

## EINE MEHRKANALIGE, VIBROTAKTILE HÖRPROTHESE

H. Leysieffer

Institut für Elektroakustik, Technische Universität München

## 1. EINLEITUNG

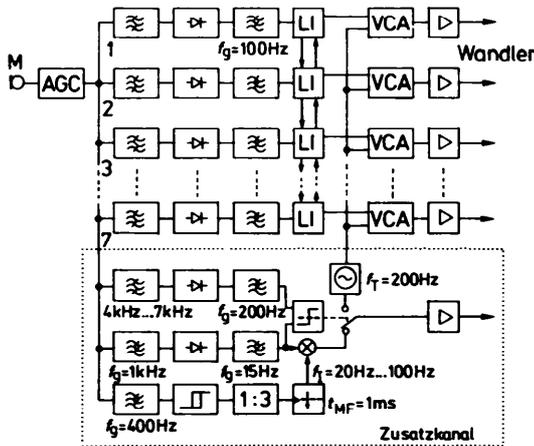
Eine Sinnesprothese, die Sprachinformation durch elektrische oder mechanische Reizung der Hautoberfläche überträgt, sollte klein, leicht und tragbar sein, damit Gehörlose das Gerät im alltäglichen Leben benutzen können. Nur so besteht die Möglichkeit, daß im Laufe der Zeit die Informationsaufnahme im Unterbewußtsein erfolgt und dadurch gegenüber stationären Laborsystemen ein deutlich höherer Grad an Sprachverständlichkeit erreicht werden kann. Dieser Lernprozeß sollte am besten bereits im Vorschulalter taub geborener Kinder beginnen /1/. Obwohl die mechanische Reizung der Hautoberfläche deutliche Vorteile gegenüber der elektrischen Stimulation hat /1/,/2/, scheiterte die sinnvolle, technische Realisierung eines vibrotaktilen Systems zumeist an der hohen Leistungsaufnahme der elektromechanischen Wandler /2/,/3/. Daher wurden neuartige, auf dem piezoelektrischen Material PVDF basierende Wandler entwickelt, die als kleine, ringförmige Elemente an den Fingergliedern einer Hand getragen werden und deren Leistungsaufnahme in der gleichen Größenordnung liegt wie bei optimierten elektrokutanen Reizgebern /2/,/3/,/4/. Im folgenden wird eine tragbare, vibrotaktile Hörprothese beschrieben, die mit diesen Reizgebern arbeitet und mit einer Batterieladung über viele Stunden kontinuierlich betrieben werden kann.

## 2. DARSTELLUNG DES VERFAHRENS

Das gewählte Verfahren entspricht dem des Kanalvocoders. Es extrahiert relevante Information aus dem Sprachsignal und kann mit relativ einfachen schaltungstechnischen Mitteln realisiert werden. Die Hüllkurveninformationen aus mehreren Frequenzbändern des Sprachsignals werden mit Amplitudenmodulation einer in diesem Anwendungsfall für den Tastsinn passend gewählten Trägerfrequenz auf verschiedene Hautorte übertragen. Sprachverständlichkeitsuntersuchungen mit Normalhörenden bei akustisch dargebotener Vocodersprache ergaben, daß bereits mit sechs Kanälen eine Einzelwortverständlichkeit von über 90 % erreicht wird /5/. Aufgrund dieser Ergebnisse wurden im vorliegenden System sieben Kanäle gewählt; in einem achten Zusatzkanal werden Informationen über die Grundfrequenz und über Frikative und Plosive übertragen, die in den Vocoderkanälen nicht enthalten sind.

In Fig.1 ist das Gesamtsystem als Blockschaltbild dargestellt. Ein kleines Lavaliermikrofon wandelt den Sprachschall in ein elektrisches Signal, das in einem Regelverstärker (AGC) vorverstärkt wird. Der Regeleinsatz des AGC liegt bei ca. 55 dB SPL, der Regelumfang beträgt 35 dB. Das Zurückregeln der Verstärkung erfolgt mit einer Zeitkonstante  $t_{ab} = 0.5$  ms, um auch kurze Verschlusslaute zu erfassen. Das Hochregeln erfolgt langsamer ( $t_{auf} = 200$  ms), um die Phonedynamik nicht zu stark zu verändern. Der AGC beinhaltet zusätzlich eine Höhenanhebung 2. Ordnung ( $f_g = 1$  kHz), um den Höhenabfall im Langzeitspektrum der Sprache zu kompensieren /6/.

Das geregelte Signal wird einer Filterbank mit sieben Bandpässen vierter Ordnung zugeführt. Die Filtergrenzfrequenzen sind in der Tabelle 1 zusammengestellt. Die Filterbank umfaßt einen Spektralbereich, innerhalb dessen die wesentlichen ersten beiden Formanten des Sprachsignals liegen. Die Bandfilter-



Filter	$f_u$ (Hz)	$f_o$ (Hz)
1	160	350
2	350	550
3	550	750
4	750	1050
5	1050	1450
6	1450	2100
7	2100	2900

Fig. 1 : Blockschaltbild des Gesamtsystems

Tabelle 1 : Grenzfrequenzen der Filterbank

ausgangssignale werden über Einweggleichrichter und nachfolgende Tiefpässe erster Ordnung hüllkurvendemoduliert. Die Einschwingzeit der Tiefpässe (ca.10ms) ist dem Zeitaufklärungsvermögen des Tastsinns bei mechanischer Stimulation angepaßt. Voruntersuchungen ergaben, daß kurze, aufeinanderfolgende Impulse dann als Einzelereignisse wahrgenommen werden, wenn sie mit ca. 15 ms Verzögerungszeit dargeboten werden, und zwar unabhängig davon, ob die Reizung mit einem Impulspaar an einem Hautort (Fingerglied) oder an zwei verschiedenen Orten erfolgt.

Die demodulierten Filtersignale werden in ein spezielles Netzwerk (LI) eingespeist, das die einzelnen Kanäle mit ihren in der Frequenzlage direkt benachbarten Kanälen vermascht und dessen Funktionsweise eng an das aus der Psychophysik bekannte Prinzip der "lateralen Inhibition" angelehnt ist /7/. Dieses Netzwerk verbessert durch Subtraktion der mit einem wählbaren Gewicht bewerteten Hüllkurvensignale der benachbarten Kanäle die spektrale Selektion der Filterbank. Die Gewichte der Inhibition sind so gewählt, daß bei einer Pegeldifferenz zwischen zwei direkt benachbarten Kanälen von ca. 5 dB der schwächere Kanal vollständig unterdrückt wird. Versuche mit rechner-synthetisierten Vokalspektren zeigten, daß am Ausgang dieses Netzwerks im wesentlichen nur noch zwei Maxima auftreten, die den ersten beiden Formanten des jeweiligen Vokals entsprechen.

Die Ausgangsspannungen der sieben Inhibitionsstufen gelangen als Steuereingänge auf spannungsgesteuerte Verstärker (VCA). Das Trägersignal dieser Amplitudenmodulatoren wird von einem Sinusoszillator ( $f_p = 200$  Hz) geliefert; diese Trägerfrequenz entspricht dem Fühlschwellenminimum bei sinusförmigen Hautoberflächen-deformationen /3/. Die Ausgangsspannungen der Modulatoren werden in Endstufen verstärkt und speisen über Resonanzübertrager die PVDF-Wandler.

Mit diesem Vocodersystem wird eine Transformation der spektralen Energieverteilung des Sprachsignals in ein örtliches Erregungsmuster erreicht, die in grober Analogie der Frequenz-Orts-Transformation des menschlichen Gehörs ähnelt. Die Optimierung der Frequenz-Orts-Zuordnung folgte aus Mustererkennungs- und Maskierungsexperimenten :

- Örtlich direkt benachbarte Reizorte werden am häufigsten verwechselt
- Örtlich direkt benachbarte Reizorte maskieren sich gegenseitig am stärksten.

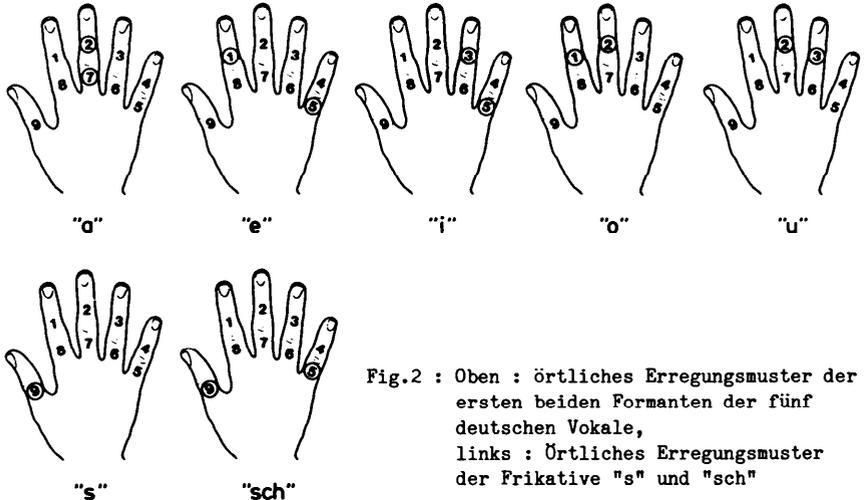


Fig.2 : Oben : örtliches Erregungsmuster der ersten beiden Formanten der fünf deutschen Vokale, links : Örtliches Erregungsmuster der Frikative "s" und "sch"

Daraus ergab sich folgende Vocoderkanal-Reizort-Zuordnung :

Kanal :	1	2	3	4	5	6	7
Reizort :	3	1	8	2	7	6	5

Die aus dieser Zuordnung resultierenden örtlichen Erregungsmuster der ersten beiden Formanten der fünf deutschen Vokale "a", "e", "i", "o" und "u" sind im oberen Teil von Fig.2 dargestellt. Der Hautort "4" wird nicht verwendet.

Informationen über Grundfrequenz und stimmlose Laute werden mit dem in Fig.1 punktiert umrahmten Zusatzkanal übertragen. Das geregelte Signal wird im unteren Schaltungszweig zur Extraktion der Grundfrequenz tiefpassgefiltert (Grad 3,  $f_g=400$  Hz), mit einem Schmitt-Trigger in eine Rechteckschwingung umgeformt und mit einem Frequenzteiler durch drei geteilt. Dieses geteilte Rechtecksignal wird über ein Monoflop (Laufzeit 1 ms) in eine Impulsfolgefrequenz  $f_I$  umgesetzt und einem Multiplizierer zugeführt. Diese der Grundfrequenz proportionale Impulsfolgefrequenz liegt in einem Frequenzbereich ( $f_I=20\text{Hz} \dots 100\text{Hz}$ ), innerhalb dessen für den Tastsinn etwa sechs bis sieben unterscheidbare Frequenzstufen liegen. Die Amplitude dieser Impulsfolge wird über das Multiplizierglied mit der Hüllkurve des Sprachsignals moduliert, die über einen Tiefpaß (Grad 3,  $f_g=1$  kHz) mit Hüllkurvendemodulation ( $f_g=15$  Hz) gewonnen wird. Im oberen Schaltungszweig des Zusatzkanals wird über einen Bandpaß ( $f_u=4$  kHz,  $f_o=7$  kHz) mit nachfolgender Demodulation ( $f_g=200$  Hz) der Energieinhalt der höherfrequenten Anteile des Sprachsignals (Frikative und Plosive) ermittelt. In einer Komparatorstufe wird durch Vergleich der Energieinhalte der höherfrequenten und der tieffrequenten Sprachsignalanteile entschieden, ob ein stimmhafter oder stimmloser Laut vorliegt. Dieser Vergleichler steuert einen Umschalter, der bei stimmhaften Lauten die Impulsfolge, bei stimmlosen Lauten das sinusförmige Trägersignal des Vokoderteils ( $f_T=200$  Hz) über eine Endstufe auf den Reizgeber des Zusatzkanals führt. Als Reizort wurde der Daumen (Ort "9", siehe Fig.2) gewählt. Die Arbeitsweise dieses Zusatzkanals

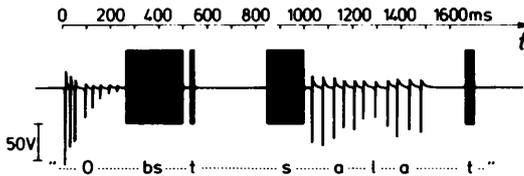


Fig.3 : Beispiel für die Zeitfunktion der Wandlerspannung des Zusatzkanals

ist mit einem Wortbeispiel ("Obstsalat") in Fig.3 dargestellt.

Aufgetragen ist die Zeitfunktion der Wandlerspannung. Bei stimmhaften Wortanteilen wird die amplitudenmodulierte Impulsfolge zugeschaltet, bei frikativem oder plosivem Anteil das sinusförmige 200 Hz-Ersatzsignal konstanter Amplitude. Zur Veranschaulichung der Impulsfolgefrequenzmodulation wurde bei Wortbeginn mit erhöhter Stimmlage gesprochen.

In Fig.2 sind im unteren Teil die örtlichen Erregungsmuster der Frikative "s" und "sch" dargestellt; beim stimmlosen "s" wird nur das Ersatzsignal des Zusatzkanals dargeboten (Reizort "9"), beim breitbandigeren Frikativ "sch" fallen Energieanteile zusätzlich in den obersten Kanal des Vokoderteils (Reizort "5"). Durch diese deutlich verschiedenen Reizmuster wird mit Hilfe des Zusatzkanals eine grobe Frikativklassifizierung erreicht.

### 3. SCHLUSSBEMERKUNGEN

Das gesamte System ist in aktiver RC-Technik aufgebaut, hat etwa die Größe eines Taschenbuchs und wiegt ca. 700 g. Mit einer Batterieladung kann das Gerät 15-20 Stunden kontinuierlich betrieben werden. Erste Vorversuche mit Gehörlosen zeigen, daß das System nicht als übermäßig hinderlich oder lästig empfunden wird und daß die mechanischen Hautstimulationen in einem angenehmen Empfindungsstärkebereich liegen.

### LITERATUR :

- /1/ Sherrick, C.E.(1984). Basic and applied research on tactile aids for deaf people: Progress and prospects. J.Acoust.Soc.Am. 75 (5).
- /2/ Hoffmann, C.(1981). Eine tragbare Sinnesprothese für Gehörlose zur Übertragung von Sprachinformation durch elektrische Anregung der Haut. Dissertation TU München.
- /3/ Leysieffer, H.(1985). Polyvinylidenfluorid als elektromechanischer Wandler für taktile Reizgeber. Acustica 58 (3).
- /4/ Leysieffer, H.(1985). Vibrotaktile Reizgeber mit PVDF (Polyvinylidenfluorid) als elektromechanischem Wandler. Fortschritte der Akustik, DAGA 85, VDE Verlag Berlin.
- /5/ Zollner, M.(1979). Verständlichkeit der Sprache eines einfachen Vocoders. Acustica 43.
- /6/ Tarnoczy, T.(1971). Das durchschnittliche Energie-Spektrum der Sprache (für 6 Sprachen). Acustica 24.
- /7/ Zwicker, E. (1963). Möglichkeiten zur Spracherkennung über den Tastsinn mit Hilfe eines Funktionsmodells des Gehörs. Elektron. Rechenanl. 7.