Trennung von tonalen und geräuschhaften Anteilen im Sprachsignal

M. MUMMERT
(Lehrstuhl für Elektroakustik der Technischen Universität München)

Einleitung

Im Sprechorgan kommen mit Glottisschwingung, Turbulenzbildung an Engstellen und Verschlußbildung drei grundsätzlich verschiedene Anregungsquellen zum Einsatz. Entsprechend nimmt das Gehör im Sprachsignal "tonale" sowie als "geräuschhaft" zusammenfaßbare "rausch-" und "impulshafte" Anteile wahr. Ein getrennter Zugriff auf die Signalanteile, die diesen Empfindungen entsprechen, ist für sprachverarbeitende Systeme wünschenswert. Der unterschiedliche Frequenz- und Zeitauflösungsbedarf für die drei Anteile kann z.B. zur Datenreduktion genutzt werden.

Das hier beschriebene Verfahren stützt sich auf den Gedanken der spektralen Konturisierung nach Terhardt [1] und stellt eine Erweiterung des durch Heinbach [3] geschaffenen Teiltonzeitmuster-Verfahrens dar. Ausgangspunkt ist eine gehörgerechte Spektralanalyse [2] mit einem Fenster höherer Ordnung [4].

Spektrale Konturisierung

Sucht man im Pegelspektrum der Fourier-t-Transformation (FTT) [1] den Zeitverlauf der lokalen Maxima über der Frequenz, so erhält man ein Frequenzkonturmuster, das den Zeitverlauf der Teiltöne eines Schallsignals in Frequenz und Pegel beschreibt. Dieses sog. Teiltonzeitmuster bewahrt die wesentliche akustische Information des Sprachsignals, wie durch Synthese nachgewiesen wurde [3]. Somit sind neben tonalen auch geräuschhafte Anteile repräsentiert, die impulshaften Anteile erscheinen jedoch stark unterbewertet.

Das FTT-Pegelspektrum L(f,t) eines 1kHz-Sinustons, der in stationärem weißen Rauschen hart ein- (Fig.1a) und ausgeschaltet wird, erzeugt als Frequenzkontur eine gerade Teiltonlinie, das Rauschen zeigt sich in regellosen, dicht liegenden kurzen Teiltönen (Fig.2a). Die Spektrale Verbreiterung (vergl. Fig.1b, ohne Rauschen) kurz nach dem Schaltzeitpunkt wird durch Frequenzkonturisierung direkt nicht wiedergegeben; auch der Pegelverlauf an den Enden der 1kHz-Teiltonline enthält hierüber keine Information, da er durch die maximale Einschwingzeit des Analysefilters als Folge der begrenzten Bandbreite vorgegeben ist.

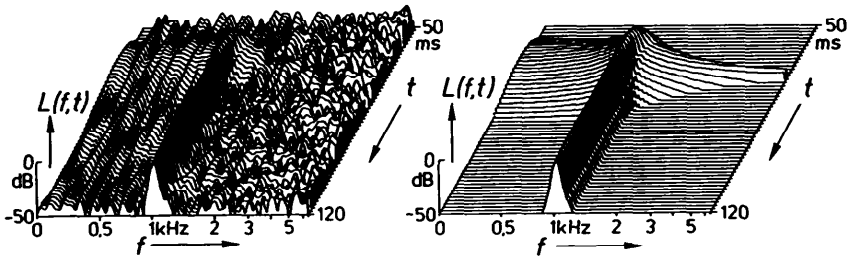

Fig.1a,1b: FTT-Pegelspektrum eines zu t=50ms eingeschalteten 1kHz-Sinustons, links mit bzw. rechts ohne stationäres weißes Rauschen (S/N = 22dB für 0-5kHz); Analysefenster ähnlich 4. Ordnung nach [4] mit Laufzeitausgleich.

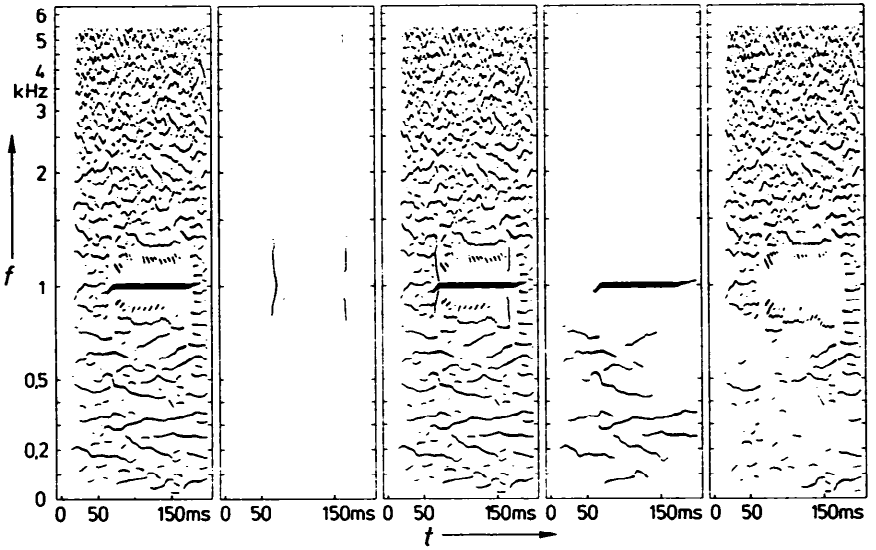

Fig.2a-e: Frequenzkontur (Teiltonzeitmuster), Zeitkontur (impulshafte Anteile), gesamtes Konturmuster, tonale bzw. rauschhafte Anteile der Frequenzkontur; zugrundeliegendes FTT-Pegelspektrum s. Fig.1a (Ton wird zu t=150ms wieder abgeschaltet); die Breite der Frequenzkonturlinien kennzeichnet den Pegel.

Durch Übertragung des Gedankens der spektralen Konturisierung auf die Zeitrichtung kann die spektrale Verbreiterung erfaßt und dadurch ein Zugang zu impulshaften Anteilen gewonnen werden. Hierzu wird die Anstiegsgeschwindigkeit $\partial L(f,t)/\partial t$ des Pegelspektrums $L(f,t)$ mit einem Schwellwert λ überwacht, der zum Zeitpunkt t_1 des Unterschreitens einen Konturpunkt in Frequenz und Pegel bestimmt (Fig.3). Der Pegel des Konturpunktes liegt dadurch um einen festen Wert ΔL unterhalb des Maximums bei Impulsanregung, welche für Frequenzen, die von der spektralen Verbreiterung betroffen sind, angenommen werden kann. Die sich auf diese Weise ergebende Zeitkonturlinie (Fig.2b) gibt in ihrem Frequenzverlauf die spektrale Verbreiterung annähernd pegelgetreu wieder. Das Zusammenfügen von Zeit- und Frequenzkonturmuster ergibt das spektrale Konturmuster (Fig.2c).

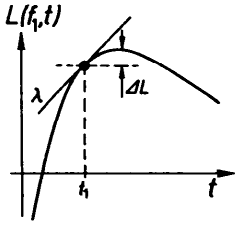

Fig.3: Konturpunktdetektion des FTT-Pegelspektrums $L(f,t)$ in Zeitrichtung bei $f=f_1$ (s.Text).

Isolierung tonaler Anteile

Die im Frequenzkonturmuster repräsentierten tonalen Anteile zeichnen sich durch glattere und längere Linienstücke gegenüber den meist kurzen und regellosen aus, die den rauschhaften Anteilen zuzurechnen sind. Die Einführung einer Mindestdauer erlaubt die Trennung dieser Anteile (Fig.2d,e). Für sie dient als Anhaltspunkt der Wert 50ms, er entspricht nach Messungen der Ausgeprägtheit von Sinustönen in

Abhängigkeit von der Dauer einer Halbierung der Ausgeprägtheit [5]. Linien tieffrequenter Rauschanteile können diese Länge überschreiten (vergl. Fig.2d), sie wurden in Sprachsignalen jedoch nicht beobachtet.

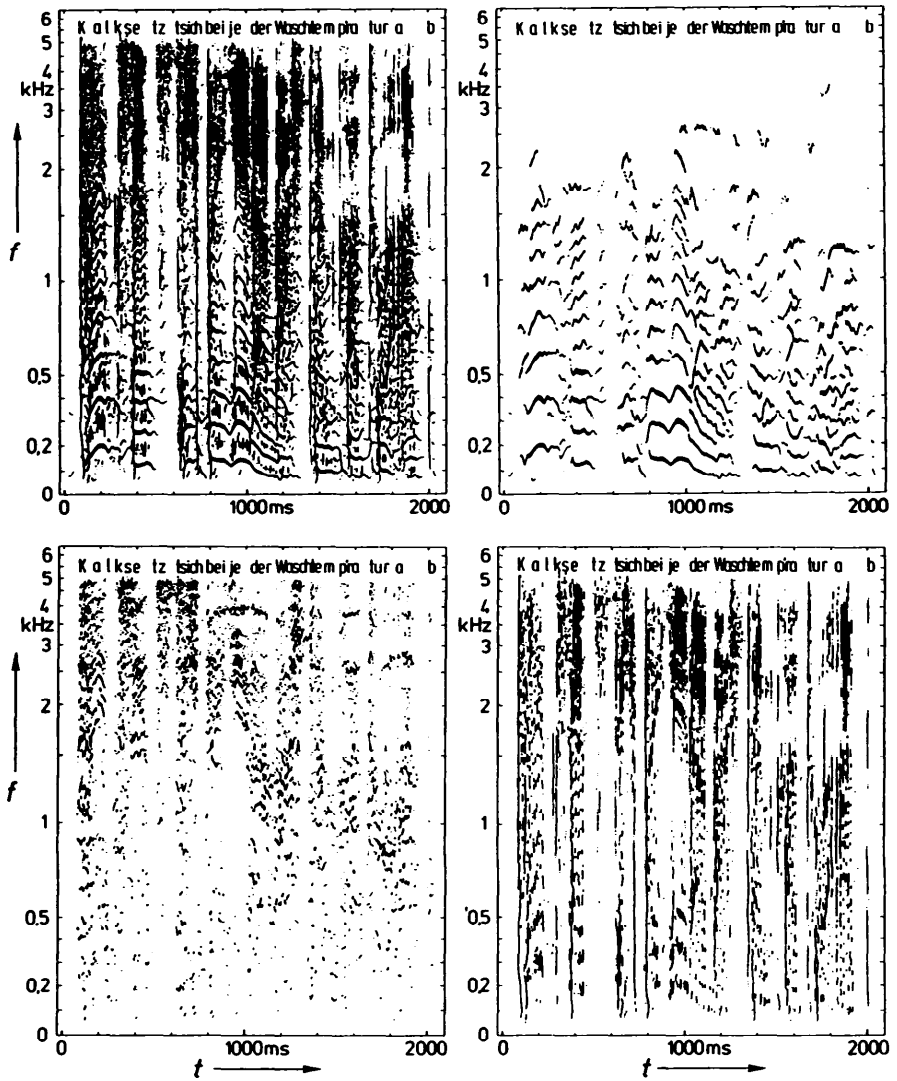

Fig.4a-d: Konturmuster, tonale bzw. rauschhafte Anteile der Frequenzkontur, Zeitkontur (impulshafte Anteile), jeweils für den Satz "Kalk setzt sich bei jeder Waschtemp'ratur ab", gesprochen von einem männl. Sprecher; die Breite der Frequenzkonturlinien kennzeichnet den Pegel.

Anwendung auf Sprache

Als Mindestlänge für die Isolierung der tonalen Anteile wurden ca. 50ms bei tiefen Frequenzen (0–500Hz) mit einer Abnahme zu hohen Frequenzen (25ms bei 5kHz) ermittelt, bei einer 3dB-Analysebandbreite B von 0.25 bis 0.3 Bark, und Fensterordnungen größer eins [4]. Für die Anstiegsschwelle der Zeitkontur ergibt sich ein günstiger Wert zu λ = 25dB·B.

Die Synthese der aus der Frequenzkontur gewonnenen Anteile erfolgt analog zum ungetrennten Fall wie in [4] in Ergänzung zu [3] beschrieben. Dieses Verfahren eignet sich auch zur ungefähren Synthese der Zeitkontur, wenn die den Zeitkonturpunkten zugeordneten Sinusgeneratoren durch eine ausreichend breite zeitliche Hüllkurve gesteuert werden.

Das Ergebnis des Trennverfahrens anhand des Testsatzes "Kalk setzt sich bei jeder Waschtemp'ratur ab", gesprochen von einem männlichen Sprecher, zeigt (Fig.4a-d), daß der Beitrag der Glottisschwingung gut durch die Frequenzkonturlinien mit einer Mindestlänge repräsentiert ist. Der Frikativanteil befindet sich fast ausschließlich im restlichen Frequenzkonturmuster. Das durch Zeitkonturisierung gewonnene Muster markiert deutlich den breiten Anstieg des Spektrums infolge impulsartiger Anregungen z.B. bei den Plosiven.

Zusammenfassung

Die Einführung einer Mindestdauer für Linien in der Frequenzkontur des FTT-Pegelspektrums, bekannt als Teiltonzeitmuster, erlaubt die Abtrennung tonal empfundener Anteile. Übrig bleiben die geräuschhaften Anteile, wobei impulshafte Anteile nur sehr schwach vertreten sind. Diese können durch eine spektrale Zeitkontur, die das Teiltonzeitmuster zu einem spektralen Konturmuster verallgemeinert, berücksichtigt werden. Im Gegensatz zum Frequenzkonturpunkt als lokales Maximum in Frequenzrichtung ergibt sich der Zeitkonturpunkt in Zeitrichtung im Augenblick des Unterschreitens eines Schwellwerts für die Anstiegsgeschwindigkeit. Die Anwendung des Verfahrens auf Sprache erlaubt eine Trennung der Beiträge von Glottisschwingung, Turbulenz- und Verschlußbildung im Sprechorgan. Eine akustische Überprüfung durch Synthese ist möglich.

Literatur

[1] Terhardt, E. : "Psychophysics of audio signal processing and the role of pitch in speech", The Psychophysics of Speech Perception (M. E. H. Schouten, Ed.), M. Nijhoff Publ., Dordrecht (1987) S. 271-283.

[2] Terhardt, E. : "Fourier tranformation of time signals, conceptual revision", Acustica vol. 57 (1985) S. 242-256.

[3] Heinbach, W. : "Gehörgerechte Repräsentation von Audiosignalen durch das Teiltonzeitmuster", Dissertation Technische Universität München (1988).

[4] Schlang, M., Mummert, M. : "Die Bedeutung der Fensterfunktion für die Fourier-t-Transformation als gehörgerechte Spektralanalyse", Fortschritte der Akustik, DAGA'90, Bad Honnef (1990), in diesem Tagungsband.

[5] Fastl, H. : "Pitch strength of pure tones", Proc. 13. ICA Belgrade, vol.3 (1989) S. 11-14.

Die Untersuchungen wurden im Sonderforschungsbereich 204 "Gehör", München, gefördert durch die Deutsche Forschungsgemeinschaft, durchgeführt.