

## EIN VIELKANALIGES INHIBITIONSNETZWERK FÜR SPRACHÜBERTRAGUNG

Thomas Beckenbauer  
Institut für Elektroakustik, Technische Universität München

### 1. Einleitung

Auf der DAGA des vergangenen Jahres wurde ein analoges Netzwerk vorgestellt, das auf stark vereinfachte Weise inhibitorische Vorgänge in Sinnesorganen nachbildet [1]. Es wird eingesetzt, um den spektralen Kontrast verrauschter Sprachsignale zu erhöhen und damit eine Störverminderung herbeizuführen.

Sprache weist kontrastreiche Spektren auf, die zeitlich stark variieren. Es ist deshalb notwendig, mit Hilfe eines zeitlich dynamischen Verfahrens das Eingangssignalspektrum zunächst in schmale Teilfrequenzbereiche zu zerlegen, um dann lokale Maxima im Spektrum, die zeitabhängig immer wieder in anderen Teilfrequenzbereichen liegen, zu detektieren und zur Steuerung der inhibitorischen Vorgänge heranzuziehen. Die Wirkung beruht in der Reduzierung intensitätsärmerer Spektralanteile.

Hinsichtlich der spektralen Struktur des Eingangssignals erzielt das Netzwerk die beabsichtigte Wirkung. Hinsichtlich der zeitlichen hat sich gezeigt, daß rauschhafte Störsignale, die ebenso wie das Sprachsignal starke, kurzzeitige Schwankungen aufweisen, zu dauernden, schnellen wechselseitigen Hemmungen der Kanalsignale führen und sich akustisch störend bemerkbar machen. Das zeitliche Verhalten der Steuersignale wurde deshalb modifiziert.

### 2. Beschreibung des modifizierten Inhibitionsnetzwerks

Wie das Blockschaltbild in Fig.1 zeigt, wird zur Steuerung der einzelnen Kanäle ein Signal aus der Differenz der Verhältnisse der Hüllkurvensignale in den benachbarten Kanälen zum Hüllkurvensignal im betrachteten Kanal generiert (vgl. [1]). Jeder Kanal ist mit maximal vier Nachbarkanälen nach höheren und niedrigeren Frequenzen vermascht. Jedes ankommende Hüllkurvensignal wird mit einem Gewicht  $g_{\pm j}$  bewertet, das zwischen 0 und 1 eingestellt werden kann. Insgesamt stehen 20 Kanäle in einem Frequenzbereich von 50 Hz bis 7,5 kHz zur Verfügung.

Betrachtet man die zeitliche Dynamik des Steuersignals  $U_{St}$  stellt man fest, daß sie zum einen von der Einschwingzeit des Bandpasses, zum anderen von der Zeitkonstante des Tiefpasses bei der Hüllkurvenbildung bestimmt ist. Wird das Signal in einem betrachteten Kanal durch den sprunghaften Anstieg des Hüllkurvensignals in einem Nachbarkanal vollständig unterdrückt, geht das Steuersignal mit einer Zeitkonstante von 30 bis 12 ms bei tiefen, 10 bis 5 ms bei mittleren und 4 bis 1 ms bei hohen Frequenzen auf seinen kleinsten Wert zurück. Das bedeutet, daß auch kurzzeitige Änderungen des Eingangssignalspektrums, wie sie nicht nur von Sprache, sondern auch von rauschhaften Störsignalen hervorgerufen werden, zu inhibitorischen Wirkungen im Netzwerk führen. Diese Eigenschaft bewirkt eine starke zeitliche Strukturierung von sonst gleichmäßig wahrzunehmenden, breitbandigen Rauschen. Der positive Effekt der Unterdrückung von Störsignalanteilen während gleichzeitiger Darbietung von Sprache wird durch den negativen Effekt der starken zeitlichen Zergliederung des Störsignals wieder aufgehoben, was in den Ergebnissen der Sprachverständlichkeitstests, die später beschrieben werden, zum Ausdruck kommt.

Die einzelnen Kanäle wurden daraufhin so erweitert, daß das Steuersignal  $U_{St}$  unterschiedliche Ein- und Ausschwingzeiten aufweist. Mit Hilfe eines nichtlinearen

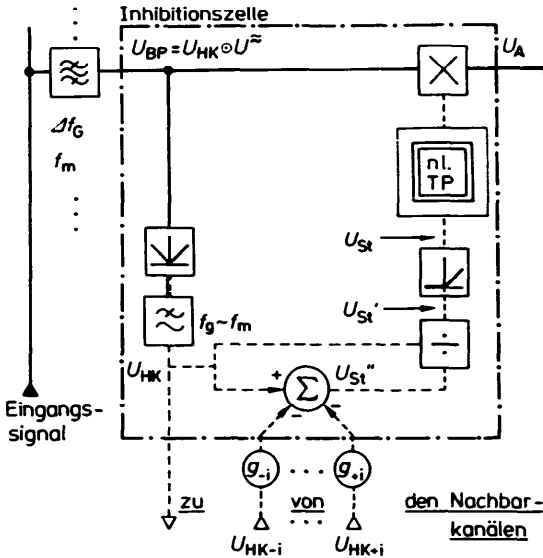


Fig.1: Blockschaltbild einer um den nichtlinearen Tiefpaß erweiterten Inhibitionszelle. Es gilt:

$$\begin{aligned}
 &U_{HK} \geq 0; \\
 &U_{\Sigma} = \sum g_{\pm i} U_{HK \pm i} \geq 0 \\
 &\text{mit } i=1.2.3.4; \\
 &U_{St}'' = U_{HK} - U_{\Sigma}; \\
 &U_{St}' = U_{St}'' U_{HK} \leq 1; \\
 &0 \leq U_{St} = \max(0; 1 - U_{\Sigma} U_{HK}) \leq 1; \\
 &U_A = U_{BP}'' \max(0; 1 - U_{\Sigma}'' U_{HK});
 \end{aligned}$$

— Wechsel-  
 - - - Gleich-  
 signalpfad.

Tiefpasses, dessen Eingangssignal das Steuersignal  $U_{St}$  darstellt (Fig.1), konnte erreicht werden, daß die Einschwingzeitkonstante, also diejenige Zeitkonstante, die für den Anstieg des Steuersignals und damit für die Aufhebung der Hemmung maßgeblich ist, unverändert bleibt, während die Ausschwingzeitkonstante, also diejenige Zeitkonstante, die für den Abfall des Steuersignals  $U_{St}$  und damit für die Zunahme der Hemmung maßgeblich ist, einstellbar und beliebig groß ist. Der Steuerungsmechanismus verhält sich zeitlich adaptiv.  $\tau_{EnlT}$  sei die Einschwingzeitkonstante und  $\tau_{AnlT}$  die Ausschwingzeitkonstante des nichtlinearen Tiefpasses. Fig.2 zeigt das durch den Einbau des nichtlinearen Tiefpasses veränderte Verhalten des Ausgangssignals eines Kanals.

**3. Sprachverständlichkeitsmessungen**

Ein Anwendungsgebiet für das vorliegende Netzwerk könnte die Versorgung von Hörbehinderten sein. Speziell die Gruppe der Innenohrgeschädigten weist häufig den Effekt des Recruitments auf, einen versteilten Anstieg der Lautheitsempfindung mit dem Schallpegel, der schnell zu einer Überlastung des geschädigten Gehörs führt. Schon aus diesem Grund wäre eine Verminderung der Lautheit von Umgebungsgereuschen alleine wünschenswert.

In einem ersten Experiment wurde Sprache gleichzeitig mit einem 140 Hz Rechteckpuls mit 50  $\mu$ s Impulsdauer als Störgeräusch dargeboten. Die gefüllten Kreise in Fig.3, Teilbild (a) kennzeichnen die Lautheit des Störgeräusches  $N_{ml}$  mit Inhibition in Abhängigkeit von der Lautheit  $N_{ol}$  dieses Störgeräusches ohne Inhibition. Die Ausschwingzeitkonstante  $\tau_{AnlT}$  der nichtlinearen Tiefpässe wurde 5 mal größer gewählt als die Einschwingzeitkonstante  $\tau_{EnlT}$ . Wie aus dem Bild hervorgeht, nimmt die Lautheit des Störgeräusches am Ausgang des Netzwerks nicht in dem Maße zu wie am Eingang, obgleich die Gesamtlautheit des Schalles, der sich aus Sprache und Störgeräusch zusammensetzt, gleich bleibt. Zur Orientierung sind die Gesamtlautheiten beider Schalle für den Fall ohne Inhibition nach rechts und für den Fall mit Inhibition nach oben mit offenen Quadraten eingezeichnet. Die dazu-

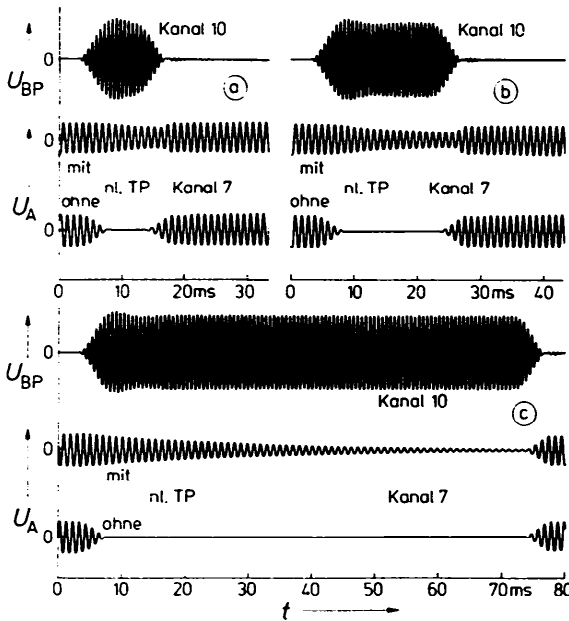
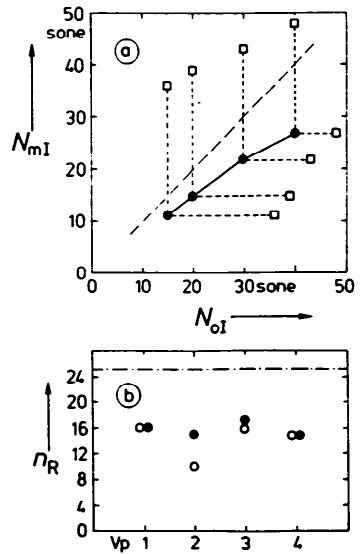


Fig.2: Zeitverläufe des Bandpaßausgangssignals  $U_{BP}$  in Kanal 10 und des Kanalausgangssignals  $U_A$  in Kanal 7 mit und ohne nichtlinearem Tiefpaß. Das Eingangssignal setzt sich aus einem Tonimpuls (970 Hz) und einem Dauerton (1570 Hz) halber Amplitude zusammen. Die Frequenzen entsprechen den Mittenfrequenzen der Kanäle 7 und 10. Vermaschung der beiden Kanäle gegenseitig mit Gewicht  $g_{\pm 3} = 0.5$ . Bezeichnungen der Signale aus Fig.1. (a) Impulsdauer  $t_i = 10\text{ms}$ . (b)  $t_i = 20\text{ms}$ . (c)  $t_i = 70\text{ms}$ .

Fig.3: (a) Lautheit eines Rechteckpulses (Folgefrequenz 140 Hz, Impulsdauer 50  $\mu\text{s}$ ) (geschlossene Kreise) und Gesamtlautheit des Rechteckpulses bei gleichzeitig dargebotener Sprache (offene Rechtecke)  $N_{mI}$  mit Inhibition in Abhängigkeit von der Lautheit der Schalle  $N_{oI}$  ohne Inhibition. Gewichtung der Nachbarkanalhüllkurven mit 0,0135 (Kanäle -4 und +4), 0,045 (Kanäle -3 und +3), 0,15 (Kanäle -2 und +2) und 0,5 (Kanäle -1 und +1).  $\tau_{AnIT} = 5 \cdot \tau_{EnIT}$ . (b) Anzahl der richtig erkannten Substantive  $n_R$  in 25 vollständigen Sätzen für 4 verschiedene Versuchspersonen (offene Antwortmöglichkeiten). Offene Kreise: ohne, geschl. Kreise: mit Inhibition (vgl. (a)).

gehörigen Zahlenwerte können auf der Abszisse bzw. auf der Ordinate abgelesen werden. Die Länge der Strecken zwischen den geschlossenen Kreisen und den offenen Quadraten sind ein Maß für den Anteil des Störgeräusches an der Gesamtlautheit. Je kürzer die Strecke ist, desto mehr wird die Gesamtlautheit des Schalles vom Störgeräusch bestimmt.

Um sicherzugehen, daß die auf diese Weise bearbeitete Sprache nicht in der Verständlichkeit gelitten hat, wurden vier normalhörenden Versuchspersonen 25 verschiedene, kurze Sätze dargeboten, in denen das Hauptwort erkannt werden sollte. Die Ergebnisse sind in Fig.3, Teilbild (b) ein-

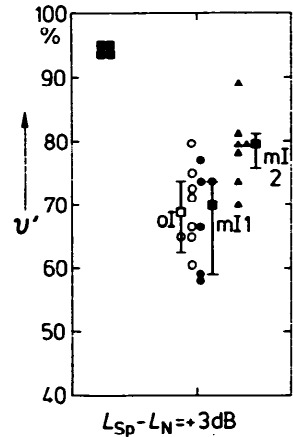


getragen. Außer bei Versuchsperson 2. die bei dem Versuch mit Inhibition deutlich besser abschnitt, verändert sich die Erkennungsrate mit und ohne Inhibition nicht.

Verwendet man anstelle des eben angegebenen Störgeräusches mit diskrettem Spektrum konstanter Amplitude ein stochastisches Störgeräusch, läßt sich der Lautheitsgewinn ohne zusätzlich dargebotene Sprache nicht mehr angeben. Das Störgeräusch wird nur dann stark unterdrückt, wenn auch ein Sprachsignal vorhanden ist. Die Wahl der Ausschwingzeitkonstante des nichtlinearen Tiefpasses spielt ebenfalls eine wichtige Rolle.

In Fig.4 sind Meßergebnisse dargestellt, die mit dem Sotscheck Reimtest /2/ und normalhörenden Versuchspersonen gewonnen wurden. Als Störgeräusch diente Sprachsimulierendes Rauschen, dessen Schallpegel 67 dB betrug. Die Sprachschalle wurden mit einem mittleren maximalen Pegel von 70 dB dargeboten. Bietet man die Sprache ohne Störgeräusch mit Inhibition dar wird eine Verständlichkeit  $v'$  von 94 % erreicht.  $v'$  repräsentiert in dem Diagramm die um die Ratewahrscheinlichkeit korrigierte Rohverständlichkeit  $v'_{21}$ .

**Fig.4:** Reimtestverständlichkeit  $v'$  normalhörender Versuchspersonen für den Sotscheck-Reimtest bei Störung durch Sprachsimulierendes Rauschen. Rauschpegel: 67 dB. Sprachpegel: 70 dB. Geschlossene Quadrate, links oben: Einzelwerte für Sprache ohne Störgeräusch mit Inhibition; Offene Kreise, offenes Quadrat, Bezeichnung "o1": Einzelwerte für Sprache mit Störgeräusch ohne Inhibition mit dazugehörigem Zentralwert und Wahrscheinl. Schwankungen; Geschlossene Kreise, geschlossenes Quadrat, Bezeichnung "m1": Einzelwerte, Zentralwert und Wahrscheinl. Schwankungen mit Inhibition ohne nichtlineare Tiefpässe; geschlossene Dreiecke, geschlossenes Quadrat, Bezeichnung "m2": Einzelwerte, Zentralwert und Wahrscheinl. Schwankungen mit Inhibition mit nichtlinearen Tiefpässen ( $\tau_{AnIT} = 10 \cdot \tau_{EnIT}$ ). Die Gewichte sind wie in Fig.3 angegeben belegt.



Bei gleichzeitig vorhandenem Störgeräusch ohne Inhibition beträgt der Zentralwert der Verständlichkeit  $v' = 68.5\%$ . Bei Betrieb des Netzwerks ohne nichtlineare Tiefpässe ändert sich die Sprachverständlichkeit nicht, der Zentralwert liegt bei  $v' = 70\%$ . Führt man die nichtlinearen Tiefpässe mit einer Ausschwingzeitkonstante  $\tau_{AnIT}$  ein, die um den Faktor 10 höher liegt als die Einschwingzeitkonstante  $\tau_{EnIT}$ , verbessert sich die Sprachverständlichkeit  $v'$  auf einen Zentralwert von 79.5 %.

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des Sonderforschungsbereichs 204 "Gehör", München, unterstützt.

#### Literatur

- /1/ Beckenbauer, Th., Einfluß vielkanaliger Inhibition auf die Sprache. Fortschritte der Akustik - DAGA '88, DPG-GmbH, Bad Honnef, 713, 1988.
- /2/ Pfeiffer, B. und Sotscheck, J., Versuch zur Sprachaudiometrie bei Lärm-schwerhörigkeit mit einem Reimtest aus der Nachrichtentechnik, BIA-Report 1/84, Berufsgenoss. Inst. f. Arbeitssicherheit, St. Augustin, 1984.