

Dynamikreduktion von Sprachsignalen in konturierter Darstellung

STEFAN WARTINI, CLAUS v. RÜCKER

Fachgebiet Akustische Kommunikation, Technische Universität München

In der Regel ist der von Gehörgeschädigten nutzbare Dynamikbereich von Schallen stark eingeschränkt. Daher erscheint es sinnvoll, durch geeignete Verfahren die Dynamik des Signals zu reduzieren, möglichst ohne dadurch die informationstragenden Anteile zu beeinflussen. Die meisten herkömmlichen Verfahren bearbeiten das Zeitsignal, ohne die spektrale Information einzelner Signalanteile zu berücksichtigen. Diese Verfahren leiten aus der zeitlichen Hüllkurve des gesamten Signals oder einzelner Teilbänder eine Verstärkungsregelung ab, welche auf das Gesamtsignal oder einzelne Teilbänder wirkt [1]. Durch eine Spektraltransformation mit anschließender Konturierung kann eine gehörgerechte Repräsentation von Sprachschallen gewonnen werden, in der nur die wesentlichen Signalanteile enthalten sind. In der vorliegenden Arbeit werden sechs Verfahren zur Bearbeitung dieser als Teiltonzeitmuster (TTZM) bezeichneten Spektralkonturen vorgestellt und im Hinblick auf ihre Dynamikreduktion und ihren Einfluß auf Wiedergabequalität und Verständlichkeit untersucht.

Bearbeitungsverfahren

Das TTZM entsteht aus dem Zeitsignal durch Fourier-t-Transformation mit anschließender Glättung und Konturierung [2, 3]. Die Konturen werden in konstanten Zeitabschnitten, den sogenannten Auswerteintervallen abgetastet. Das so gewonnene TTZM kann auf einfache Weise bearbeitet werden. Sechs Bearbeitungsverfahren zur Dynamikreduktion durch Pegeländerung einzelner Teiltöne wurden auf die TTZM verschiedener Sprachschalle angewandt und hinsichtlich ihrer Verwendbarkeit überprüft:

- Entfernen von Teiltönen niedrigen Pegels;
- Entfernen von Teiltönen geringer Dauer;
- Pegelabsenkung des Teiltönen mit dem größten Pegel eines Auswerteintervalles (CUT);
- Veränderung der Teiltonpegel gemäß einer linearen Kennlinie mit Begrenzung (CLIP);
- Veränderung der Teiltonpegel gemäß einer quadratischen Kennlinie (POW);
- Veränderung der Teiltonpegel gemäß einer wurzelförmigen Kennlinie (SQRT).

Das verwendete Sprachmaterial stammte aus dem „Freiburger Wörkertest“ [4]. Es zeigte sich, daß die Verfahren, bei denen Teiltöne niedrigen Pegels oder geringer Dauer entfernt werden, die Signale entweder kaum in ihrer Dynamik beeinflussen oder bei Anhebung des Schwellenkriteriums das Gesamtsignal zeitweise vollständig auslöschen. Daher wurden diese Verfahren nicht zu Hörversuchen herangezogen.

Das Verfahren zur Pegelabsenkung des Teiltönen mit dem höchsten Pegel im Auswerteintervall, im folgenden mit CUT bezeichnet, wurde mehrfach auf die Testschalle angewandt, um eine Reduktion des Gesamtpegels zu erreichen. Dadurch wird dieses Verfahren stark von der Pegelverteilung in einem Auswerteintervall abhängig. Auswerteintervalle in denen ein einzelner Teilton im Pegel weit über den übrigen Teiltönen liegt erfahren durch die wiederholte Anwendung des Verfahrens eine hohe Pegelabsenkung. Treten dagegen mehrere Teiltöne ähnlichen Pegels auf, so werden sie durch die Mehrfachanwendung abwechselnd abgesenkt, woraus insgesamt eine geringere Absenkung des Pegels im Auswerteintervall resultiert. Die Prominenz einzelner Spektralbereiche wird durch die mehrfache Anwendung dieses Verfahrens je nach Teiltonpegelverteilung unterschiedlich beeinflusst.

Die Bearbeitung der Teiltöne durch eine lineare Kennlinie mit Begrenzung (Fig. 1 a), im folgenden mit CLIP bezeichnet, begrenzt hohe Teiltonpegel oberhalb eines Schwellwertes L_{grenz} auf einen Grenzwert L_{max} , während Teiltöne niedrigeren Pegels unverändert bleiben. Das Verfahren mit quadratischer Kennlinie (POW) senkt Teiltöne geringen und mittleren Pegels gemäß der Gleichung $L_A = L_{max} \cdot (L_E/L_{Emax})^2$ ab (Fig. 1 b), wobei L_{Emax} den maximalen Pegel des Eingangssignals bezeichnet. Auf diese Weise sollen Teiltöne erhalten bleiben, die auf Grund ihres hohen Pegels einen hohen Informationsgehalt vermuten lassen, während Teiltöne niedrigen Pegels weiter reduziert werden. Die Bearbeitung der Teiltonpegel gemäß einer Wurzelkennlinie (SQRT) hebt kleine und mittlere Pegel nach der Gleichung $L_A = L_{max} \cdot \sqrt{L_E/L_{Emax}}$ an, während hohe Teiltonpegel abgesenkt werden (Fig. 1 c). Daher verspricht dieses Verfahren eine erhebliche Reduktion des Dynamikbereiches.

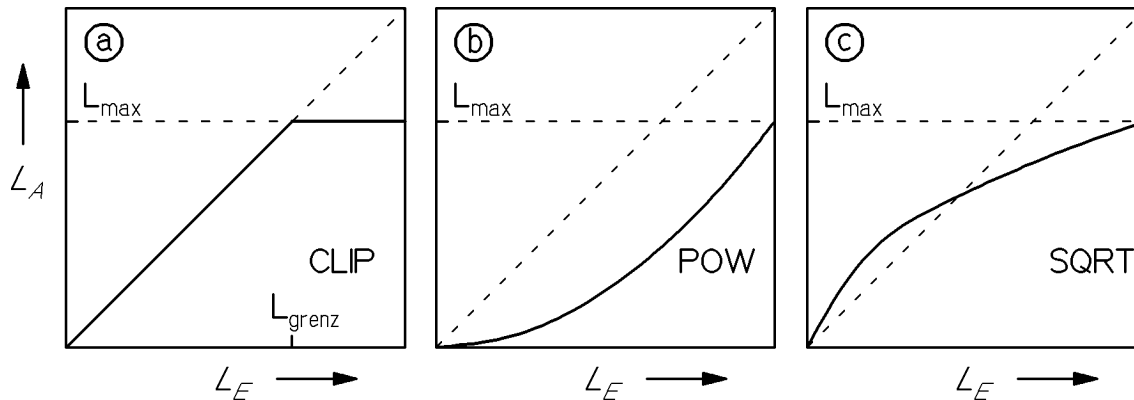


Fig. 1: Verlauf der verwendeten Kennlinien. a) Begrenzung der Teiltonpegel: Verfahren CLIP, b) Quadratische Kennlinie: Verfahren POW, c) Wurzelförmige Kennlinie: Verfahren SQRT.

Dynamikreduktion der Sprachschalle

Mit den Verfahren CLIP, CUT, POW und SQRT wurden Worte nach [4] bearbeitet, deren Maximalpegel ungefähr 70 dB betrug. Das Verfahren CUT wurde zwanzigmal auf die TTZM angewandt, wobei die Pegelabsenkung 3 dB betrug. Die Parameter L_{max} und L_{grenz} wurden für das Verfahren CLIP zu jeweils 50 dB gewählt. Für die Verfahren POW und SQRT wurden der maximale Eingangssignalpegel L_{Emax} auf 90 dB und der maximale Ausgangssignalpegel L_{max} auf 90 dB (POW) bzw. 60 dB (SQRT) gesetzt. Zur Beurteilung der erzielten Dynamikreduktion wurde aus den bearbeiteten TTZMn der Verlauf des Gesamtpegels über der Zeit durch Intensitätsaddition der einzelnen Teiltöne berechnet. Da es sich bei den untersuchten Signalen um Einzelwörter handelt, wurde die Dynamik durch die Pegeldifferenz zwischen dem Maximum und einem lokalen Minimum des Pegelverlaufs im Bereich der Wortmitte definiert. Fig. 2 zeigt exemplarisch die resultierenden Pegelverläufe des Wortes 'Durst', die sich aus der Anwendung der verwendeten Verfahren ergeben. Die Kurven wurden zur besseren Vergleichbarkeit mit dem Pegelverlauf des unbearbeiteten TTZMs auf den gleichen Maximalpegel verschoben. Während die Verfahren POW, CLIP und SQRT im Mittel über die vier im Hörversuch verwendeten Wörter zur Reduktion des Dynamikumfangs von ungefähr 3 dB, 9 dB und 16 dB führen, erzeugt das Verfahren CUT im Mittel eine leichte Dynamikzunahme um 0,5 dB. Dieser Wert resultiert aus der oben angesprochenen Problematik, daß die Gesamtpegelreduktion des Verfahrens stark von der Anzahl der Teiltöne mit hohem Pegel je Auswertintervall abhängt. Allgemein zeigte sich eine mäßige Abhängigkeit der Veränderung des Dynamikumfangs von den Wörtern. Nicht bei allen Wörtern konnte eine Dynamikreduktion erreicht werden. Vielmehr führten die Verfahren CUT und POW bei einzelnen Wörtern sogar zu einer Zunahme des Dynamikumfangs.

Hörversuche

Zum Vergleich der Verfahren wurden Hörversuche durchgeführt, in denen normalhörende Versuchspersonen die Wörter hinsichtlich ihrer Wiedergabequalität und ihrer Verständlichkeit beurteilten. Für die Versuche wurden die Einsilber 'Durst', 'Farm', 'Schreck' und 'Spott' aus [4] verwendet. Nach der Bearbeitung im TTZM wurden die Wörter durch Fouriersynthese resynthetisiert und in ihrem Gesamtpegel auf subjektiv gleiche Lautheit untereinander angeglichen. Die Wörter wurden mit einem Abhörpegel von 75 dB in einer schallisolierten Meßkabine über einen freifeldentzerrten Kopfhörer paarweise dargeboten (Abtastfrequenz 44,1 kHz, Quantisierung 16 Bit), wobei unbearbeitete und bearbeitete Worte mit Ausnahme der Paarung gleiches Wort und gleiches Verfahren miteinander kombiniert wurden.

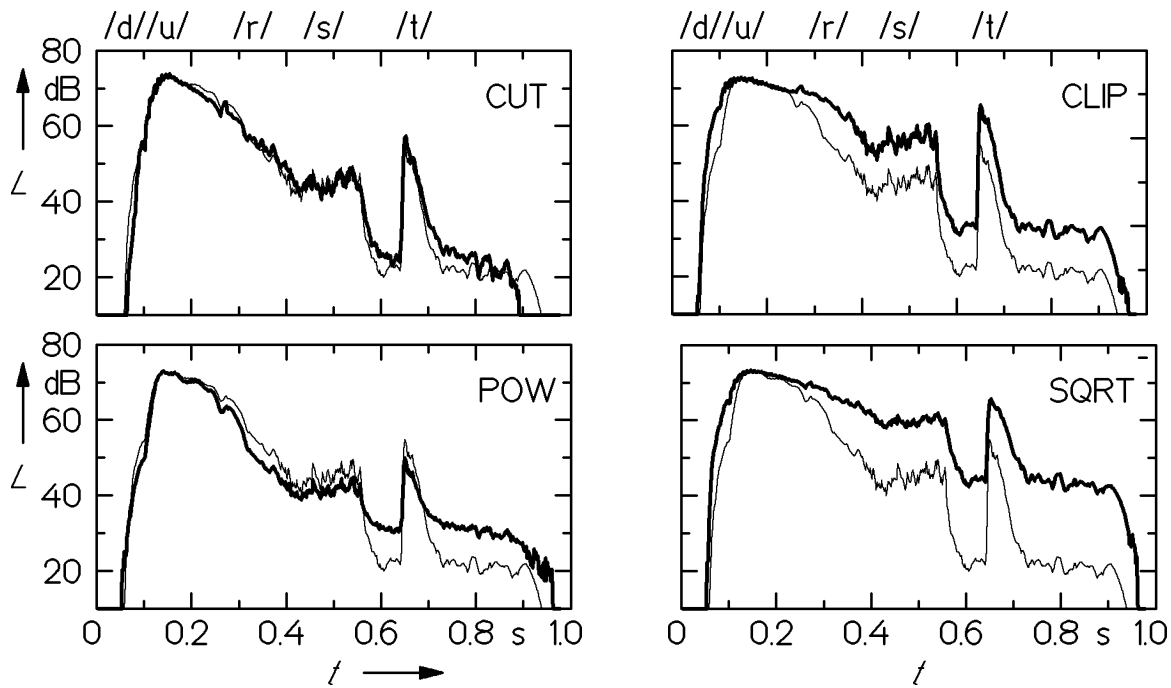


Fig. 2:

Verlauf des Gesamtpegels des Wortes 'Durst' als Funktion der Zeit. Die einzelnen Bilder zeigen die durch Intensitätsaddition der Teiltöne im TTZM gewonnenen Pegelverläufe des Originalsignals (dünne Linie) und der mit den Verfahren CUT, CLIP, POW und SQRT bearbeiteten Signale (dicke Linie). Die Pegelverläufe der bearbeiteten Signale wurden für die Darstellung auf gleichen Maximalpegel verschoben.

Die Versuchspersonen beurteilten den Unterschied der Wiedergabequalität dieser Wortepaarungen auf einer Kategorienskala zwischen 0 (kein Unterschied) und 6 (riesiger Unterschied). An den Versuchen nahmen fünf normalhörende Versuchspersonen teil. Für jede Versuchsperson wurde der Median der Beurteilungen jeder Verfahrenskombinationen unabhängig von den dargebotenen Worten ermittelt. Tab. I zeigt Mediane und Interquartilbereiche, die sich aus diesen Werten aller fünf Versuchspersonen ergeben. Die Qualitätsunterschiede zwischen den Verfahren wurden allgemein als 'gering' bis 'mittel' beurteilt. Der größte Unterschied in der Beurteilung verschiedener Verfahren entstand dort, wo gegensätzliche Kennlinien verwendet wurden (POW-SQRT) und bei der Darbietung von bearbeiteten Wörtern mit unbearbeiteten Wörtern. Es zeigt sich, daß eine Anhebung von Teiltönen im unteren und mittleren Pegelbereich (SQRT) die größte Dynamikreduktion (16 dB), aber auch den größten Qualitätsverlust im Vergleich mit allen anderen Verfahren verursacht. Die von den übrigen verwendeten Verfahren erzeugten Veränderungen scheinen keine so großen Qualitätsunterschiede untereinander

hervorzurufen, obwohl sie bis auf das Verfahren CUT den Dynamikumfang im Mittel um 3 – 9 dB reduzieren. Die Unterschiede ihrer Wiedergabequalität wurden als 'gering' bis 'mittel' beurteilt. Die Tabelle zeigt auch, daß die Wiedergabequalität auch bei der Darbietung von verschiedenen Wörtern, die mit demselben Verfahren bearbeitet waren, nicht mit 'kein Unterschied' beurteilt wurde. Dieser Umstand läßt sich darauf zurückführen, daß in die Beurteilungen der Versuchspersonen wohl nicht nur der Qualitätsunterschied sondern auch die mittlere Wiedergabequalität des Wortpaares einfließen. Da die unbearbeiteten TTZM durch das Resyntheseverfahren ebenfalls leichte Qualitätsverluste aufweisen, wurden auch die Beurteilungen der Paarungen unbearbeiteter Wörter (ORI-ORI) nicht mit 'kein Unterschied' beurteilt.

	ORI	CUT	CLIP	POW	SQRT	Kategorie	Wert
SQRT					$1 \begin{smallmatrix} 1 \\ 2,5 \end{smallmatrix}$	kein Unterschied	0
POW				$2 \begin{smallmatrix} 1 \\ 2,25 \end{smallmatrix}$	$3,5 \begin{smallmatrix} 3 \\ 4,5 \end{smallmatrix}$	sehr geringer Unterschied	1
CLIP			$2 \begin{smallmatrix} 1,25 \\ 2,75 \end{smallmatrix}$	$3 \begin{smallmatrix} 2 \\ 3,5 \end{smallmatrix}$	$3 \begin{smallmatrix} 2,5 \\ 4,25 \end{smallmatrix}$	geringer Unterschied	2
CUT		$2 \begin{smallmatrix} 1,75 \\ 3 \end{smallmatrix}$	$2 \begin{smallmatrix} 1,75 \\ 2 \end{smallmatrix}$	$2,5 \begin{smallmatrix} 2,5 \\ 3 \end{smallmatrix}$	$3 \begin{smallmatrix} 1,5 \\ 4,25 \end{smallmatrix}$	mittlerer Unterschied	3
ORI	$0,5 \begin{smallmatrix} 0 \\ 1,5 \end{smallmatrix}$	$2 \begin{smallmatrix} 2 \\ 3 \end{smallmatrix}$	$2 \begin{smallmatrix} 1,25 \\ 2,5 \end{smallmatrix}$	$2 \begin{smallmatrix} 1,5 \\ 2,5 \end{smallmatrix}$	$4 \begin{smallmatrix} 2,75 \\ 4,5 \end{smallmatrix}$	großer Unterschied	4
						sehr großer Unterschied	5
						riesiger Unterschied	6

Tab. I: Beurteilung von Unterschieden der Wiedergabequalität verschiedener Verfahrenskombinationen auf einer Kategorienskala. Die Abkürzung ORI kennzeichnet die aus den unbearbeiteten TTZMn resynthetisierten Schalle. Dargestellt sind die Mediane und Interquartilbereiche der Beurteilungen ermittelt aus den Beurteilungen aller Versuchspersonen. Trotz erheblich unterschiedlicher Dynamikreduktionen werden die Unterschiede zwischen den Verfahren als 'mittel' bis 'gering' beurteilt. Die größten Unterschiede ergeben sich zwischen Verfahren mit gegenläufiger Kennlinie (SQRT – POW).

In einer weiteren Versuchsreihe, die mit der 2I3AFC-Methode durchgeführt wurde, verglichen die Versuchspersonen die Verständlichkeit der Worte eines Wortpaares. Neben der Entscheidung, welches der Worte besser verständlich war, konnten auch beide Worte als gleich verständlich beurteilt werden. Die Bewertung der unterschiedlichen Verfahren zeigt, daß mit dem Verfahren CLIP bearbeitete Wörter im Vergleich zu allen anderen Verfahren als besser verständlich beurteilt wurden. Stellt man eine Rangfolge der Bearbeitungsverfahren mit abnehmender Verständlichkeit auf, so ergibt sich die Reihenfolge CLIP, CUT, POW und SQRT. Auch diese Bewertungen bestätigen die Tendenz der Qualitätsbeurteilungen, wonach die Anhebung niedriger und mittlerer Pegel zu einer Verschlechterung der Verständlichkeit führen.

Die Untersuchungen bestätigen die bereits früher getroffene Annahme, daß die Tonhöhen eine weitaus wichtigere Rolle in der Informationsübertragung spielen, als die Absolutwerte ihrer Amplituden. Alle Versuchspersonen konnten trotz der Qualitätsunterschiede die mit den untersuchten Verfahren bearbeiteten Wörter einwandfrei verstehen.

Die Untersuchungen wurden im SFB 204 „Gehör“ mit Unterstützung der DFG durchgeführt.

Literatur

- [1] Zwicker, E. und Zollner, M. (1984). Elektroakustik. Springer Verlag, 1. Auflage, S.278 ff.
- [2] Terhardt, E. (1985). Fourier Transformation of Time Signals: Conceptual Revision. In: *Acustica*, 57, S. 242–256.
- [3] Heinbach, W. (1988). Aurally adequate signal representation: The Part-Tone-Time-Pattern. In: *Acustica*, 67, S.113–121.
- [4] Westra Audiometric Disc Nr.11, (1992). Zahlen- und Wörtertest nach DIN 45621 mit Störgeräusch nach Prof. Dr.-Ing. H. Fastl.