernandes, M. A. (1984). Detection in noise by spectro-temporal pattern analysis. *Journal of the Acoustic Society of America*, 76, 50–60.
- Hanke, M. & Krüger, S. (2005). *PsychoExp - Classes for psychological experiments*. [Computer Software]. Dept. of Psychology, University of Halle, Germany.
- Hartmann, W. M. (1998). *Signals, sound, and sensation*. New York: Springer.
- Hartmann, W. M. (2000). Binaural coherence edge pitch. *Journal of the Acoustic Society of America*, 109, 294–305.
- Hartmann, W. M. & Zhang, P. X. (2003). Binaural models and the strength of dichotic pitches. *Journal of the Acoustic Society of America*, 114(6), 3317–3326.
- Harvey, L. O. (1997). Efficient estimation of sensory thresholds with ML-PEST. *Spatial Vision*, 11(1), 121–128.
- Helmholtz, H. L. F. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Vieweg.
- Henning, G. (1966). Frequency discrimination of random amplitude tones. *Journal of the Acoustic Society of America*, 39, 336–339.
- Hermes, D. J. (1988). Measurement of pitch by subharmonic summation.

- Journal of the Acoustic Society of America*, 83, 257–264.
- Houtsma, A. J. M. & Fleuren, J. F. M. (1991). Analytic and synthetic pitch of two-tone complexes. *Journal of the Acoustic Society of America*, 90, 1674–1676.
- Houtsma, A. J. M. & Goldstein, J. L. (1971). The central origin of the pitch of complex tones: Evidence from musical interval recognition. *Journal of the Acoustic Society of America*, 51, 520–529.
- Houtsma, A. J. M. & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustic Society of America*, 87, 304–310.
- Jammalamadaka, S. R. & SenGupta, A. (2001). *Topics in circular statistics*. Singapore: World Scientific Press.
- Jones, J. A. & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustic Society of America*, 108, 1246–1251.
- Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: The University of Chicago Press.
- Kaernbach, C. & Bering, C. (2001). Temporal mechanisms of pitch perception. *Journal of the Acoustic Society of America*, 110(2), 1039–1048.
- Kaernbach, C. & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *Journal of the Acoustic Society of America*, 104(4), 2298–2306.
- Karjalainen, M. & Tolonen, T. (1999). Multi-pitch and periodicity analysis model for sound separation and auditory scene analysis. In M. Baumann, A. Keinath & J. F. Krems (Hrsg.), *In Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*. Phoenix, Arizona, USA: ICASSP.
- Klapuri, A. (2006). Multiple analysis of polyphonic music and speech signals using an auditory model. *IEEE Transaction on Audio, Speech, and Language Processing*, 16(2), 255–266.
- Klein, M. A. & Hartmann, W. M. (1980). Binaural edge pitch. *Journal of the Acoustic Society of America*, 70, 51–61.
- Krüger, S. (2003). *Das Tritonusparadox in der Tonhöhenwahrnehmung*. Unveröffentlichte Diplomarbeit, Institut für Psychologie, Martin-Luther-Universität Halle-Wittenberg.
- Krüger, S. (2006). *The perception of pitch of octave-complex tones in a dichotic listening experiment*. Paper presented at the DAGA 2006 (Jahrestagung der Gesellschaft für Akustik), Braunschweig.

- Krüger, S. & Lukas, J. (2002). Zirkuläre Urteile bei der Tonhöhenwahrnehmung: Ein internationaler Vergleich. In M. Baumann, A. Keinath & J. F. Krems (Hrsg.), *Experimentelle Psychologie - Beiträge zur 44. Tagung experimentell arbeitender Psychologen - TeaP* (S. 153). Lengerich: Pabst.
- Krüger, S. & Lukas, J. (2005). Dichotische Reizdarbietung von Oktavkomplexen Reizen: zentrale oder periphere Verarbeitung? In K. W. Lange, K. Bäuml, M. W. Greenlee, M. Hammerl & A. Zimmer (Hrsg.), *Experimentelle Psychologie - Beiträge zur 47. Tagung experimentell arbeitender Psychologen - TeaP* (S. 117). Lengerich: Pabst.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, *63*(8), 1279–1292.
- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*, 414–423.
- Liang, C. A. & Chistovich, L. A. (1961). Frequency difference limen as a function of tonal duration. *Sov. Phys. Acoust.*, *6*, 75–80.
- Licklider, J. C. R. (1959). Three auditory theories. In S. Koch (Hrsg.), *Psychology, a study of a science* (S. 41–144). New York: McGraw-Hill.
- Lockhead, G. R. & Byrd, R. (1981). Practically perfect pitch. *Journal of the Acoustic Society of America*, *70*, 387–389.
- Lukas, J. (2004). *Zur Trennung der Komponenten für das rechte und linke Ohr im Experiment 'binaural'*. Unveröffentlichtes Manuskript, Martin-Luther-Universität Halle-Wittenberg.
- Lukas, J. & Krüger, S. (2001a). Das Tritonus-Paradox bei der Tonhöhenwahrnehmung und seine Beziehung zur Bewegungswahrnehmung beim Phi-Phänomen. In H. H. Bülthoff, K. R. Gegenfurtner, H. A. Mallot & R. Ulrich (Hrsg.), *TWK 2001. Beiträge zur 4. Tübinger Wahrnehmungskonferenz* (S. 161). Kirchentellinsfurt: Knirsch.
- Lukas, J. & Krüger, S. (2001b). Zirkuläre Urteile bei der Tonhöhenwahrnehmung und das Tritonus-Paradox. In A. Zimmer, K. Lange, K. Bäuml, R. Loose, R. Scheuchenpflug, O. Tucha, H. Schnell & R. Findl (Hrsg.), *Experimentelle Psychologie* (S. 94). Lengerich: Pabst.
- McKay, C. M., McDermott, H. J. & Carlyon, R. P. (2000). Place and temporal cues in pitch perception: Are they truly independent? *Acoustics Research Letters Online*, *1*, 25–30.
- Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustic Society of America*, *79*, 702–

711.

- Meddis, R. (1988). Simulation of auditory-neural transduction: Further studies. *Journal of the Acoustic Society of America*, *83*, 1056–1063.
- Meddis, R. & Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. *Journal of the Acoustic Society of America*, *89*, 2883–2894.
- Meddis, R. & Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustic Society of America*, *89*, 2866–2882.
- Meddis, R. & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustic Society of America*, *102*, 1811–1820.
- Micheyl, C. & Oxenham, A. J. (2005). Comparing F0 discrimination in sequential and simultaneous conditions. *Journal of the Acoustic Society of America*, *118*, 41–44.
- Miller, R. L. (1970). Performance characteristics of an experimental harmonic identification pitch extraction (HIPEX) system. *Journal of the Acoustic Society of America*, *47*, 1593–1601.
- Moore, B. C. J. (2003). *Psychology of hearing*. London: Academic Press.
- Moore, B. C. J. & Glasberg, B. R. (1983). Suggested formulae for calculating auditory filter bandwidths and excitation patterns. *Journal of the Acoustic Society of America*, *74*, 770–773.
- Moore, B. C. J. & Glasberg, B. R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing Research*, *28*, 209–225.
- Nakajima, Y., Tsumura, T., Matsuura, S., Minami, H. & Teranishi, R. (1988). Dynamic pitch perception for complex tones derived from major triads. *Music Perception*, *6*(1), 1–20.
- Noll, A. M. (1967). Cepstrum pitch determination. *Journal of the Acoustic Society of America*, *41*, 293–309.
- Normann, I., Purwins, H. & Obermayer, K. (2001). Interdependence of pitch and timbre perception for octave ambiguous tones. *Jahrestagung der Gesellschaft für Akustik (DAGA)*.
- O'Mard, L. & Meddis, R. (2007). *Development system for auditory modelling (DSAM), version 2.8*. [Computer Software]. Department of Psychology, University of Essex, UK.
- Pantev, C., Elbert, T., Ross, B., Eulitz, C. & Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hearing Research*, *100*, 164–170.

- Parncutt, R. (1993). Pitch properties of chords of octave-spaced tones. *Contemporary Music Review*, 9, 35–50.
- Parncutt, R. & Tan, A. P. (2005). *The tritone paradox and the dominance regions of spectral and virtual pitch*. Paper presented at the 6th Biennial Congress of the Australian Institut of Physics, Canberra, Australia.
- Patterson, R. D. (1987). A pulse ribbon model of monaural phase perception. *Journal of the Acoustic Society of America*, 82, 1560–1586.
- Patterson, R. D. (2000). Auditory images: How complex sounds are represented in the auditory system. *Journal of the Acoustic Society Japan*, 21(4), 183–190.
- Patterson, R. D., Allerhand, M. H. & Giguere, C. (1995). Time domain modeling of auditory processing: A modular architecture and a software platform. *Journal of the Acoustic Society of America*, 98, 1890–1894.
- Patterson, R. D., Handel, S., Yost, W. A. & Datta, A. J. (1997). The relative strength of the tone and noise components in iterated rippled noise. *Journal of the Acoustic Society of America*, 100, 3286–3294.
- Patterson, R. D. & Irino, T. (1998). Auditory temporal asymmetry and auto-correlation. In A. Palmer, A. Rees, Q. Summerfield & R. Meddis (Hrsg.), *Psychophysical and physiological advances in hearing: Proceedings of the 11th international symposium on hearing* (S. 554–562). London: Whurr.
- Piszczałski, M. & Galler, B. A. (1979). Predicting musical pitch from component frequency ratios. *Journal of the Acoustic Society of America*, 66, 710–720.
- Plack, C. J. (2005). *The sense of hearing*. Hillsdale, N.J.: Lawrence Erlbaum.
- Pollack, I. (1978). Decoupling of auditory pitch and stimulus frequency: The shepard demonstration revisited. *Journal of the Acoustic Society of America*, 63(1), 202–206.
- Purwins, H. (2005). *Profiles of pitch classes circularity of relative pitch and key: Experiments, models, computational music analysis, and perspectives*. Dissertation. [http://opus.kobv.de/volltexte/2005/1085/pdf/purwins_hendrik.pdf]. Technische Universität Berlin.
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria. (ISBN 3-000051-07-0)
- Raatgever, J. & Bilsen, F. A. (1986). A central spectrum theory of binaural processing. Evidence from dichotic pitch. *Journal of the Acoustic Society of America*, 80, 429–441.
- Ragozzine, F. (2001). The tritone paradox and perception of single octave related complexes. *Music Perception*, 19(2), 155–168.

- Rensink, R. A., O'Regan, J. K. & Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368–373.
- Repp, B. H. (1994). The tritone paradox and the pitch range of the speaking voice: A dubious connection. *Music Perception*, 12(2), 227–255.
- Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception*, 26, 645–665.
- Repp, B. H. & Thompson, J. M. (2009). Context sensitivity and invariance in perception of octave-ambiguous tones. *Psychological Research*, [Epub ahead of print].
- Révész, G. (1972). *Einführung in die Musikpsychologie*. Bern: Francke Verlag.
- Risset, J. C. (1971). Paradoxes de hauteur: Le concept de hauteur de sonore n'est pas le meme pour tout le monde. *Proceedings of the Seventh International Congress on Acoustics, Budapest*, 10, 613–616.
- Rousseau, L., Peretz, I., Liegeois-Chauvel, C., Demany, L., Semal, C. & Larue, S. (1996). Spectral and virtual pitch perception of complex tones: An opposite hemispheric lateralization? *Brain and Cognition*, 30, 303–308.
- Ruckmick, C. (1929). A new classification of tonal qualities. *Psychological Review*, 36, 172–180.
- Scheffers, M. T. (1983). Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter. *Journal of the Acoustic Society of America*, 74, 1716–1725.
- Schouten, J. F. (1940). The perception of pitch. *Philips Techn. Rev.*, 5, 286–294.
- Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. *Journal of the Acoustic Society of America*, 43, 829–834.
- Schroeder, M. R. (1986). Auditory paradox based on fractal waveform. *Journal of the Acoustic Society of America*, 79(1), 186–189.
- Seneff, S. (1978). Real-time harmonic pitch detector. *IEEE ASSP*, 26, 358–365.
- Shaw, E. A. G. (1974). Transformation of sound pressure level from the free field to the horizontal plane. *Journal of the Acoustic Society of America*, 56, 1848–1861.
- Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustic Society of America*, 36(12), 2346–2353.
- Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Hrsg.), *The psychology of music* (S. 344–390). New York: Academic Press.

- Shofner, W. P. & Selas, G. (2002). Pitch strength and the Stevens's power law. *Perception & Psychophysics*, *64*(3), 437–450.
- Smooenburg, G. F. (1970). Pitch perception of two-frequency stimuli. *Journal of the Acoustic Society of America*, *48*, 924–942.
- Stevens, S. S. (1935). The relation of pitch to intensity. *Journal of the Acoustic Society of America*, *6*, 150–154.
- Stevens, S. S., Volkman, J. & Newman, E. B. (1937). A scale for measurement of the psychological magnitude of pitch. *Journal of the Acoustic Society of America*, *8*, 185–190.
- Strasburger, H. (2001). Converting between measures of slope of the psychometric function. *Perception & Psychophysics*, *63*(8), 1348–1355.
- Stumpf, C. (1883). *Tonpsychologie*, 2 Bd. Leipzig: Hirzel.
- Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P. & Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *Journal of the Acoustic Society of America*, *111*, 2178–2188.
- Taylor, M. M. & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustic Society of America*, *41*, 782–787.
- Teranishi, R. (2001). *Endlessly rising and falling chordal tones wich can be played on the piano; another variation of the Shepard demonstration*. Paper presented at the 12th Internation Congress on Acoustics, Toronto, Canada.
- Terhardt, E. (1971). Pitch shifts of harmonics, an explanation of the octave enlargement phenomenon. In *Proc. 7th Intern. Congr. Acoust.* (S. 621–624). Budapest.
- Terhardt, E. (1972a). Zur Tonhöhenwahrnehmung von Klängen. II Ein Funktionsschema. *Acustica*, *26*, 187–199.
- Terhardt, E. (1972b). Zur Tonhöhenwahrnehmung von Klängen. I Psychoakustische Grundlagen. *Acustica*, *26*, 173–186.
- Terhardt, E. (1974). Pitch of pure tones: Its relation to intensity. In E. Zwicker & E. Terhardt (Hrsg.), *Facts and models in hearing*. Berlin: Springer.
- Terhardt, E. (1979). Calculating virtual pitch. *Hearing Reseach*, *1*, 155–182.
- Terhardt, E. (1989). On the role of ambiguity of perceived pitch in music. In *13th international conference on acoustics* (S. 34–38). Belgrad.
- Terhardt, E. (1998). *Akustische Kommunikation*. Heidelberg: Springer Verlag.
- Terhardt, E. (2004). *VPITCH - Version 0.4*. [Computer Software]. Technische Universität München, Germany.
- Terhardt, E. & Grubert, A. (1988). Zur Erklärung des Tritonusparadoxons. In (DAGA) (Hrsg.), *Fortschritte der Akustik* (S. 717–720). Bad Honnef:

DPD.

- Terhardt, E., Stoll, G., Schermbach, R. & Parncutt, R. (1986). Tonhöhenmehrfachdeutigkeit, Tonverwandschaft und Identifikation von Sukzessivintervallen. *Acustica*, 61, 57–66.
- Turner, R. S. (1977). The Ohm-Seebeck Dispute, Hermann von Helmholtz, and the origins of physiological Acoustics. *The British Journal for the History of Science*, 10(1), 1–24.
- Ueda, K. & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. *Journal of the Acoustic Society of America*, 82, 1193–1200.
- van den Brink, G., Sintnicolaas, K. & van Stam, W. (1976). Dichotic pitch fusion. *Journal of the Acoustic Society of America*, 59, 1471–1476.
- Walliser, K. (1969). Über ein Funktionsschema für die Bildung der Periodentönhöhe aus dem Schallreiz. *Kybernetik*, 6, 65–72.
- WHO. (1992). *International statistical classification of diseases and health related problems. 10. Rev. (ICD-10)*. Geneva: World Health Organisation (WHO).
- Wichmann, F. & Hill, N. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics*, 63(8), 1293–1313.
- Wier, C., Jesteadt, W. & Green, D. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustic Society of America*, 61, 178–184.
- Wightman, F. L. (1973). Pitch and stimulus fine structure. *Journal of the Acoustic Society of America*, 54, 397–406.
- Wightman, F. L. & Green, D. M. (1974). The perception of pitch. *American Scientist*, 62, 208–215.
- Wilkening, F. & Krist, H. (2002). Entwicklung der Wahrnehmung und Psychomotorik. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (S. 395–417). Weinheim: Beltz.
- Yost, W. & Hill, R. (1979). Models of pitch and pitch strength of ripple noise. *Journal of the Acoustic Society of America*, 66, 400–410.
- Zwicker, E. (1954). Die Verdeckung von Schmalbandgeräuschen durch Sinustöne. *Acustica*, 8, 237–258.

Anhang

- Instruktionen
- Abbildungen: Daten und Schätzungen aller Versuchspersonen für das Experiment 6.4: Paarvergleich zum Einfluss der Teiltonanzahl
- Abbildungen: Daten aller Versuchspersonen für das Experiment 6.5: Paarvergleich zum Einfluss der Hüllenposition und -form

Instruktionen

Experiment: Hörschwellenaudiometrie

Instruktion

*Es werden Töne in verschiedenen Lautstärken
auf dem rechten oder auf dem linken Ohr dargeboten.*

Einzelne Töne können sehr leise sein.

Drücke eine Maustaste sobald du einen Ton hörst.

*(Mündlicher Hinweis des Versuchsleiters: Bevor du den Hörtest durch
drücken der Leerstaste startest, warte eine Minute, um dich an die Stille zu
gewöhnen.)*

Experiment: Paarvergleiche zur Beurteilung der Tonhöhe (Sinustöne aus Vorexperiment II, binaurale Reize, oktav-komplexe Töne aus dem Experiment zum Tritonusparadox und den Experimenten zum vollständigen Paarvergleich)

Instruktion

Es werden Topaare dargeboten.

*Drücke die Hoch-Taste,
wenn du die Tonhöhe als **aufsteigend** wahrnimmst.*

*Drücke die Runter-Taste,
wenn du die Tonhöhe als **absteigend** wahrnimmst.*

Um zu beginnen drücke die Leertaste.

Abbildungen zum vollständigen Paarvergleich zum Einfluss der Teiltonanzahl

Für jede Versuchsperson sind die relativen Häufigkeiten der Antwort 'Tonhöhe ist aufsteigend' je Startton und Distanz dargestellt (linke Seite) und die Ergebnisse der Anpassung des P4-Modells (rechte Seite). Die fünf Bedingungen (Teiltonanzahl) finden sich untereinander. Die Abkürzungen der Parameter sind wie folgt zu lesen:

A Amplitude (α)

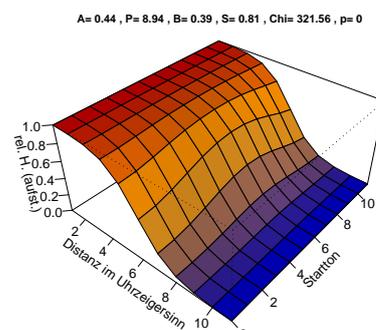
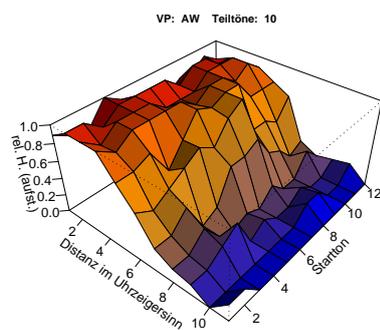
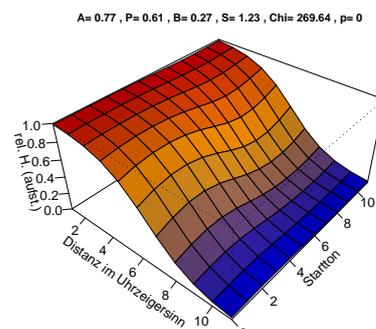
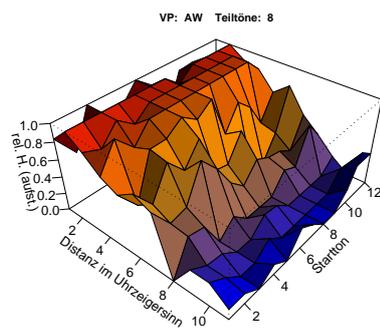
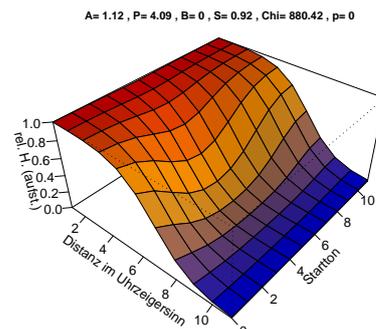
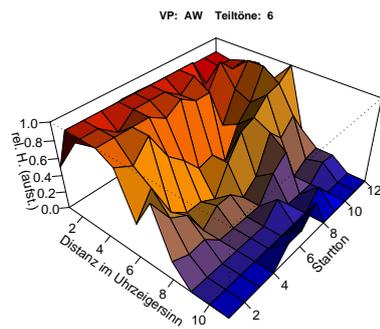
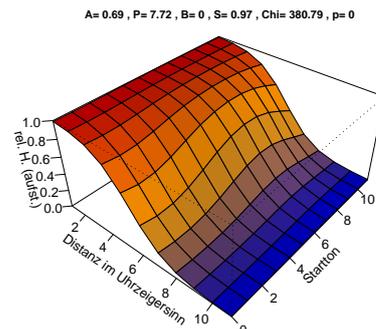
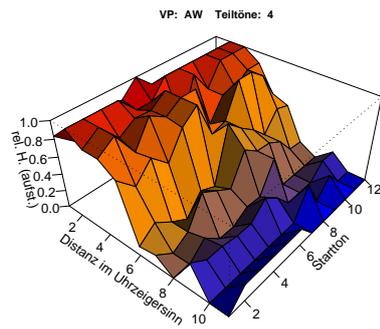
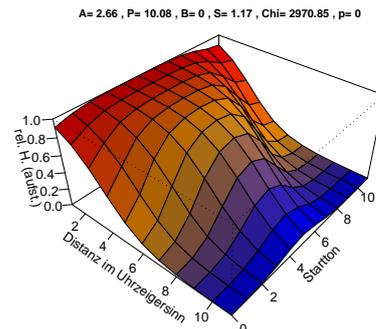
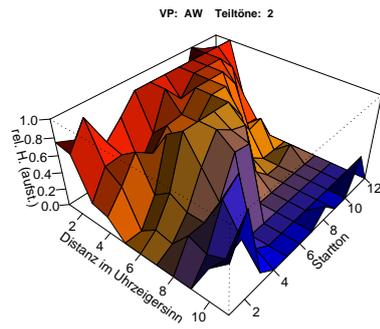
P Phase (ϕ)

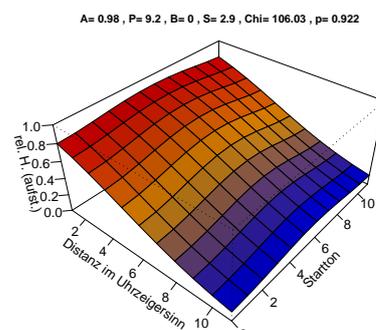
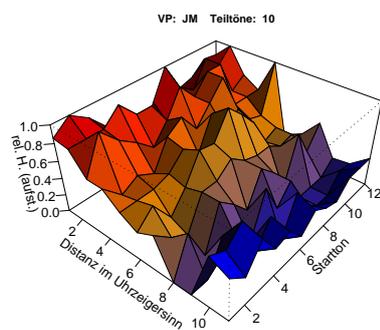
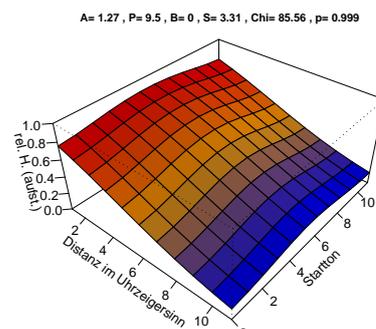
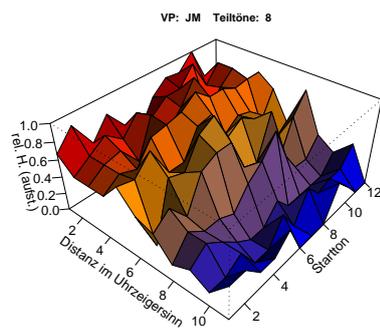
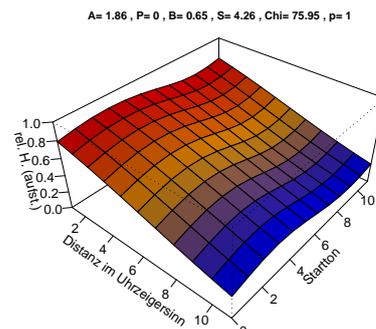
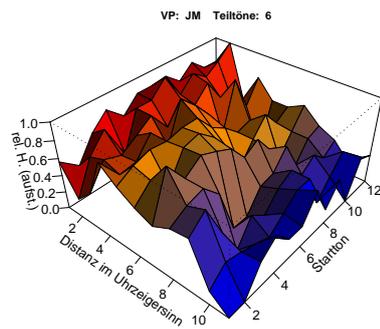
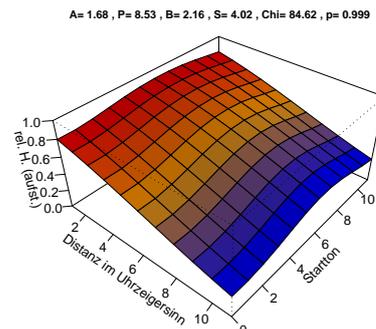
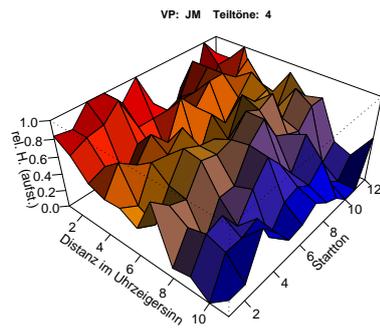
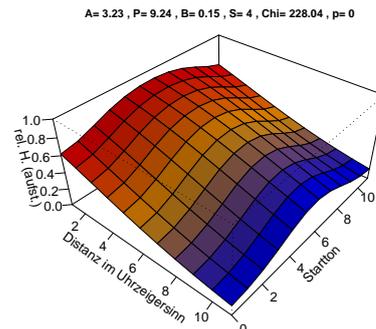
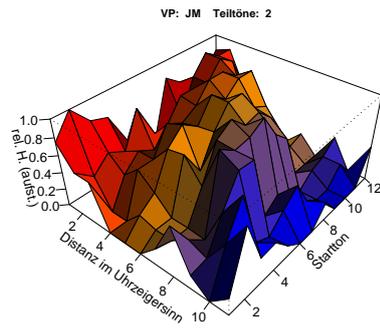
B Bias (σ)

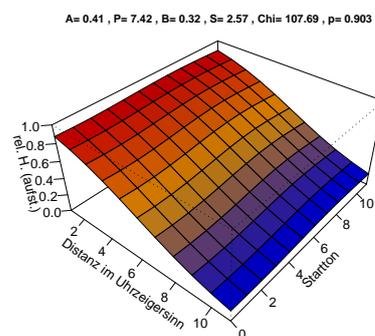
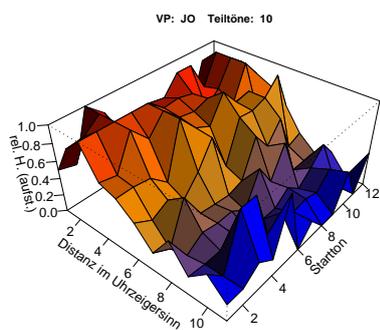
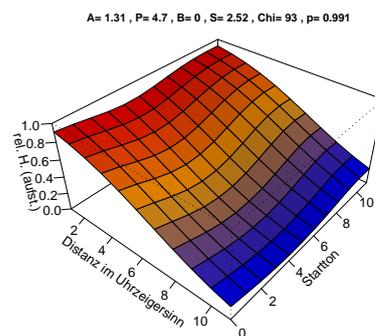
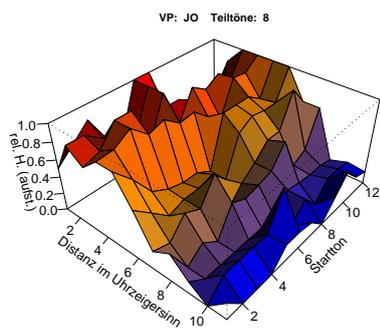
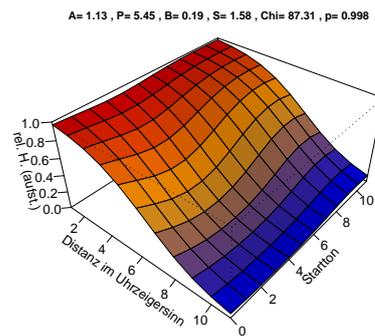
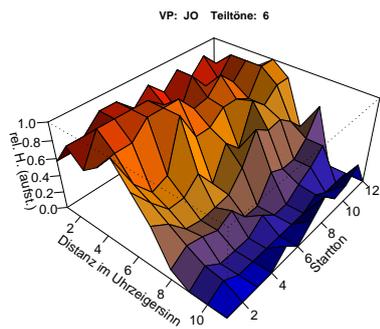
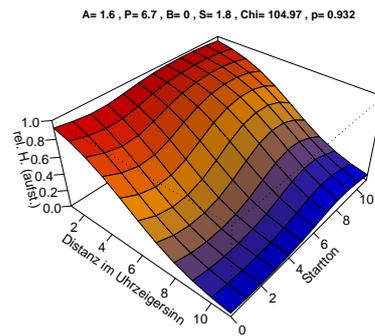
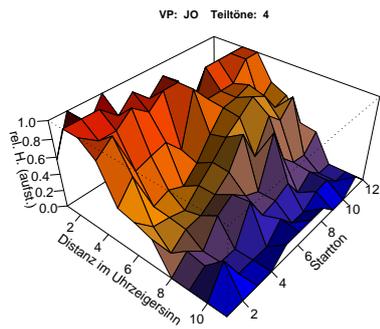
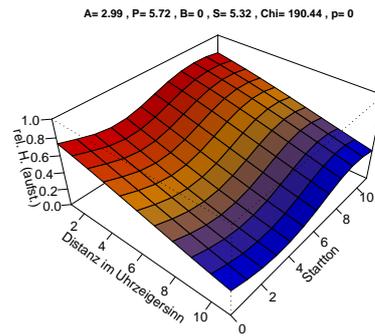
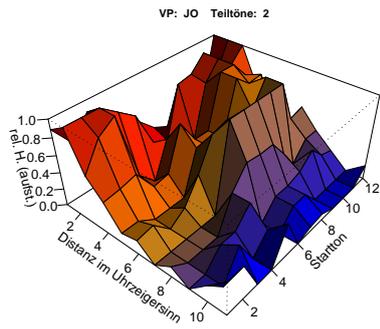
S Steigung (β)

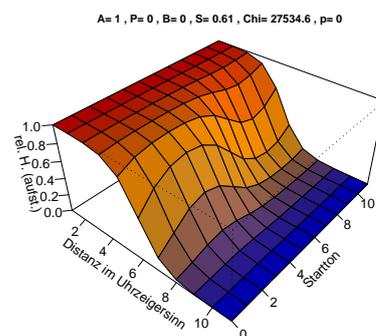
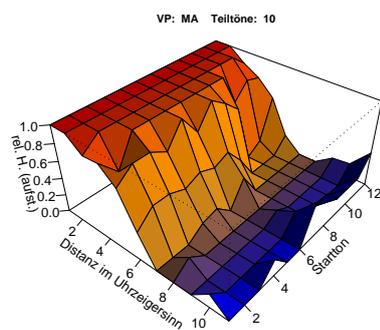
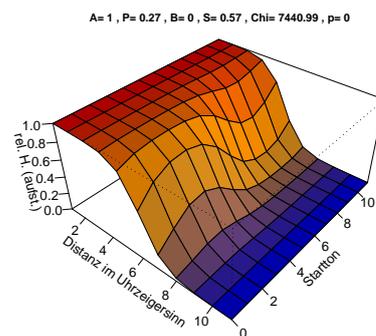
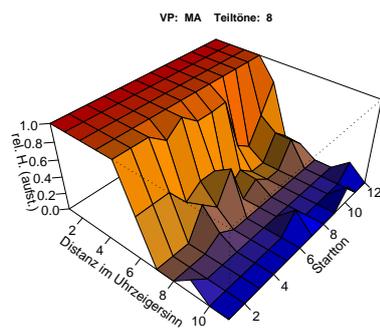
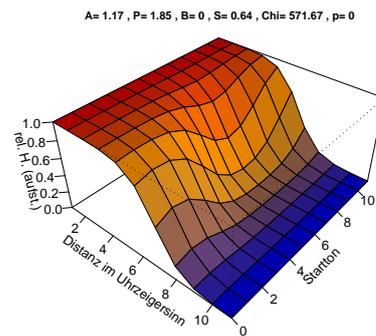
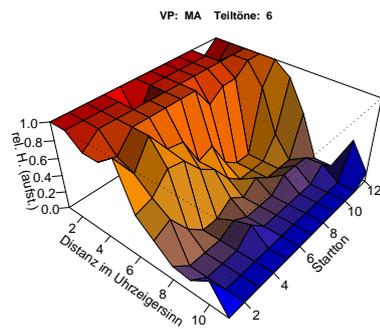
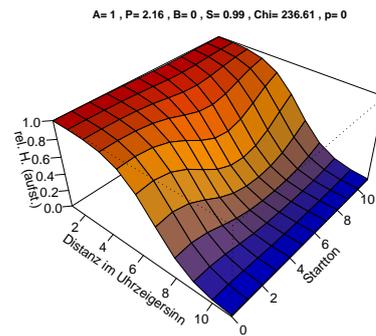
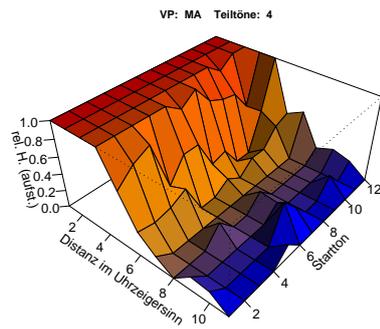
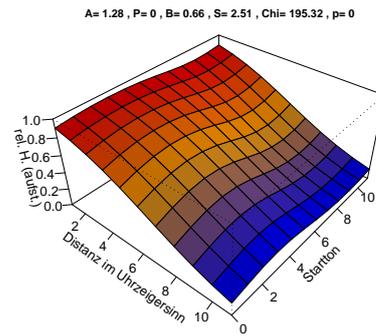
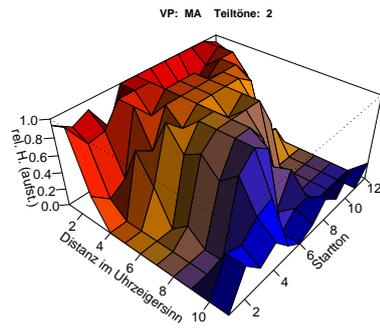
C Chi-Quadrat Statistik (Abweichung von Daten zur Schätzung)

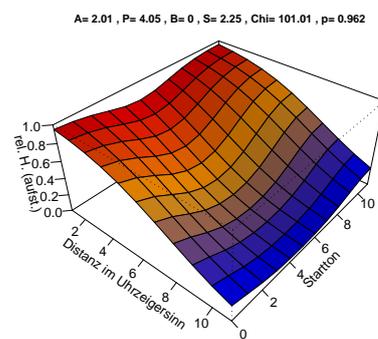
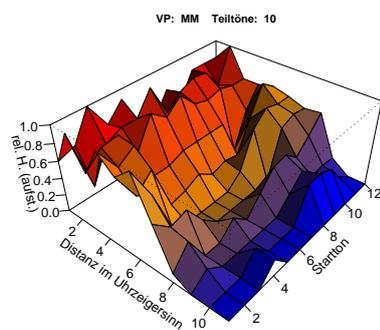
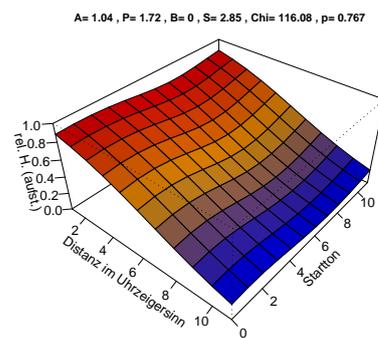
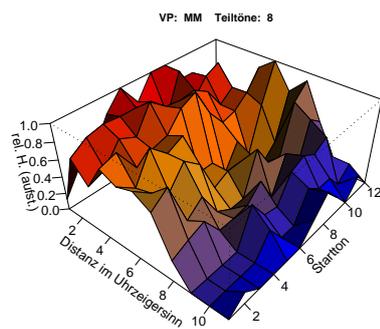
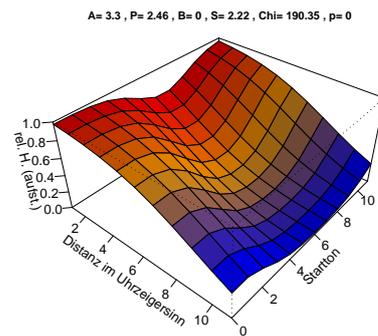
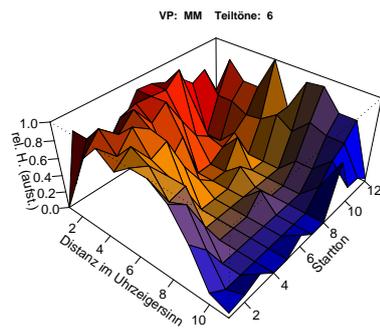
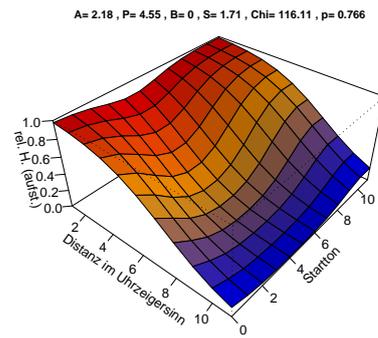
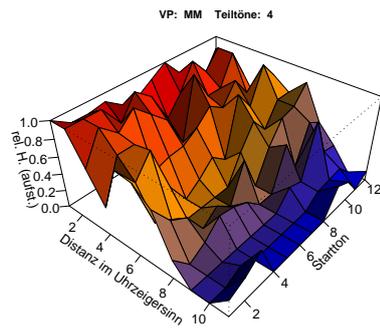
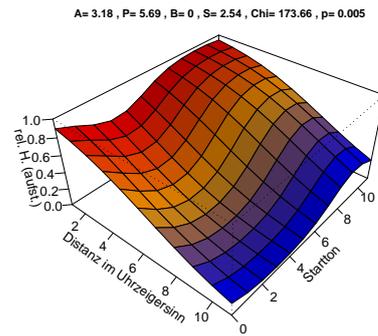
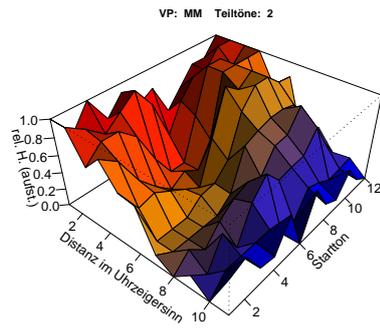
p Wahrscheinlichkeit, diese oder extremere Abweichungen zu erhalten

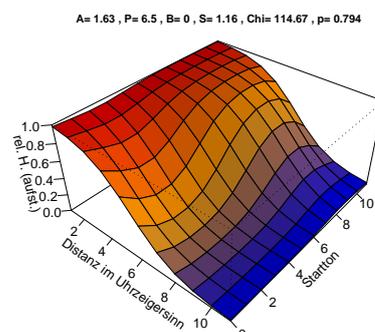
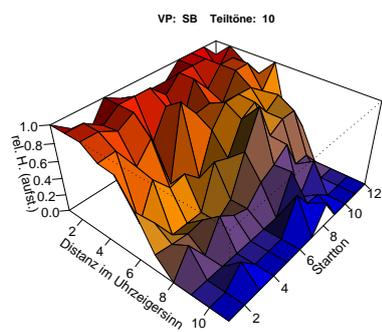
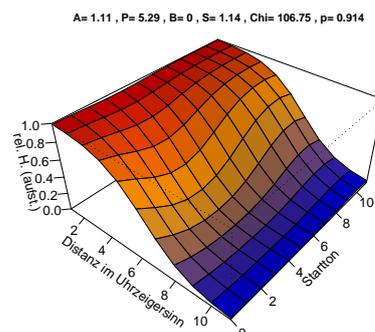
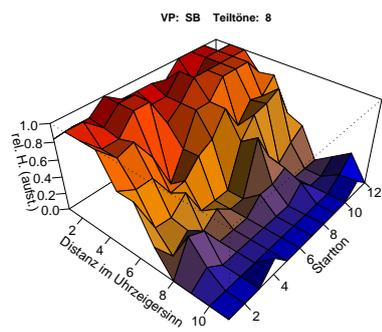
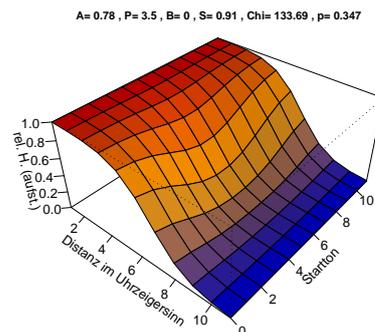
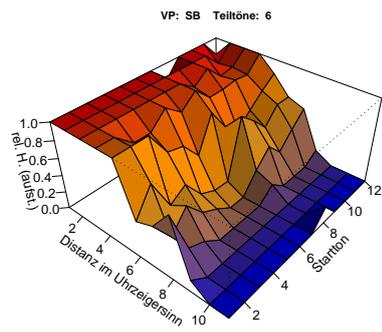
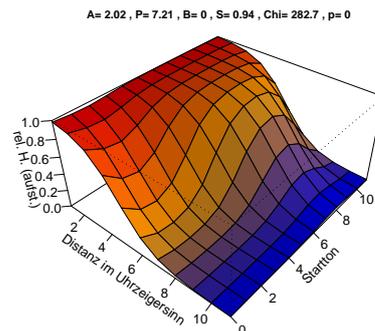
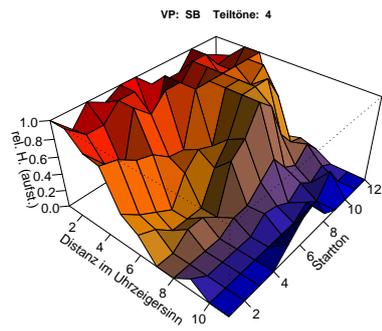
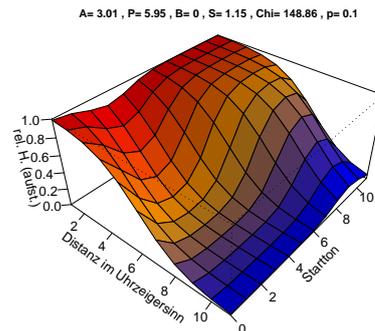
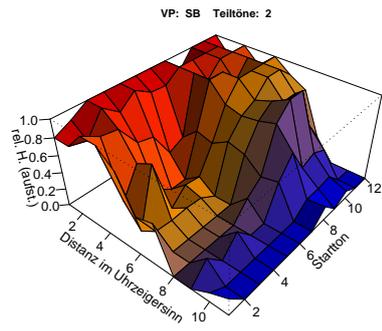


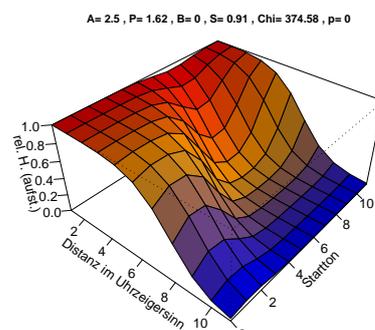
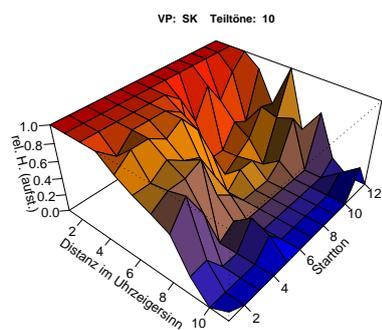
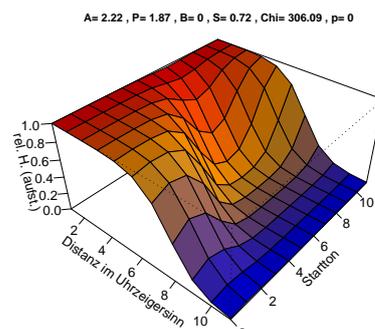
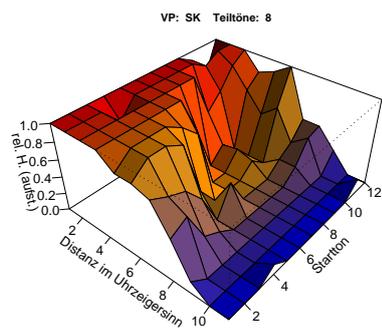
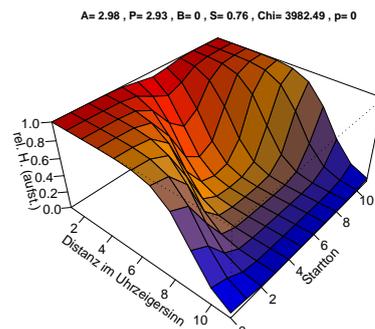
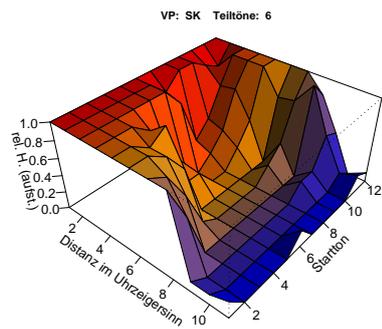
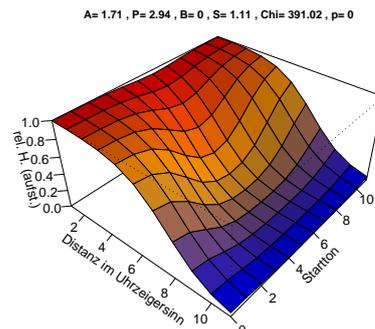
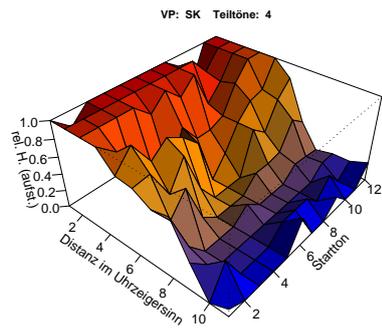
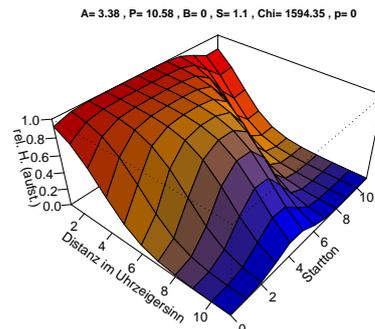
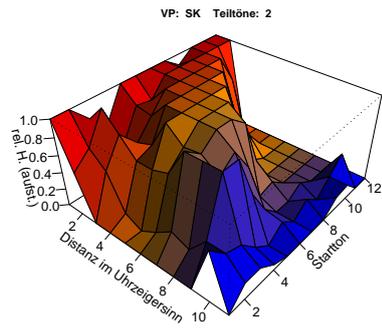


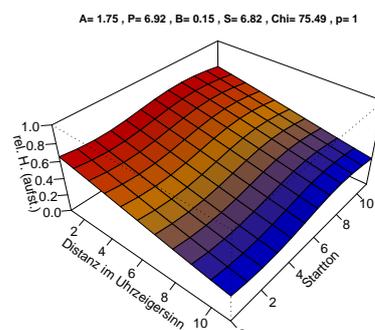
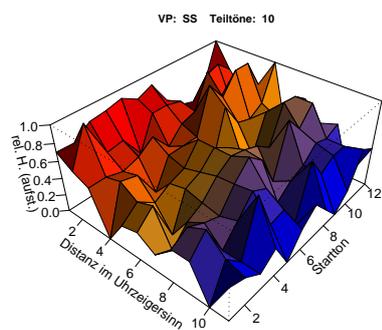
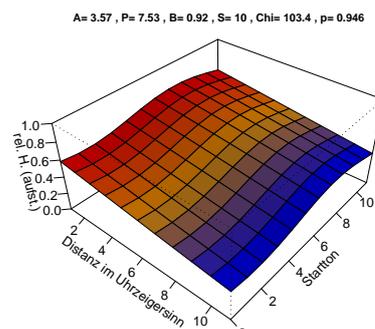
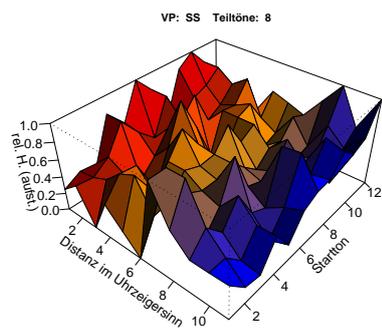
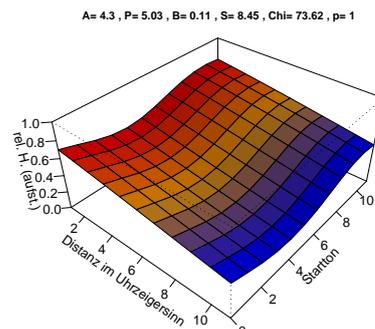
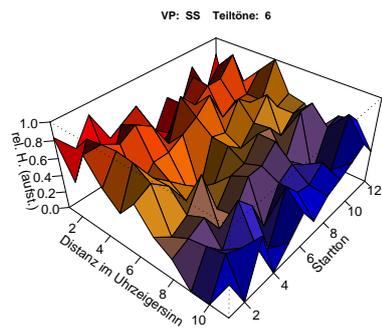
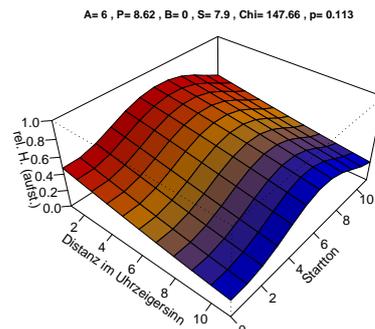
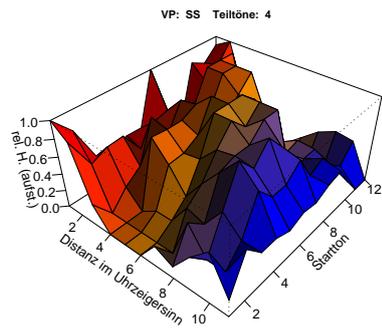
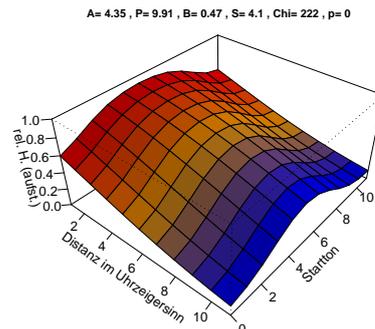
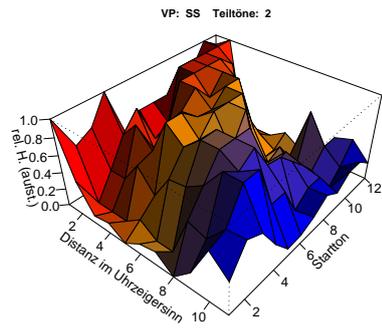


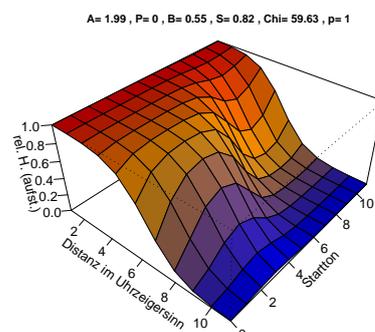
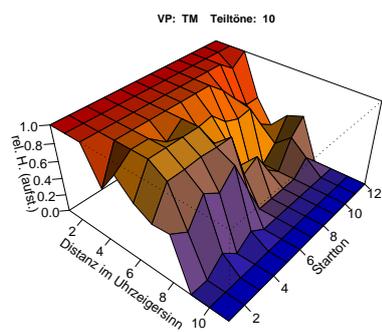
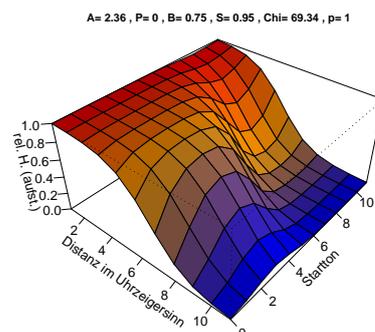
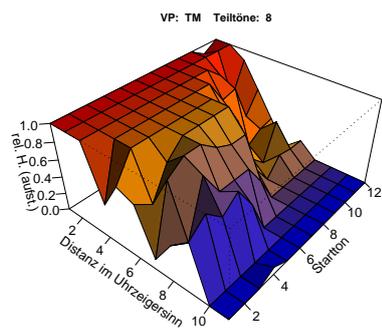
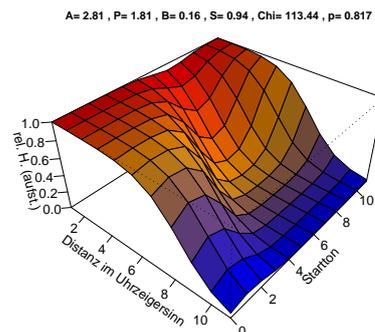
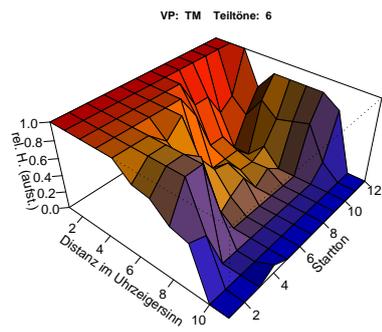
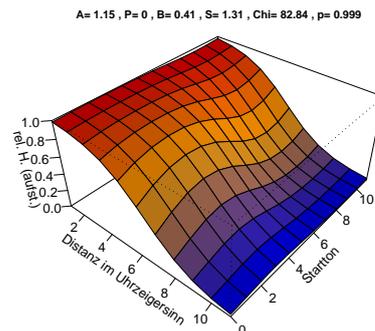
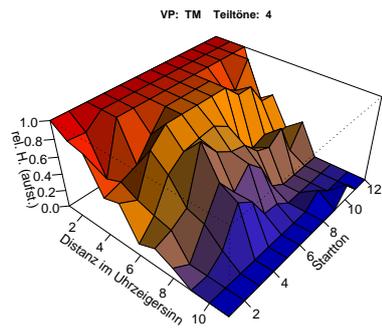
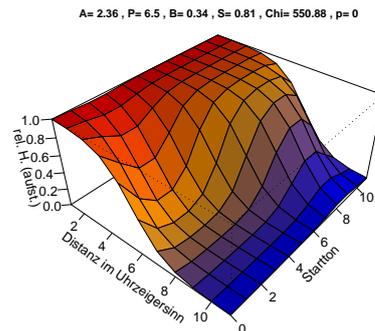
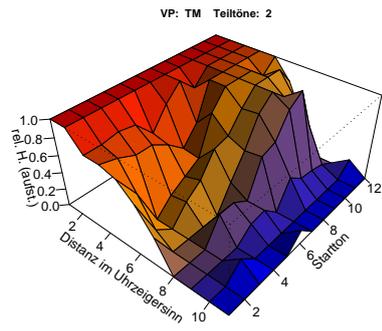


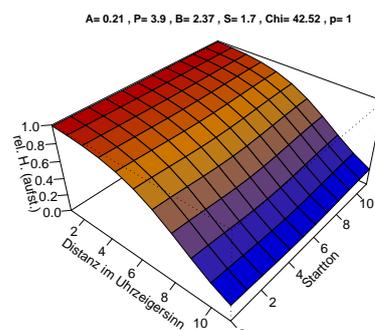
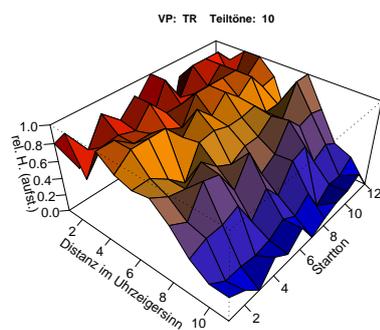
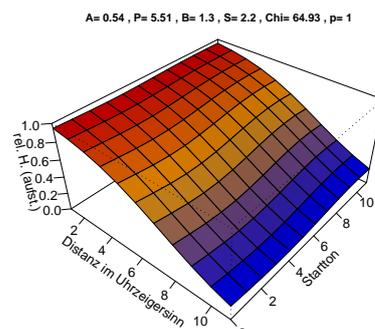
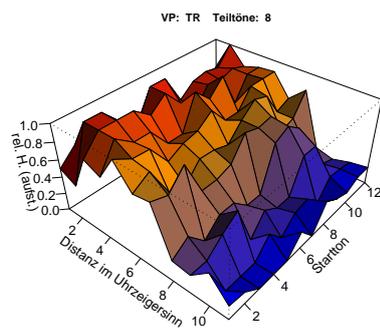
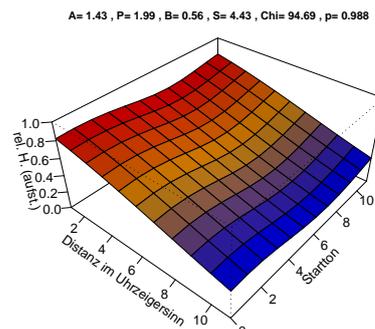
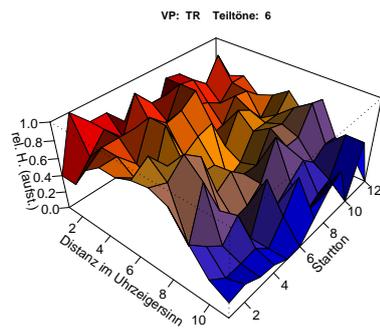
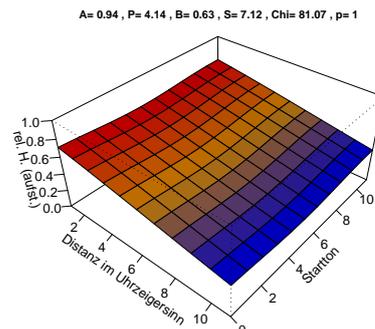
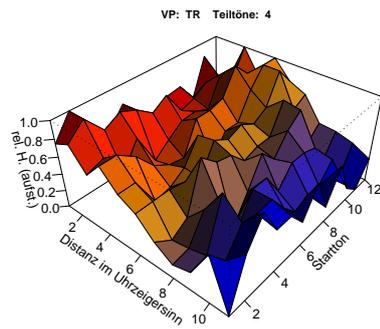
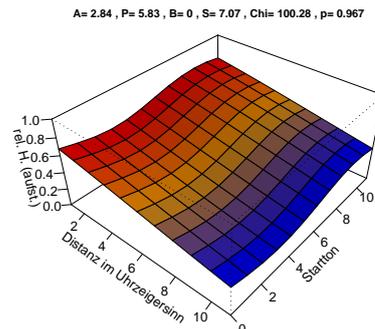
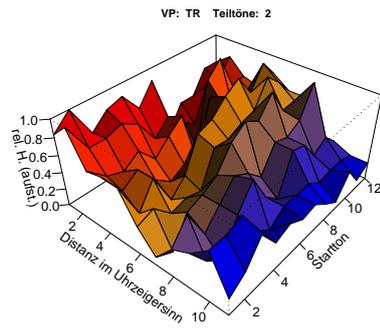












**Abbildungen zum vollständigen Paarvergleich zum
Einfluss der Hüllenposition und -form**

Für jede Versuchsperson sind die relativen Häufigkeiten für
Startton und Distanz dargestellt.

