

Sprachverständlichkeitsmessungen mit einer vibrotaktilen Hörprothese

H. Leysieffer

Institut für Elektroakustik, Technische Universität München

1. Einleitung

Bei der DAGA' 86 wurde ein tragbares, mehrkanaliges Sprachübertragungssystem für Gehörlose vorgestellt, das mit mechanischer Hautreizung an acht Fingergliedern einer Hand arbeitet /1/. Das auf dem Vocoderprinzip basierende Verfahren überträgt spektrale Information des Sprachschalles in sieben Kanälen auf elektromechanische PVDF-Wandler, die wie Fingerringe getragen werden /2, 3/. Der Energieinhalt der Bandfilter des Analyseteils wird durch Amplitudenmodulation einer sinusförmigen Trägerschwingung mit der Bestfrequenz der Vibrationsstärkeempfindung (200 Hz) übermittelt. Die Spektralanalyse wurde so konzipiert, daß im wesentlichen die ersten zwei Formanten der Vokale auf zwei verschiedene Reizorte abgebildet werden. In einem Zusatzkanal werden Informationen über die Sprachgrundfrequenz und über stimmlose Laute, die in dem Vocoderteil nicht verarbeitet werden, auf einen weiteren Wandler am Daumen übertragen; in diesem ergänzenden Kanal wird bei stimmhaften Lauten die Grundfrequenz mit Hilfe einer frequenzmodulierten Impulsfolge übermittelt, bei stimmlosen Plosiven und Frikativen wird ein sinusförmiges Ersatzsignal ($f = 200$ Hz) dargeboten. Das Gerät hat die Größe einer Zigarrenschachtel, wiegt ca. 700 g und kann mit einer Batterieladung etwa 15 Stunden kontinuierlich betrieben werden. Mit diesem System wurden Sprachverständlichkeitsuntersuchungen mit normalhörenden und tauben oder schwerst hörgeschädigten Personen durchgeführt, die im folgenden beschrieben werden.

2. Versuche mit normalhörenden Personen

Das verwendete Sprachmaterial wurde auf einem Rechner synthetisiert oder von einem oder mehreren Sprechern in einem echoarmen Raum auf Band aufgezeichnet. Die Darbietungen wurden den Versuchspersonen ohne das Mikrofon direkt in das Gerät eingespielt; bei allen Versuchen wurde ausschließlich die rechte Hand als Reizfläche gewählt. Bei den Sitzungen wurde in einer Lernphase das Sprachmaterial neunmal in bekannter Reihenfolge dargeboten; in der anschließenden Testphase erfolgten pro Sitzung 100 statistisch gleichverteilte Abfragen des Sprachmaterials. Die Ergebnisse sind als Häufigkeit h_r richtiger Antworten über

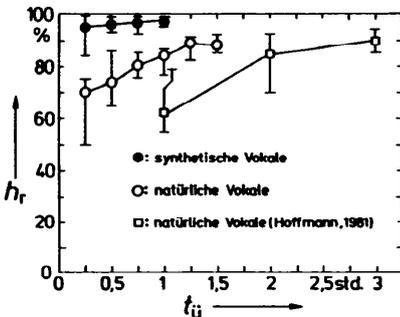


Abb. 1: Häufigkeit h_r der richtigen Erkennung von fünf langen Vokalen bei synthetischer Erzeugung (gefüllte Kreise) und bei Darbietung von einem männlichen Sprecher (offene Kreise, sechs Versuchspersonen) über der Übungszeit $t_{\ddot{u}}$. Offene Quadrate: Daten von Hoffmann (1981) bei der Darbietung des gleichen Sprachmaterials mit einem 12-kanaligen, elektroakustischen Verfahren (acht Versuchspersonen)

der akkumulierten Übungs- und Testzeit t_{ij} als Zentralwerte und Wahrscheinliche Schwankungen dargestellt. Für den ersten, einfachen Test wurden die fünf langen Vokale "a:", "e:", "i:", "o:" und "u:" rechner-synthetisiert und daher identisch reproduzierbar sechs Versuchspersonen nur über den Vocoderteil des Systems (ohne Zusatzkanal) dargeboten. Bereits nach einer Sitzung ($t_{ij} = 0,25$ std., vgl. Abb. 1: gefüllte Kreise) liegen die Erkennungsquoten um 95 % und nehmen im Laufe weiterer Versuche noch zu. Werden die Vokale von einem männlichen Sprecher dargeboten, so daß sie der natürlichen Aussprachevariation unterliegen, ergaben sich Erkennungsdaten zu Beginn der Versuchsreihe von 70 % (offene Kreise), die im Laufe einer Stunde auf ca. 90 % ansteigen. Zum Vergleich sind in Abb. 1 Versuchsergebnisse von Hoffmann /4/ bei gleichem Testmaterial mit einer 12-kanaligen, elektrokutanen Übertragung in der Unterarmregion eingetragen (offene Quadrate). Seine Versuchspersonen erreichten nach etwas längerer Übungszeit (ca. 3 std.) den gleichen Endwert der Erkennungsquoten um 90 %. In einer weiteren Versuchsreihe wurden mit dem vollständigen System (Vocoderteil + Zusatzkanal) acht Versuchspersonen zehn einsilbige Zahlworte (0...9; 7 als "sie:b" gesprochen) von einem männlichen Sprecher mit gleicher Statistik in der Lern- und Testphase wie bei den Vokalversuchen dargeboten. Die Ergebnisse sind als Zentralwerte (gefüllte Kreise) und Mittelwerte (punktierte Kurve) in Abb. 2 dargestellt:

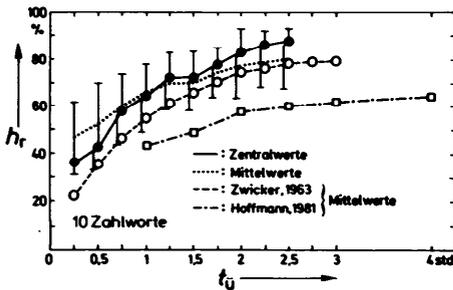


Abb. 2: Häufigkeit h_r der richtigen Erkennung zehn einsilbiger Zahlworte über der Übungszeit t_{ij} (gefüllte Kreise: Zentralwerte, punktierte Kurve: Mittelwerte von acht Versuchspersonen); offene Kreise: Daten nach Zwicker (1963) bei einem 24-kanaligen vibrotaktilen Verfahren; offene Quadrate: Daten nach Hoffmann (1981) bei einer 12-kanaligen, elektrokutanen Übertragung

Nach etwas längerer Übungszeit ($t_{ij} = 2,5$ std.) als bei den Vokalversuchen werden auch bei diesem Versuch Erkennungsdaten um 88 % (Zentralwert) erreicht. Wird der Versuch mit einem anderen Sprecher ohne Lernphase wiederholt, sinken die Ergebnisse um 10...20 % bei der ersten Sitzung, steigen aber rasch wieder auf die gleichen Werte an wie bei dem ersten Sprecher. Zum Vergleich sind in Abb. 2 Mittelwerte der Versuchsergebnisse nach Zwicker /5/ (offene Kreise, vier Versuchspersonen) bei gleichem Sprachmaterial dargestellt, die sich bei einer 24-kanaligen vibrotaktilen Übertragung mit einem Funktionsmodell des Gehörs in der Unterarmregion ergaben. Die große Ähnlichkeit der Lernkurven und der erreichten Endwerte der Erkennungsquoten deuten darauf hin, daß zumindest bei einfachen Worten acht Kanäle ausreichend sind, um wesentliche Informationen zu übertragen. Die Erkennungsquoten bei der 12-kanaligen elektrokutanen Übertragung nach Hoffmann /4/ (offene Quadrate, Mittelwerte von acht Versuchspersonen) bei dem gleichen Sprachtest bleiben auch nach langer Übungszeit ($t_{ij} = 4$ std.) signifikant unter den Ergebnissen, die mit den vibrotaktilen Verfahren erreicht werden.

In weiterführenden Tests mit zwei geübten Versuchspersonen mit je zehn zwei- und dreisilbigen Worten zeigte sich, daß die Erkennungsdaten mit der Silbenzahl der Testworte abnehmen ($h_r = 80$ % bei zweisilbigen und $h_r = 65$ % bei dreisilbi-

gen Worten nach $t_{ij} = 2$ std.). Dies liegt einerseits daran, daß die durch den Zusatzkanal übermittelte prosodische Information bei mehrsilbigen Worten überbewertet wird; andererseits konnte Bäuerte /6/ zeigen, daß die übertragbare Information (Synentropie) elementarer taktiller Zeichen mit steigender Zahl dieser Zeichen in einem Wort signifikant abnimmt. Weitere Kontrollmessungen mit den zwei trainierten Versuchspersonen mit einem Reintestverfahren nach Sotschek /7/ ergaben, daß die Vokalverständlichkeit bei Einbettung der Vokale in eine konsonantische Umgebung gegenüber isolierter Darbietung deutlich verschlechtert wird; dies ist ein weiterer Grund für die niedrigeren Erkennungsraten mehrsilbiger Worte gegenüber den einsilbigen Zahlworten, die etwa zur Hälfte mit einem langen Vokal beginnen oder enden. In einem abschließenden Test mit den zwei trainierten, normalhörenden Versuchspersonen wurden die zehn einsilbigen Zahlworte mit den Operanden "plus", "mal", "minus", "durch" und "ist" zu einfachen algebraischen Aufgaben oder Aussagen verknüpft und so 96 verschiedene kurze Sätze (je vier oder fünf Worte) zusammengestellt; die Lernphase dieser Versuchsreihe bestand nur in der neunmaligen Darbietung der Operanden. In Abb. 3 sind die Ergebnisse der zwei Versuchspersonen als Häufigkeit h_r der vollständig richtigen Wiedergabe jedes Testsatzes über der Übungszeit t_{ij} dargestellt:

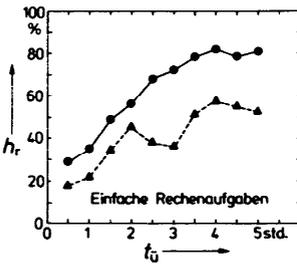


Abb. 3: Häufigkeit h_r der richtigen, vollständigen Erkennung einfacher algebraischer Aufgaben und Aussagen über der Übungszeit t_{ij} dargestellt sind die Ergebnisse von zwei trainierten, normalhörenden Versuchspersonen. Drei Beispiele der Testsätze aus insgesamt 96 verschiedenen Sätzen: "2 · 4 = 8", "6 = 8 + 3", "8 - 9 = 1"

Nach $t_{ij} = 2$ std. erreichen beide Versuchspersonen Erkennungsquoten um 50 %; eine Versuchsperson (gefüllter Kreis) konnte ihre Ergebnisse im Laufe weiterer Sitzungen auf ca. 80 % steigern, bei der anderen blieben die Erkennungsraten auch nach längerer Übung bei etwa 55 %. Diese Ergebnisse deuten darauf hin, daß die Informationsaufnahme zu einem großen Teil bereits nach relativ kurzer Übungszeit im Unterbewußtsein erfolgt, da ein "Nachdenken" über einzelne Informationseinheiten (Phoneme, Silben oder Worte) bei der annähernd normalen Sprechgeschwindigkeit bei der Darbietung nicht möglich ist.

3. Versuche mit tauben und schwerst hörgeschädigten Personen

Folgende vier Personen stellten sich für einige einfache Tests mit natürlichen Vokalen und den zehn einsilbigen Zahlworten zur Verfügung:

- VP1: männlich, 58 Jahre, Ingenieur, postlingual ertaubt,
- VP2: männlich, 33 Jahre, Feinmechaniker, prälingual ertaubt,
- VP3: weiblich, 18 Jahre, Schülerin, prälingual ertaubt,
- VP4: weiblich, 61 Jahre, Hausfrau, postlingual ertaubt.

Alle Versuche wurden in normalen Wohnräumen durchgeführt; das Sprachmaterial wurde von zwei männlichen Sprechern in einem Abstand von ca. einem Meter zu dem Mikrophon des Gerätes dargeboten. Nach einer Lernphase mit Lippenablesen erfolgten je Sitzung zwischen 50 und 140 statistisch verteilte Einzelabfragen ohne Lippenablesen. Die Versuchspersonen konnten beliebige Wiederholungen der

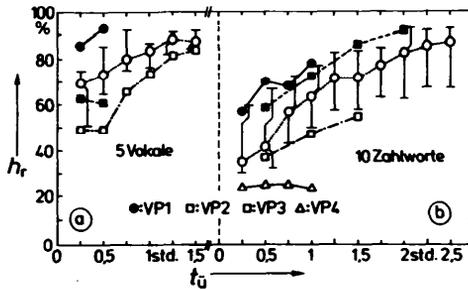


Abb. 4: Häufigkeit h_r der richtigen Erkennung von fünf langen Vokalen (a) und zehn einsilbigen Zahlworten (b) über der Übungszeit t_{ij} bei vier gehörlosen oder schwerst hörgeschädigten Personen (VP1...VP4); zum Vergleich sind die Ergebnisse normalhörender Personen als Zentralwerte (offene Kreise) und Wahrscheinliche Schwankungen eingetragen (vgl. Abb. 1 und Abb. 2)

Darbietungen anfordern. Die in Abb. 4 dargestellten Ergebnisse zeigen, daß die erzielten Erkennungsquoten sehr unterschiedlich sind. Zum Vergleich sind als Zentralwerte (offene Kreise) und Wahrscheinliche Schwankungen die Ergebnisse normalhörender Personen eingetragen. Bei dem Vokaltest (Abb. 4a) erreicht VP1 bereits in den ersten zwei Sitzungen Erkennungsquoten, die deutlich über denen der Normalhörenden liegen. Diese guten Ergebnisse werden auch in dem Zahlworttest erzielt (Abb. 4b). VP2 erreicht bei dem Vokaltest nur Raten um 60 %, dagegen liegen die Ergebnisse bei den Zahlworten in jeder Sitzung wie bei VP1 über den Zentralwerten der Ergebnisse Normalhörender. VP3 zeigt in beiden Tests "normale" Lernkurven. VP4, die nur an dem Zahlwortversuch teilnahm, erzielt auch nach längerer Übung Erkennungsquoten, die nur unwesentlich über der Ratewahrscheinlichkeit (10 %) liegen. Diese Versuche zeigen einerseits, daß die bei tauben oder schwerst hörgeschädigten Personen erzielten Ergebnisse mit denen Normalhörender vergleichbar sind und daß andererseits eine signifikante Abhängigkeit der Erkennungsleistungen von der Art der Ertaubung (prä- oder postlingual) nicht zu erkennen ist. Vielmehr können diese ersten Resultate dahingehend gedeutet werden, daß die Konzentrationsfähigkeit, Motivation und eine positive Einstellung zum Erlernen dieser "neuen Sprache" wesentliche Rollen bei dem erfolgreichen Einsatz des beschriebenen Verfahrens spielen.

Literatur

- /1/ Leysieffer, H. (1986): Eine mehrkanalige, vibrotaktile Hörprothese. In: Fortschritte der Akustik - DAGA'86 Oldenburg; DPG-Kongreß GmbH, Bad Honnef
- /2/ Leysieffer, H. (1985): Vibrotaktile Reizgeber mit PVDF (Polyvinylidenfluorid) als elektromechanischem Wandler. Fortschritte der Akustik - DAGA'85 Stuttgart; DPG-Kongreß GmbH, Bad Honnef
- /3/ Leysieffer, H. (1985): Polyvinylidenfluorid als elektromechanischer Wandler für taktile Hörprothesen. *Acustica*, Vol. 58(3), 196
- /4/ Hoffmann, C. (1981): Eine tragbare Sinnesprothese für Gehörlose zur Übertragung von Sprachinformation durch elektrische Anregung der Haut. Dissertation TU München
- /5/ Zwicker, E. (1963): Möglichkeiten zur Spracherkennung über den Tastsinn mit Hilfe eines Funktionsmodells des Gehörs. *Elektron. Rechenanl.* 7, 239
- /6/ Bäuerle, R. (1974): Vibrotaktile Informationsübertragung mit Folgen binärer Zeichen. *Kybernetik* 15, 85
- /7/ Sotschek, J. (1982): Ein Reimtest für Verständlichkeitsmessungen mit deutscher Sprache als ein verbessertes Verfahren zur Bestimmung der Sprachübertragungsgüte. *Der Fernmelde-Ing.* 36, Vol. 4(5)