Prinzipien der Aufnahme und Verarbeitung von Information durch das Gehör

E. TERHARDT

Lehrstuhl für Elektroakustik der Technischen Universität München

Die Robustheit, mit welcher die von Sprach- beziehungsweise Musiksignalen getragene Information selbst unter ungünstigen akustischen Bedingungen vom Gehör aufgenommen wird, entzieht sich bisher weitgehend dem wissenschaftlichen Verständnis. Man kann dies daran erkennen, daß es bisher nicht gelungen ist, ein sprach- beziehungsweise musikerkennendes System zu konstruieren, welches auch nur annähernd mit den entsprechenden Leistungen des Gehörs konkurrieren kann. Angesichts der außerordentlichen Variabilität des akustischen Inputs des Gehörs (d.h., der beiden Ohrsignale) wird das Verständnisdefizit nach verbreiteter Ansicht darauf zurückgeführt, daß jene Leistung in hohem Maße von "zentralen" beziehungsweise "kognitiven" Vorgängen abhänge und daß die Komplexität dieser Prozesse deren Erforschung behindere. Dieser Standpunkt ist aber insofern wenig hilfreich, als er das Problem lediglich in unbekannte Regionen der zentralen Verarbeitung verschiebt, ohne einen Beitrag zur Lösung zu leisten oder einen Lösungsweg erkennen zu lassen. Insbesondere lenkt er möglicherweise davon ab, daß man auf keinen Fall um die Klärung der Frage herumkommt, welches die informationstragenden Merkmale der Ohrsignale sind.

Im vorliegenden Beitrag wird ein Ansatz zur Lösung des letzteren Problems skizziert. Er beruht auf einer Analyse der Rolle, welche Information in biologischen sensorischen Systemen spielt (vgl. Terhardt, 1989; 1991). Die Analyse ergibt im wesentlichen folgendes.

a) Information und Informationsverarbeitung werden dadurch charakterisiert, daß *Objekte* (kategoriale Einheiten, im Gegensatz zu kontinuierlichen Signalen) und *bedingte Entscheidungen* im Spiel sind.
b) Sensorische Informationsverarbeitung bildet die Voraussetzung für sinnvolle physische Reaktion; erstere ist als Bestandteil der letzteren aufzufassen.
c) Sensorische Informationsverarbeitung ist hierarchisch organisiert. (Dies scheint überhaupt ein notwendiges Merkmal jeglicher Art von Informationsverarbeitung zu sein.) Sie beginnt bereits in der Peripherie, also beim Gehör im Corti'schen Organ.
d) Psychophysikalisch-funktional gesehen wird die erste Stufe sensorischer Informationsverarbeitung durch die Bildung *primärer Konturen* gebildet. Die Konturbildung stellt die erste Abstraktions-, das heißt, Informationsgewinnungs-Stufe dar. Daraus folgt, daß diejenigen physikalischen Parameter des Stimulus informationstragend sind, welche die primären Konturen hervorrufen bzw. beeinflussen.
e) Wie man am Beispiel der Konturen visueller Gestalten erkennt, sind die Konturen insbesondere in dem Sinne informationstragend, daß sie auch unter ungünstigen und sehr unterschiedlichen physikalischen Bedingungen das Erkennen von Objekten der Umgebung ermöglichen.
f) Als die Primärkonturen des Gehörs sind die Spektraltonhöhen anzusehen, das heißt, die Tonhöhen der Teiltöne von Schallen.

Die Analyse ergibt demnach (unter anderem), daß den Teiltönen von Schallen im Sinne informationstragender Signalmerkmale eine hervorragende Bedeutung zukommt. Das Konzept, wonach man sich Klänge aus Teiltönen zeitvariabler Frequenzen und Amplituden zusammengesetzt denken kann — das "zeitvariante Fourier-Synthese-Modell" — ist zwar gerade in der Akustik seit langem eingebürgert; jedoch war es bisher weitgehend auf die

Rolle einer intuitiven, allgemeinen und unpräzisen Beschreibungsweise beschränkt. Einige seiner Implikationen seien wie folgt kurz beleuchtet.

Das zeitvariante Fourier-Synthese-Modell für ein Signal $s(t)$ wird durch die Formel

$$s(t) = \sum_{\nu=1}^{n} \hat{s}_\nu(t) \cos 2\pi f_\nu(t) t \tag{1}$$

beschrieben, wobei $\hat{s}_\nu(t)$ bzw. $f_\nu(t)$ die zeitvariante Amplitude bzw. Frequenz des ν-ten Teiltones bedeuten. Es ist einsichtig, daß man beispielsweise das zu einer beliebigen ein- oder mehrstimmigen musikalischen Tonfolge gehörende Schallsignal derart beschreiben kann.

Erfahrungsgemäß kann man eine Anzahl der Teiltöne auch dann noch hören, wenn das Signal eine Übertragungsstrecke mit erheblichen linearen Verzerrungen durchlaufen hat, beispielsweise im Konzertsaal. Es ist also evident, daß die betreffenden Teiltöne durch die Übertragung zwar in Amplitude und Phase verfälscht werden, jedoch ihre ursprünglichen Frequenzen beibehalten. Dies kann man durch die Beziehung

$$a(t) \approx \sum_{\nu=1}^{n} \hat{s}_\nu(t) |H(f_\nu, t)| \cos[2\pi f_\nu(t) t + \phi(f_\nu, t)], \tag{2}$$

ausdrücken, wobei das am Ohr wirksame Signal mit $a(t)$ bezeichnet wurde, und $H(f,t)$ die komplexe, im allgemeinen zeitvariante Übertragungsfunktion der Strecke mit dem Absolutbetrag $|H(f,t)|$ und der Phase $\phi(f,t)$ bedeutet.

Läßt man der Einfachheit halber die Weg-Laufzeit des Schalles außer acht, so kann man sagen, daß Gl.(2) umso genauer zutrifft, je kürzer die Impulsantwort der Übertragungsstrecke (ihr "Gedächtnis") im Vergleich zur Änderung der Amplituden und Frequenzen des ursprünglichen Signals $s(t)$ ist. M. a. W.: Wenn sich innerhalb der Dauer der Impulsantwort die genannten Parameter von $s(t)$ nicht merklich ändern, gilt Gl.(2) mit entsprechender Genauigkeit. Unter dieser Voraussetzung kann man sich das Ohrsignal $a(t)$ in der Tat aus "denselben" zeitvariablen Teiltönen zusammengesetzt denken, aus welchen $s(t)$ besteht. Ein wirksames Verfahren, aus dem Ohrsignal $a(t)$ "durch die Übertragungsstrecke hindurch" die Eigenschaften und das zeitliche Verhalten der Schallquellen zu erschließen, besteht demnach konsequenterweise in einer zeitvarianten Teiltonanalyse — insbesondere der Extraktion der Teiltonfrequenzen als Funktion der Zeit. In der Tat hat sich gezeigt, daß die Teiltonzeitmuster, welche man durch Spektralanalyse und nachgeschaltete "spektrale Konturisierung" gewinnen kann, die gesamte gehörrelevante Information beliebiger Audiosignale enthalten (vgl. Heinbach, 1988; Mummert, 1990).

Es ist evident, daß das Gehör eben diese Möglichkeit nutzt, indem es die Spektralfrequenzen als Spektraltonhöhen abbildet. Damit wird zum einen die Grundlage dafür geschaffen, das Verhalten einer einzelnen Schallquelle zu analysieren; zum anderen dafür, die Beiträge und Charakteristika mehrerer überlagerter Schallsignale in gewissem Umfange voneinander zu trennen und auf das Verhalten der beteiligten Einzelquellen zurückzuschließen (vgl. Bregman & Campbell, 1971; Hartmann, 1988; McAdams, 1989).

Fig.1 illustriert den ersten Gesichtspunkt an einem konstruierten Beispiel. Dargestellt ist das Spektraltonhöhen-Zeitmuster des Diphtongs /a-i/, und zwar für einen stark eingeschränkten Frequenzbereich (0–2 kHz) und in einer nach Frequenz und Zeit diskretisierten Form. Jedes einzelne Spektraltonhöhenmuster bildet ein binäres Muster und ist vollkommen analog einem binären Rechner-"Wort" (in diesem Fall mit 40 Binärstellen).

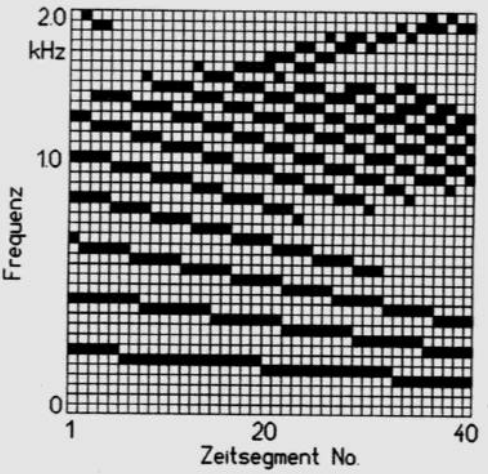

Fig.1. Vergröbertes Spektraltonhöhen-Zeitmuster des Diphtongs /a-i/ (Rechnersimulation von 40 aufeinanderfolgenden Mustern) mit abwärts gleitender Grundfrequenz (200–100 Hz). Das Teiltonspektrum wurde nach der Vokaltheorie von Fant berechnet; anschließend wurden die Ausgeprägtheiten der Spektraltonhöhen nach dem Verfahren von Terhardt (1979) bestimmt. Überschwellig ausgeprägte Spektraltonhöhen sind als schwarze Quadrate dargestellt. Die Frequenz ist nach der Tonheitsfunktion skaliert und zwischen 0 und 2 kHz in 40 Abtastwerte unterteilt, um den binären Charakter der einzelnen Spektraltonhöhenmuster zu verdeutlichen.

Der durch 40 Segmente dargestellte Zeitbereich entspricht einer Dauer von einigen hundert Millisekunden. Die Grundfrequenz ändert sich proportional zur Zeit von 200 auf 100 Hz. Die Formantfrequenzen ändern sich wie folgt: F1 = 750–300 Hz; F2 = 1200–2000 Hz; F3 = 2500–2800 Hz; F4 = 3500 Hz.

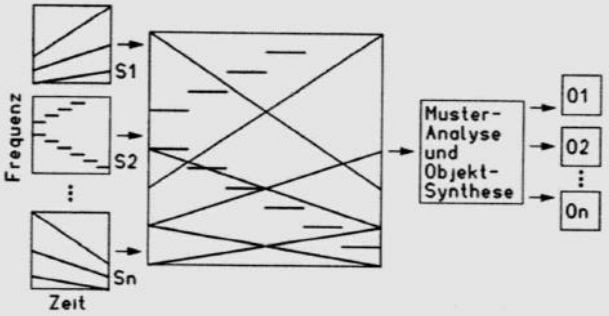

Fig.2. Schematische Darstellung der auditiven Analyse und Resynthese akustischer Objekte S1–Sn mit Hilfe der Spektraltonhöhen-Zeitverläufe. Für drei Signalquellen sind einfache, willkürlich gewählte Muster dargestellt. Dieselben liegen nach der auditiven Fourier-Analyse und Primärkonturisierung als Spektraltonhöhen-Zeitmuster in überlagerter Form vor. Durch aktive, intelligente Verarbeitung gewinnt das Gehör daraus subjektive Repräsentanten O1–On der externen Schallobjekte.

Den Gesichtspunkt der Schallsignaltrennung und Rekonstruktion externer Schallobjekte auf der Grundlage zeitlicher Teiltonfrequenzverläufe illustriert schematisch Fig.2. Die auditive Informationsverarbeitung kann man sich offenbar im Prinzip als eine Kette von Entscheidungsoperationen auf "Objekte" vorstellen, wobei die Teiltöne die Rolle von Elementarobjekten der untersten Hierarchiestufe spielen. Fig.3 illustriert diese Vorstellung durch ein allgemeines Modell. Es enthält insbesondere die folgenden Gesichtspunkte beziehungsweise Annahmen.

- Die Hierarchie von Entscheidungen ist "nach oben offen".
- Die Entscheidungsprozesse (Proz) arbeiten autonom und besitzen das dazu benötigte "Wissen".
- Physikalische Reaktionen auf einen Stimulus können im Prinzip von jeder Stufe aus eingeleitet werden.
- Die subjektive, bewußte Wahrnehmung wird durch einen "Beobachter" repräsentiert, der die Objekte aller Stufen zur Beobachtung selektieren kann.

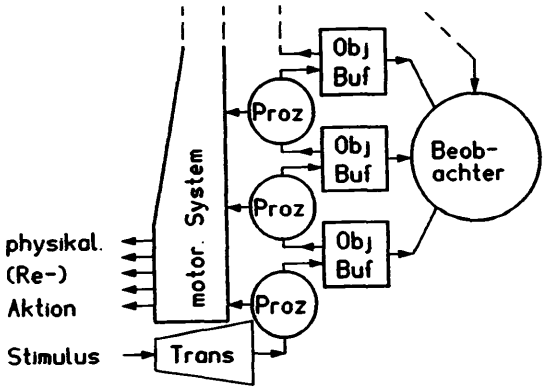

Fig.3. Allgemeines Modell der hierarchischen sensorischen Informationsverarbeitung. Trans: Peripheres Antransportorgan (Außen-Mittel-, Innenohr). Proz: "Intelligente" Prozessoren. Obj Buf: Kurzzeitspeicher für sensorische Objekte.

Literaturangaben
Hartmann, W.M., in W.E.Gall & W.M.Cowan (Hrsg.), *Auditory Function.* Wiley, New York (1988), S. 623–645.
Heinbach, W., Acustica 67, 113–121 (1988).
Bregman, A.S. & Campbell, J., J. Exp. Psychol. 89, 244–249 (1971).
McAdams, S., J. Acoust. Soc. Am. 86, 2148–2159 (1989).
Mummert, M., in *Fortschritte der Akustik (DAGA 90)*, Bad Honnef—Wien, S. 1047–1050.
Terhardt, E., Hearing Research 1, 155–182 (1979).
Terhardt, E., Naturwissenschaften 76, 496–504 (1989).
Terhardt, E., Music Perception 8, 217–239 (1991).