

Verh. 5. int. Kongr. Phon. Wiss., Münster 1964, pp. 578-581
(S. Karger, Basel/New York 1965).

Die Tonhöhen-Unschärfe von Sprachlauten

Von F. WINCKEL, Berlin

Der Grund für den geringen Erfolg der Sprechmelodie-Forschung ist darin zu erblicken, daß Sprechtonhöhen absolut und auch in ihren Intervallbeziehungen in laufender Rede auditiv kaum erkannt werden, andererseits sind die psychoakustischen Zusammenhänge von Tonhöhenverläufen bisher wenig geklärt. Dies erscheint verwunderlich, da Tonhöhenmesser von hoher Meßgenauigkeit seit langem vorhanden sind.

Der Tonhöhenumfang in normaler Unterhaltung umfaßt eine Oktave, in der Emphase bis zu zwei Oktaven. Erstaunlich ist, daß der Sprechumfang von einer Oktave kaum bemerkt wird, während einfache Kinderlieder vom Umfang einer Quinte sofort eine einprägsame Melodie ergeben. Tonhöhenverläufe in der Sprache haben die Tendenz einer gewissen Stetigkeit, sprunghafte Intervalle werden möglichst vermieden. Aus einem auf Tonband gesprochenen Text hat der Verfasser Schnitte von je 100 ms angefertigt und diese stochastisch wieder zusammengesetzt. Es ist nunmehr ein deutliches Intervallspringen wahrzunehmen, das man wegen des statistischen Verhaltens natürlich nicht als «Melodie» bezeichnen darf. Auffallend ist in der stochastischen Fassung der mit wachsendem Intervall seltener werdende Anteil höherer Töne, die demnach ziemlich isoliert aus dem übrigen Verlauf hervorstechen. Aus der stochastischen Behandlung lassen sich somit eher Individualkonstanten herleiten.

Ein weiterer Grund für die schlechte Wahrnehmbarkeit der Sprachmelodie liegt darin, daß die Konsonanten, die nicht als Tonhöhen bewertet werden, in ihrem ständigen Wechsel mit den Vokalen einen hohen Prozentsatz ausmachen. Die Vokaldauern bewegen sich in der Größenordnung von 50 bis 200 ms, was grundsätzlich unterhalb der Tondauern in musikalischen Abläufen liegt.

Die Differential-Empfindlichkeit des Ohres wird festgestellt beim Vergleich von vibratoüberlagerten sinusartigen Tönen von hinreichender Dauer. Von verschiedenen Autoren werden übereinstimmend 3‰ Frequenzschwankung für Sinustöne im Bereich der Sprechgrundtonhöhen als gerade noch bemerkbar angegeben. Macht man den Versuch mit Tonimpulsen z.B. von 250 Hz, so wird man bei einer Impulsdauer von 200 ms gerade noch 3 Hz Abweichung unterscheiden können, wogegen bei einer Dauer von 5 ms das Unterscheidungsvermögen gemäß Abbildung 1 zehnmal schlechter

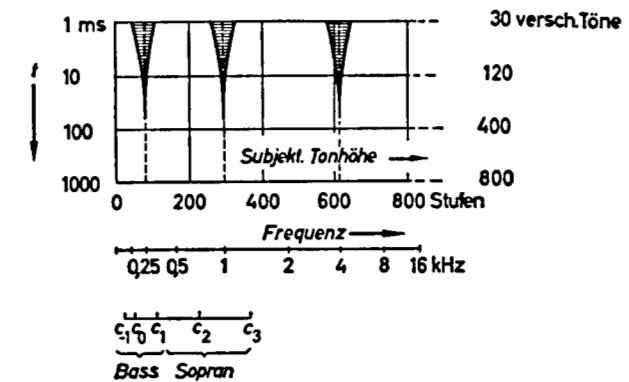


Abb. 1. Hörschärfe für Töne kurzer Dauer. Mit wachsender Dauer (Ordinate) nimmt die Hörschärfe zu, unabhängig vom Grundton, gezeigt für 250, 1000 und 4000 Hz.

wird (Feldtkeller). Dies ist ein Grund dafür, warum die gegenüber Musiklauten relativ kürzeren Sprachlaute im Sprechablauf so wenig melodiebildend wirken. Ferner werden Vokallaute niemals mit konstanter Tonhöhe intoniert. Anstieg und Abfall ereignen sich in großer Häufigkeit mit einer Gleitgeschwindigkeit von 1000 Hz/sec, wobei Schwankungen in den Grenzen 500 bis 2000 Hz/sec und gelegentlich auch darüber hinaus vorkommen (Frequenzmodulation). Derartige Gleitgeschwindigkeiten können trotzdem noch als konstante Tonhöhen wahrgenommen werden, wenn nur die zeitlichen Ausschnitte aus den Gleitlauten genügend klein bleiben.

Der Verfasser hat Versuche angestellt, aus Gleittönen einer Posaune, die eine Geschwindigkeit von 0,112 Oktaven/sec haben, Ausschnitte herzustellen, die als Tonhöhen konstant beurteilt werden konnten. Es ergab sich eine Dauer von etwa 0,5 sec. Eine Tonhöhenunschärfe ist zunächst aus der Heisenbergschen Unschärfe-Relation abzuleiten, woraus eine physikalische Tonhöhen-Unbestimmtheit des Glissandotones hergeleitet werden kann (vgl. Abb. 2).

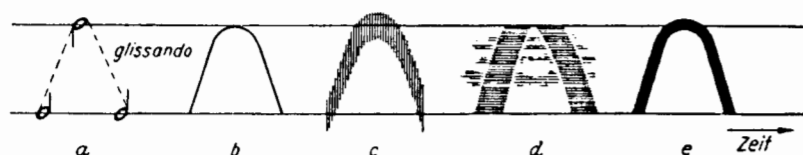


Abb. 2. Die Unbestimmtheit eines Glissandotones nach Zeit und Frequenz. a) Notierung, b) theoretischer Verlauf, c) bei exakter zeitlicher Bestimmung wird die Tonhöhenbestimmung ungenau, d) bei höchster Frequenz-Selektivität wird die zeitliche Abgrenzung ungenau, e) praktischer Fall der mittleren Zeit- und Frequenz-Ungenauigkeit (nach F. Winckel, Phänomene des musikalischen Hörens, Berlin 1960).

Wenn man die Frequenzmodulationsschwellen nach *Feldtkeller* und *Zwicker* heranzieht, so ergibt sich z.B. für eine Modulationsfrequenz von 16 Hz im Bereich eines 60 dB lauten Tones von 250 Hz eine Änderungsgeschwindigkeit von 80 Hz/sec. Hier handelt es sich um die periodische Wiederholung mit einem Frequenzhub von 2,5 Hz (Vibrato), also einer wesentlich geringeren Anforderung als bei der Vokalintonation mit extrem hoher Gleitgeschwindigkeit. Allerdings lagen den Messungen der Frequenzmodulationsschwellen Sinustöne zugrunde, dem Vokalmaterial dagegen obertonreiche Spektren mit 30 Teiltönen und mehr. Das gilt auch für den Posanenversuch. In beiden Fällen sind die Teiltöne nicht exakt Harmonische, so daß an der Differenztonbildung bzw. dem Residuum sich eine Verwischung der Grundtonhöhe ergibt, während im Falle streng Harmonischer sich eine Verschärfung der Tonhöhe ergeben hätte. Anhand von Sonagrammen nach *Stevens* und *House* läßt sich der unabhängig vom Grundtonverlauf verschieden geneigte Verlauf, z.B. des 1. und 2. Formanten zeigen. Bei Vibratotönen entstehen außerdem Seitenbänder.

Die erwähnten Daten nach *Feldtkeller* und *Zwicker* können daher nur in grober Annäherung für Vokalintonation herangezogen werden. Dem Sprachcharakter wird man besser gerecht, wenn man Ausschnitte aus weißem Rauschen anstelle von Sinustönen – nach *Zwicker* – heranzieht, weil wegen der Einschwingvorgänge der Vokale zwischen den Konsonanten das Oszillogramm mehr das Aussehen permanenter Ausgleichvorgänge (transients) hat. Für die Messung mit Rauschen ergibt sich ein Unempfindlichkeitsbereich der Tonhöhenänderung – bei einer Modulationsfrequenz von 4 Hz bis zu ± 10 Hz (vgl. hierzu den Beitrag dieses Kongresses: Perzeptive Grenzen der Phonem-Unterscheidung).

Bei einem derart schlechten Unterscheidungsvermögen des Ohres für Tonhöhen-Änderung bei Sprache im Gegensatz zur Dar-

bietung von Sinustönen sollte das Problem der Steuerung der Tonhöhe für synthetische Sprache keine Schwierigkeit sein. Es ergibt sich aber, daß eine Sprachübertragung ohne die volle Informationskapazität des Tonhöhenverlaufs unzureichend ist, d.h. für das Ohr einen leblosen maschinellen Eindruck ergibt. Bei Abschaltung der Tonhöhensteuerung – monotone Sprache – verliert die Sprache jeglichen Akzent, was durch unterschiedliche Tonlängen kaum auszugleichen ist.

Literatur

1. *Feldtkeller, R.* und *Zwicker, E.*: Das Ohr als Nachrichtenempfänger (Stuttgart 1956).
2. *Feldtkeller, R.*: Wechselbeziehungen zwischen Psychologie und Nachrichtentechnik. In: Aufnahme und Verarbeitung von Nachrichten durch Organismen (Stuttgart 1961).
3. *Kashikawa, T.* and *Sugimoto, T.*: Information rate of the pitch signal in speech. Speech Communication Seminar, Stockholm 1962.
4. *Winckel, F.*: Phänomene des musikalischen Hörens (Berlin 1960).
5. *Zwicker, E.*: Informationskapazität des Gehörs. *Acustica* 6: 365-381 (1956).

Adresse des Autors: Prof. Dr.-Ing. F. Winckel, Technische Universität, Hardenbergstraße 34, 1 Berlin 12 (Deutschland).