

SCHWANKUNGSSTÄRKE UND ZEITLICHE HÜLLKURVE VON SPRACHE UND MUSIK

Hugo Fastl

Institut für Elektroakustik, Technische Universität München

1. EINLEITUNG

Die Hörempfindung Schwankungsstärke wird durch langsame (< 20 Hz) Schallfluktuationen hervorgerufen [1]; sie weist als Funktion der Modulationsfrequenz eine Bandpaßcharakteristik mit einem Maximum bei 4 Hz auf [2]. Eine ähnliche Bandpaßcharakteristik zeigt die zeitliche Hüllkurve normal gesprochener fließender Sprache [3]. Der Sender und der Empfänger akustischer Information, d.h. Stimme und Gehör sind demnach gut aufeinander abgestimmt [4]. In dieser Arbeit wird diskutiert, inwieweit unterschiedliche raumakustische Bedingungen und Sprechweisen die Anpassung von Stimme und Gehör beeinflussen. Darüber hinaus wird dargestellt, welche Korrelationen zwischen der Hüllkurvenschwankung von Musik und der Hörempfindung Schwankungsstärke bestehen.

2. MESSMETHODEN

Die Messungen zur Schwankungsstärke sinusförmig amplitudenmodulierter Breitbandrauschen wurden mit sechs Versuchspersonen nach der Methode der Größenschätzung mit Ankerschall durchgeführt. Eine detaillierte Beschreibung dieser Experimente findet sich in [2]. Die Spektren der zeitlichen Hüllkurve von Sprache und Musik wurden durch Frequenztransformation um den Faktor 100 und anschließende Terzpegelanalyse bestimmt. Im einzelnen wurden folgende Schritte durchlaufen: Von dem auf Magnetband gespeicherten Schallsignal wurde mittels eines Einweggleichrichters mit Tiefpaß (50 Hz Grenzfrequenz) die zeitliche Hüllkurve extrahiert. Dieses Signal wurde bei 63 Hz steiflankig tiefpaßbegrenzt (Antialiasingfilter) und mit 160 Hz Abtastfrequenz in den Speicher eines Mikrorechners eingeschrieben, der jeweils 85 sec Hüllkurvensignal aufnehmen konnte. Der Speicherinhalt wurde zyklisch mit einer Abtastfrequenz von 16000 Hz ausgelesen und bei 4000 Hz tiefpaßbegrenzt. Die um den Faktor 100 frequenztransformierten Hüllkurvensignale mit 850 ms Periodendauer wurden mit einem Terzpegelanalysator mit 8 sec Integrationszeit spektral analysiert. Für jedes Schallsignal wurden vier Terzpegelspektren ermittelt und bei jeder Mittenfrequenz der arithmetische Mittelwert der vier Terzpegel gebildet. Die gemittelten Terzpegel wurden auf den maximalen gemittelten Terzpegel normiert und die zugehörigen Spannungsverhältnisse längs einer linearen Skala dargestellt.

3. SCHWANKUNGSSTÄRKE UND HÜLLKURVENSCHWANKUNG

3.1 Sprache Fig. 1 zeigt den Zusammenhang zwischen der Hörempfindung

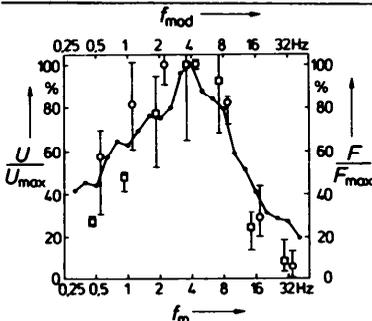


Fig. 1 Schwankungsstärke und zeitliche Hüllkurve fließender Sprache. Kurve: Terzspektrum der zeitlichen Hüllkurve fließender Sprache; auf den Maximalwert U_{max} normierte Ausgangsspannungen U von Terzfiltern der Mittenfrequenz f_m . Symbole: Auf den Maximalwert F_{max} normierte Schwankungsstärke F sinusförmig amplitudenmodulierter Breitbandrauschen als Funktion der Modulationsfrequenz f_{mod} (nach [2]).

Schwankungsstärke und der zeitlichen Hüllkurve normal gesprochener fließender Sprache eines männlichen Sprechers im reflexionsarmen Raum. Die Kurve repräsentiert das Terzspektrum der zeitlichen Sprachhüllkurve; das Verhältnis U/U_{\max} der auf den Maximalwert U_{\max} normierten Filterausgangsspannungen U ist als Funktion der Filtermittenfrequenz f_m dargestellt. Die Symbole verdeutlichen die Abhängigkeit der Schwankungsstärke sinusförmig amplitudenmodulierter Breitbandrauschen von der Modulationsfrequenz. Das Verhältnis F/F_{\max} der auf den Maximalwert F_{\max} normierten Schwankungsstärke F ist als Funktion der Modulationsfrequenz f_{mod} dargestellt (rechte Ordinate und obere Abszisse). Fig. 1 verdeutlicht die gute Übereinstimmung zwischen dem Hüllkurvenspektrum fließender Sprache und der Abhängigkeit der Hörempfindung Schwankungsstärke von der Modulationsfrequenz. Wie in der Natur zu erwarten, sind Stimme und Gehör ausgezeichnet aufeinander angepaßt.

Fig. 2a zeigt den Einfluß unterschiedlicher raumakustischer Bedingungen auf das Spektrum der Hüllkurvenschwankung normal gesprochener fließender Sprache. Von ein und demselben männlichen Sprecher wurde der gleiche Text in verschiedenen Räumen gesprochen. Die durchgezogene Kurve zeigt die Ergebnisse für den Fall, daß sich der Sprecher im reflexionsarmen Raum befindet. Die gestrichelte Kurve gilt für einen normalen Wohnraum (mittlere Nachhallzeit 0,5 sec) und die strichpunktierte Kurve für einen sehr halligen Raum (mittlere Nachhallzeit 1,8 sec). Für den reflexionsarmen Raum und für den Wohnraum ergeben sich recht ähnliche Hüllkurvenspektren mit einem Maximum bei 4 Hz, die gut mit der Schwankungsstärke übereinstimmen (vgl. Fig. 1). Dagegen zeigt sich für den halligen Raum ein völlig anderes Hüllkurvenspektrum mit einem Maximum bei 0,63 Hz und einer flachen oberen Flanke, da durch den Nachhall schnellere Schwankungen "verschmiert" werden.

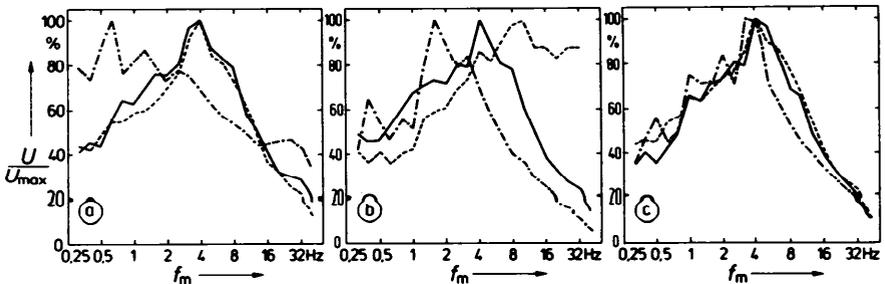


Fig. 2 Terzspektrum der zeitlichen Hüllkurve fließender Sprache.

(a) Einfluß der Raumakustik: Normal sprechender Sprecher im reflexionsarmen Raum (durchgezogen), im Wohnzimmer (gestrichelt), im halligen Raum (strichpunktiert). (b) Einfluß der Sprechweise: Normal sprechende Sprecherin (durchgezogen), flüsternde Sprecherin (gestrichelt), betont langsam und deutlich sprechender Sprecher (strichpunktiert). Reflexionsarmer Raum. (c) Fremdsprachen: Englisch (durchgezogen), Französisch (gestrichelt), Chinesisch (strichpunktiert). Nachrichtensprecher.

In Fig. 2b sind die Spektren der zeitlichen Hüllkurve fließender Sprache für verschiedene Sprechweisen dargestellt. Die durchgezogene Kurve gilt für eine Sprecherin, die den gleichen Text wie der Sprecher im reflexionsarmen Raum mit normaler Stimme liest. Die gestrichelte Kurve gilt für den Fall, daß die Sprecherin den Text flüstert. Die strichpunktierte Kurve repräsentiert das Hüllkurvenspektrum des Marburger Satzverständnistests, der zur Anpassung von Hörgeräten verwendet wird: ein männlicher Sprecher spricht in reflexionsarmer Umgebung betont langsam und deutlich. Die durchgezogenen Kurven in Fig. 2a bzw. Fig. 2b weisen eine Bandpaßcharakteristik mit einem Maximum bei 4 Hz auf und sind

praktisch identisch; für die Hüllkurvenschwankung fließender Sprache spielt das Geschlecht des Sprechers offensichtlich keine Rolle. Die strichpunktierte Kurve in Fig. 2b zeigt für den langsam sprechenden Sprecher ebenfalls eine Bandpaßcharakteristik, allerdings mit einem Maximum bei 1,6 Hz. Im Gegensatz dazu ergibt sich gemäß Fig. 2b für das Hüllkurvenspektrum der flüsternden Sprecherin (gestrichelt) eher eine Hochpaß- als eine Bandpaßcharakteristik (vgl. [5]). Das Hüllkurvenspektrum für die normal sprechende Sprecherin zeigt große Ähnlichkeit zur Abhängigkeit der Schwankungsstärke von der Modulationsfrequenz. Dies gilt auch für den langsam sprechenden Sprecher wenn die Bandpaßcharakteristik längs der Frequenzachse verschoben wird. Für die flüsternde Sprecherin weist das Hüllkurvenspektrum aber eine grundsätzlich andere Frequenzabhängigkeit als die Schwankungsstärke auf.

Fig. 2c zeigt das Terzspektrum der zeitlichen Hüllkurve fließender Sprache für verschiedene Fremdsprachen. Sämtliche analysierten Texte stammen von Nachrichtenprechern; sie wurden in reflexionsarmer Umgebung normal gesprochen. Die durchgezogene Kurve in Fig. 2c repräsentiert die Ergebnisse für englische Sprache, die gestrichelte Kurve für französische Sprache und die strichpunktierte Kurve für chinesische Sprache. Für den englischen und französischen Sprecher ergeben sich praktisch die gleichen Hüllkurvenspektren wie für den deutschen Sprecher oder die deutsche Sprecherin. Für die chinesische Nachrichtensprecherin liegt das Maximum des Hüllkurvenspektrums bei 3,15 Hz und es zeigen sich Nebenmaxima bei 2 Hz, 1 Hz und 0,5 Hz. Insgesamt weisen die zeitlichen Hüllkurvenschwankungen auch für Fremdsprachen die gleiche Frequenzabhängigkeit wie die Hörempfindung Schwankungsstärke auf.

3.2 Musik Als Musikmaterial wurden 60 je 5 sec lange Ausschnitte verschiedener Musikrichtungen verwendet, die in willkürlicher Reihenfolge ohne Pausen aneinandergereiht waren. Es handelt sich um 9 Beispiele aus klassischer Musik, 6 Märsche, 6 Schlagzeugsoli, 6 Ausschnitte aus Jazz-Musik sowie 14 Beispiele aus Pop- und 19 Beispiele aus Rock-Musik [6]. Fig. 3 zeigt das Terzspektrum der

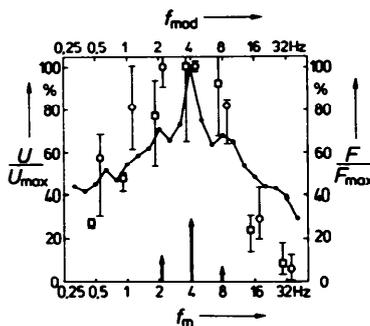


Fig. 3 Schwankungsstärke und zeitliche Hüllkurve von Musik.

Kurve: Terzspektrum der zeitlichen Hüllkurve von 60 Ausschnitten von je 5 sec Dauer aus Musikstücken unterschiedlicher Stilrichtungen. Symbole: Schwankungsstärke wie bei Fig. 1. Pfeile: Am häufigsten wahrgenommene Rhythmen der Musikstücke (nach [6]).

zeitlichen Hüllkurve dieses Musikmaterials (Kurve) sowie die Schwankungsstärke sinusförmig amplitudenmodulierter Breitbandrauschen als Funktion der Modulationsfrequenz (Symbole). Das Spektrum der zeitlichen Hüllkurve von Musik unterschiedlicher Stilrichtungen weist eine Bandpaßcharakteristik mit einem Maximum bei 4 Hz auf. Nebenmaxima treten bei etwa 2 Hz und 8 Hz auf. Insgesamt ist die Spektralverteilung der Hüllkurvenschwankungen für Musik (Fig. 3) etwas selektiver als diejenige für Sprache (Fig. 1). Dennoch zeigt auch das Hüllkurvenspektrum von Musik eine recht gute Übereinstimmung mit der Abhängigkeit der Hörempfindung Schwankungsstärke von der Modulationsfrequenz. Die Pfeile in Fig. 3 repräsentieren die Frequenz der Ereigniszeitpunkte, die Köhlmann [6] für das gleiche Musikmaterial gemessen hat, indem er Versuchspersonen den wahrgenommenen Rhythmus von Musik auf einer Morsetaste nachtasten ließ. Am häufigsten wurde die Morsetaste in Abständen von 250 msec (4 Hz) betätigt; dies entspricht sowohl der do-

minanten Hüllkurvenschwankung von 4 Hz als auch der maximalen Schwankungsstärke. Die Nebenmaxima der Hüllkurvenschwankung bei 2 Hz und 8 Hz spiegeln sich ebenfalls in den getasteten Rhythmen wider. Der Rhythmus von Musik wird demnach durch die Hüllkurvenschwankungen bestimmt. Dies gilt auch für fließende Sprache [6], wobei die Schwankungen der Lautheit [7] von Sprache und Musik besser mit dem wahrgenommenen Rhythmus korrelieren, als die Schwankungen der zeitlichen Hüllkurve. Aus diesem Grunde wurde für alle hier beschriebenen Schallsignale auch die Lautheitsschwankung analysiert, die an anderer Stelle dargestellt wird.

4. SCHLUSSBEMERKUNG

Die Hüllkurvenschwankung fließender Sprache, die in nicht zu halligen Räumen normal gesprochen wird und die Hörempfindung Schwankungsstärke zeigen praktisch die gleiche Frequenzabhängigkeit. Die Sprachverständlichkeit in Räumen wird demnach immer dann gut sein, wenn die originale Schwankungsstärke der Sprache nicht beeinträchtigt wird. Diese These wird durch Ergebnisse von Houtgast und Steeneken [3] bestätigt. Während diese Autoren von einer signalbezogenen Größe, der Hüllkurvenschwankung ausgehen, schlagen wir vor, die Sprachverständlichkeit anhand einer gehörbezogenen Größe, der Hörempfindung Schwankungsstärke, zu beurteilen. Für eine große Anzahl von Ausschnitten aus Musikstücken unterschiedlichster Musikrichtungen ließen sich ebenfalls Korrelationen zwischen der Hüllkurvenschwankung und der Schwankungsstärke aufzeigen. Allerdings handelt es sich hierbei um Mittelwerte; für einzelne Musikstücke bzw. einzelne Sätze aus klassischen Musikwerken können sich andere Hüllkurvenschwankungen ergeben (vgl. [8], S.44). Im Mittel entspricht jedoch die Frequenzabhängigkeit der zeitlichen Hüllkurvenschwankung von Musik derjenigen der Hörempfindung Schwankungsstärke. Darüber hinaus ergeben sich starke Korrelationen zur Rhythmuswahrnehmung: Im Bereich größter Empfindlichkeit der Schwankungsstärke finden sich die dominanten Hüllkurvenschwankungen von Sprache und Musik sowie deren häufigste Gliederung in rhythmische Ereignisse.

Der Autor dankt Herrn Dipl.-Ing.J.Mayer für die Durchführung eines Großteils der Hüllkurvenanalysen. Prof.Dr.-Ing.E.Zwicker und Prof.Dr.-Ing.E. Terhardt haben durch anregende Diskussionen wesentlich zum Gelingen der Untersuchungen beigetragen. Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen des SFB 204, Gehör, München gefördert.

Literatur

- [1] Terhardt, E., Über akustische Rauigkeit und Schwankungsstärke. *Acustica* 20, 215-224, 1968.
- [2] Fastl, H., Fluctuation strength of modulated tones and broadband noise. In: *Hearing - Physiological Bases and Psychophysics*, (R. Klinke & R. Hartmann, Eds.) Springer-Verlag, Berlin, 282-288, 1983.
- [3] Houtgast, T. und Steeneken, H.J., The modulation transfer function in room acoustics as a predictor of speech intelligibility. *Acustica* 28, 66-73, 1973.
- [4] Fastl, H., Beschreibung dynamischer Hörempfindungen anhand von Mithörschwellen-Mustern. Hochschul-Verlag, Freiburg 1982.
- [5] Steeneken, H.J. und Houtgast, T., The temporal envelope spectrum of speech and its significance in room acoustics. *Proc.11, ICA Paris, Vol.7*, 85-88, 1983.
- [6] Köhlmann, M., Rhythmische Segmentierung von Schallsignalen und ihre Anwendung auf die Analyse von Sprache und Musik. Dissertation, TU München, 1984.
- [7] Zwicker, E., Procedure for calculating loudness of temporally variable sounds. *J. Acoust. Soc. Am.* 62, 675-682, 1977.
- [8] Polack, J.D., Über die Modulationsübertragungsfunktion von Musiksignalen. Dissertation, Universität Göttingen, 1982.