

# Akzentuierung von Teiltönen in sprachähnlichen Schallen

STEFAN WARTINI

Fachgebiet Akustische Kommunikation, Technische Universität München

In der Regel wird ein harmonischer Klang durch ganzheitliche (synthetische) Wahrnehmung als einzelnes Ereignis (akustisches Objekt) wahrgenommen. Veränderungen einzelner Spektralanteile dieses Klanges können eine getrennte Wahrnehmung von Einzelkomponenten ermöglichen. Diese Verbesserung der Wahrnehmbarkeit (Akzentuierung) einzelner Schallanteile läßt sich durch Pegelveränderungen, Inharmonizitäten und unterschiedliche Einsatzzeitdifferenzen hervorrufen.

Betrachtet man Pegelveränderungen einzelner Komponenten in Schallen, so lassen sich zwei Arten von Akzentuierung unterscheiden. Zum einen wird ein einzelner Teilton oder ein ganzer Spektralbereich durch Pegelanhebung, also durch mehr oder weniger schnelles Auftauchen in einer akustischen Szene hervorgehoben. Zum anderen kann die Wahrnehmung von Klangkomponenten durch die Nachwirkung vorangegangener Lücken in diesen Spektralbereichen auch über eine Pause hinweg verbessert werden. Summerfield [2] zeigte, daß die Darbietung eines Klanges, dessen Teiltöne im Bereich der Formanten eines Vokals abgesenkt wurden, in einem nach kurzer Pause folgenden Klang mit glatter Spektralhüllkurve die Wahrnehmung eben dieses Vokals hervorrufen kann.

Die Absenkung der Mithörschwellen einzelner Harmonischer in einem komplexen Klang durch vorherige Darbietung desselben Klanges, bei dem diese Harmonischen unterdrückt wurden, kann als Maß für die Akzentuierung angesehen werden. Ein derart akzentuierter Teilton ruft seinerseits eine erhöhte Nachverdeckung hervor [4]. Dies legt die Vermutung nahe, daß hier ein Prozeß vorliegt, der nicht nur die Aufmerksamkeit des Beobachters leitet, sondern tatsächlich die Verstärkung im auditorischen Signalweg beeinflußt. Es ist möglich, daß die Akzentuierung bei der Wahrnehmung von Sprache zur spektralen Kontrastanhebung dient und somit den höheren Verarbeitungstufen des Wahrnehmungsapparates eine verbesserte Repräsentation des Signals zur Verfügung stellt.

In der vorliegenden Arbeit werden Untersuchungen der Mithörschwellen einzelner Teiltöne in vokalähnlichen Maskiererschallen in Abhängigkeit eines vorangehenden, das Gehör vorrerregenden Schalles (*Adaptor*) beschrieben. *Adaptor* und *Maskierer* besaßen stets dieselbe Spektralhüllkurve und gleiche Grundfrequenz. Neben dem Einfluß einer spektralen Lücke im *Adaptor* bei der Testtonfrequenz wurde auch der Einfluß unterschiedlicher Verzögerungszeiten der Teiltöne im *Maskierer* auf die Mithörschwellen untersucht.

## Versuchsschalle

Adaptoren und Maskiererschalle waren synthetisch erzeugte Vokale, bei denen der jeweils untersuchte Teilton entfernt wurde. Die Hüllkurven der Vokale /a/, /e/, /i/, /o/ und /u/ wurden nach einer auf dem Vokaltraktmodell von Fant [1] basierenden Formel [3] bestimmt. Tab. 1 zeigt die verwendeten Formantfrequenzen.

| Vokal | 1. Formant | 2. Formant | 3. Formant | 4. Formant |
|-------|------------|------------|------------|------------|
| /a/   | 800 Hz     | 1200 Hz    | 2800 Hz    | 3500 Hz    |
| /e/   | 400 Hz     | 2300 Hz    | 2900 Hz    | 3500 Hz    |
| /i/   | 170 Hz     | 2000 Hz    | 3000 Hz    | 3350 Hz    |
| /o/   | 330 Hz     | 660 Hz     | 2650 Hz    | 3120 Hz    |
| /u/   | 300 Hz     | 800 Hz     | 2200 Hz    | 3500 Hz    |

Tab. 1: Formantfrequenzen, welche zur Bestimmung der spektralen Hüllkurven der Vokale nach dem Vokaltraktmodell von Fant verwendet wurden.

Zum Erzeugen der spektralen Lücke in den Schallen wurden für die einzelnen Versuche entweder die Teiltöne im ersten Formanten, welche den höchsten Pegel besitzen oder die jeweils nächsthöhere Harmonische aus den Vokalen entfernt. Adaptoren und Maskiererschalle wurden mit den Grundfrequenzen 100 Hz und 200 Hz durch Fouriersynthese erzeugt. Alle Teiltöne der harmonischen Klänge besaßen dieselbe Startphase (Sinusphase), wodurch der zeitliche Verlauf von Adaptor und Maskierer dem der Glottisschwingung stark ähnelt. Die Dauer von Adaptor und Maskiererschall betrug 600 ms. Die Schalle wurden mit gaußförmigen Flanken der Anstiegs- bzw. Abfallszeit 5 ms versehen. Die Testtöne wurden mit unverzögerter Darbietung sowie mit einer Verzögerung von 200 ms gegenüber dem Maskiererbeginn erzeugt. Bei den Messungen mit 200 Hz Grundfrequenz wurden zusätzlich die Mithörschwellen für die Verzögerungszeit 10 ms ermittelt. Der Testton endete immer gemeinsam mit dem Maskierer, sodaß seine Dauer je nach Verzögerungszeit zwischen 600 ms und 400 ms variierte. Um bei den unterschiedlichen Verzögerungszeiten der verwendeten Testtöne definierte Phasenbeziehungen zwischen Maskierer und Testton zu erreichen, wurden die Testtöne zunächst mit derselben Dauer und Startphase wie die Maskiererschalle berechnet. Die benötigte Verzögerungszeit wurde durch Nullsetzen der entsprechenden Abtastwerte zu Beginn des Testtones erreicht. Auf diese Weise besitzt das Zeitsignal des Maskierers während der Testtondarbietung denselben Verlauf wie das Zeitsignal des vollständigen Vokals.

### Hörversuche

An den Hörversuchen nahmen insgesamt fünf normalhörende Versuchspersonen teil. Die Darbietung der Schalle erfolgte diotisch in einer schallisolierten Meßkabine über einen freifeldentzerrten Kopfhörer (Beyer DT48). Der Darbietungspegel wurde so eingestellt, daß die vollständigen harmonischen Spektren mit einem Schalldruckpegel von 80 dB wiedergegeben wurden. Die Hörversuche wurden nach der Einregelungsmethode durchgeführt. In einer ersten Versuchsreihe wurden den Versuchspersonen in zwei durch eine Pause von 300 ms Dauer getrennten Intervallen ein Vokal mit Lücke und derselbe Vokal ohne Lücke dargeboten (vgl. Fig. 1).

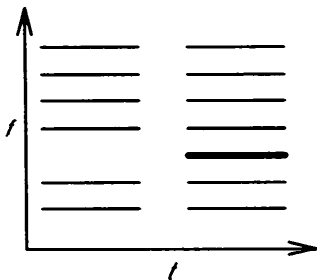


Fig. 1: Zeitliche Abfolge von Adaptor und Maskiererschall bei den Hörschwellenmessungen. Die Darbietung wurde solange wiederholt, bis die Versuchsperson den Pegel des Testtones im zweiten Darbietungsintervall (dicke Linie) so eingeregelt hatte, daß sie die Tonhöhe des Testtones gerade nicht mehr wahrnehmen konnte.

Die Darbietung des Schallpaares wurde solange wiederholt, bis die Versuchspersonen den Pegel des betreffenden Testtons so eingeregelt hatten, daß sie seine Tonhöhe gerade nicht mehr wahrnehmen konnten. Die Pegelveränderung des Testtones erfolgte durch Drücken zweier Tasten. Das Erreichen der Hörschwelle signalisierten die Versuchspersonen mit einer weiteren Taste. In weiteren Versuchsreihen wurde der dem Maskierer hinzugefügte Testton mit verschiedenen Verzögerungszeiten versehen. Die Versuchsreihen wurden mit vollständigen Vokalen als Adaptoren wiederholt, um zwischen dem Einfluß der Verzögerung des Teiltöns und dem der Akzentuierung durch die Lücke im Adaptor unterscheiden zu können.

Bei der Grundfrequenz 200 Hz konnten die Messungen mit dem Vokal /i/ nicht durchgeführt werden, da hier die Frequenz des Testtons mit der Grundfrequenz des Vokals zusammenfallen.

Der Adaptor hätte durch das Entfernen des Teiltöne keine Lücke erhalten, sondern wäre nach tiefen Frequenzen hin bandbegrenzt worden.

## Versuchsergebnisse

Da nicht alle Versuchspersonen an allen Hörversuchen teilnahmen, sind im Folgenden die Zentralwerte der Ergebnisse von vier (Grundfrequenz  $f_0 = 100$  Hz) bzw. drei ( $f_0 = 200$  Hz) Versuchspersonen angegeben. Die Reproduzierbarkeit der Messungen der Versuchspersonen lag in der Regel bei  $\pm 2$  dB, jedoch ergaben sich zum Teil erhebliche interindividuelle Schwankungen.

### Unverzögerte Testtöne

In Fig. 2 ist die Pegelabsenkung des unverzögert dargebotenen Testtones gegenüber seinem Originalpegel im Vokal dargestellt.

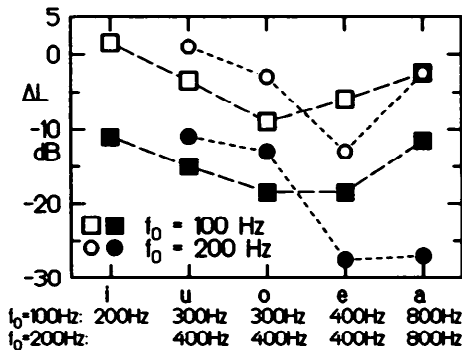


Fig. 2: Zentralwerte der Pegeldifferenz zwischen der Hörschwelle des Testtons und seinem ursprünglichen Pegel im Vokal. Die Hörschwellen wurden für Adaptoren mit (ausgefüllte Symbole) und ohne (leere Symbole) spektraler Lücke bei verschiedenen Grundfrequenzen bestimmt. Die Testtöne wurden unverzögert in den Maskierer eingefügt. Die Frequenzen geben die verwendete Testtonfrequenz (im ersten Formanten des jeweiligen Vokals) an. Die Verbindungen zwischen den Symbolen dienen der Verdeutlichung der unterschiedlichen Versuchsbedingungen.

Die Abszisse kann als nichtlineare Frequenzachse interpretiert werden. Die Verbindungen zwischen den Meßwerten stellen jedoch keine Interpolation dar, sondern sollen die Zugehörigkeit der Meßpunkte zu den verschiedenen Versuchsparametern verdeutlichen. Je nachdem, ob ein Vokal mit oder ohne Lücke (ausgefüllte bzw. leere Symbole) als Adaptor dargeboten wird, ergeben sich erhebliche Unterschiede zwischen den Mithörschwellen des Testtones. Weist der Adaptor eine Lücke auf, so sinkt die Mithörschwelle des Testtones im Maskierer bei der Grundfrequenz 100 Hz (quadratische Symbole) um 9 bis 13 dB gegenüber der Mithörschwelle bei einem Adaptor ohne Lücke. Bei Vokalen der Grundfrequenz 200 Hz (kreisförmige Symbole) ergeben sich Absenkungen von 10 bis 25 dB.

Um der Frage nachzugehen, welchen Einfluß eine Maskierung durch benachbarte Teiltöne auf die Mithörschwellen hat, wurden Messungen mit dem zum Teilton mit höchstem Pegel im ersten Formanten nächsthöheren Teilton durchgeführt. Bei den Messungen, die mit Schallen der Grundfrequenz 100 Hz durchgeführt wurden zeigte der Vergleich der Mithörschwellen bei Adaptoren mit und ohne Lücke mit Ausnahme des Vokals /o/ keine Unterschiede. Beim Vokal /o/ senkte sich die Mithörschwelle des Testtones durch die Darbietung des Adaptors mit Lücke um 6 dB ab. Während die Mithörschwellen der Teiltöne mit maximalem Pegel im 1. Formanten bei Adaptoren ohne Lücke unter den ursprünglichen Pegeln der Teiltöne liegen, sind zum Erreichen der Mithörschwellen der nächsthöheren Teiltöne selbst bei Adaptoren mit Lücke Pegelanhebungen nötig. Die Maskierung durch den benachbarten Teilton scheint so hoch zu sein, daß eine Akzentuierung des Testtones unmöglich wird. Einen Hinweis darauf liefert der Vokal /o/. Während bei allen anderen Vokalen die Pegeldifferenz zwischen dem niedrigeren Nachbaranteilton und dem Testton z.T. weit über 10 dB liegt, beträgt sie beim Vokal /o/ nur 7 dB.

### Verzögerte Testtöne

Fig. 3 zeigt die Abhängigkeiten der Mithörschwellen von der Verzögerungszeit des hinzugefügten Testtones. Die Mithörschwellen von verzögert einsetzenden Testtönen liegen sowohl bei der vorausgehenden Darbietung eines Adaptors mit Lücke als auch ohne Lücke deutlich unter denen, die sich bei unverzögerter Darbietung ergeben. Bei Vokalen der Grundfrequenz 100 Hz betrug die Absenkung durch die Verzögerung zwischen 2 und 6 bzw. 7 und 15 dB, je nachdem ob der Adaptor eine Lücke besaß oder nicht. Bei einer Grundfrequenz von 200 Hz wurden Werte zwischen 7 und 12 bzw. 15 und 24 dB ermittelt. Wird ein Vokal ohne Lücke als Adaptor dargeboten, so sinkt die Mithörschwelle mit zunehmender Verzögerung des Teiltones stark ab. Die Verzögerung des Testtones um 200 ms erzeugt im Maskierer eine spektrale Lücke, die zu einer ähnlich großen Absenkung der Mithörschwelle führt, wie die Darbietung eines Adaptors mit Lücke und unverzögertem Teilton im Maskierer, der 300 ms nach dem Adaptor dargeboten wird. Wird dagegen ein Adaptor mit Lücke dargeboten, dann bewirkt die Verzögerung des Testtons im Maskierer eine geringere zusätzliche Absenkung der Mithörschwelle.

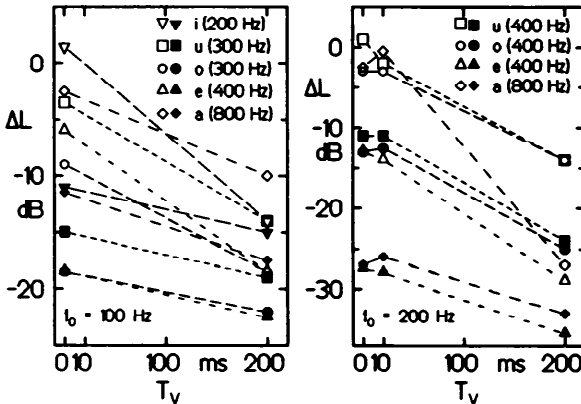


Fig. 3: Abhängigkeit der Mithörschwellen von der Verzögerungszeit eines Teiltones bei den Grundfrequenzen 100 Hz (links) und 200 Hz (rechts). Aufgetragen ist die Pegelabsenkung gegenüber dem ursprünglichen Teiltonpegel für verschiedene Vokale mit (ausgefüllte Symbole) und ohne (leere Symbole) spektraler Lücke in der vorangegangenen Darbietung des Vokals.

Zusätzliche Messungen zeigten, daß eine geringere Anstiegsgeschwindigkeit des verzögerten Testtones im Maskierer zu einer leichten Verringerung der Mithörschwellenabsenkung führt.

Die Untersuchungen zeigen, daß eine Akzentuierung einzelner Spektralbereiche durch vorangegangene Schalle, die in diesen Bereichen keine Signalanteile besitzen, auch über Pausen von 300 ms hinaus wirkt. Eine zusätzliche Verzögerung einzelner Komponenten eines Klanges bewirkt eine ähnliche Akzentuierung und damit Absenkung der Mithörschwelle, die sich derjenigen überlagert, welche durch den Adaptor verursacht wird.

Die Untersuchungen wurden im SFB 204 „Gehör“ mit Unterstützung der DFG durchgeführt.

## Literatur

- [1] Fant, G. (1960). Acoustic theory of speech production. Mouton, The Hague. (2. Auflage 1970)
- [2] Summerfield, Q., Haggard, M., Foster, J., Gray, S. (1984). Perceiving vowels from uniform spectra: phonetic exploration of an auditory aftereffect. In: *Perception & Psychophysics* 35, S. 203-213.
- [3] Terhardt, E. (1979). Calculating virtual pitch. In: *Hearing Research* 1, S. 155-182.
- [4] Viemeister, N. und Bacon, S. (1982). Forwardmasking by enhanced components in harmonic complexes. In: *J. Acoust. Soc. Am.* 71, S. 1502-1507.