VERSTÄNDLICHKEITSMESSUNGEN MIT DATENREDUZIERTEN NATÜRLICHEN EINZELVOKALEN

W. Heinbach
Lehrstuhl für Elektroakustik der Technischen Universität München

1. Einleitung

Mit Hilfe der Fourier-t-Transformation (FTT) /1,2/ kann man unter Verwendung von gehörbezogenen Analyseparametern ein Audiosignal in ein zeitvariables Muster aus Sinustönen überführen, welches als Teiltonzeitmuster (TTZM) bezeichnet wird. Aus dem TTZM läßt sich ein Audiosignal resynthetisieren, das sich gehörmäßig nur gering vom Original unterscheidet /2/. Eine Datenreduktion wird erzielt, indem zur Resynthese nur ein Teil der Teiltöne verwendet wird. Durch Verständlichkeitsmessungen mit derart resynthetisierten und datenreduzierten natürlichen Einzelvokalen wird der Einfluß der Teiltonauswahl untersucht. Im Folgenden wird über die Berechnung des Teiltonzeitmusters, die Datenreduktion durch Teiltonauswahl und die Verständlichkeitsmessungen berichtet.

2. Teiltonzeitmuster

Das Teiltonzeitmuster entsteht aus dem Zeitsignal durch Spektraltransformation, Betragsbildung, zeitliche Glättung und Maximumdetektion.

Das zeitvariable FTT-Leistungsspektrum der Analysefrequenz f berechnet man mit:

$$F(f,t) = \left[2a \int_0^t p(x)\, e^{-a(t-x)}\, e^{-j2\pi fx}\, dx \right]^2 \quad ;t>0 \qquad (1)$$

Das mit (1) berechnete Leistungsspektrum wird vor der Maximumdetektion zeitlich geglättet, um Nebenmaxima zu unterdrücken /1,2/:

$$G(f,t) = 1/T_G \int_0^t F(f,x)\, e^{-(t-x)/T_G}\, dx \quad ;t>0. \qquad (2)$$

Zur Berechnung des geglätteten Leistungsspektrums mit einem Digitalrechner wird das Audiosignal nach Tiefpassfilterung (f_g=5,6kHz) mit 1/T=12,8kHz abgetastet und mit 12Bit Auflösung analog-digital gewandelt.

Die Analyse erfolgt im Frequenzbereich von 20Hz bis 5kHz. Der Abstand der 380 Analysefrequenzen ist frequenzabhängig und entspricht 0.05Bark. Die Transformationskonstante a bestimmt die Analysebandbreite B = a/π und die effektive Analysefensterlänge T_F = 1/a. Der Wert von a hängt von der Analysefrequenz ab und wird so gewählt, daß sich eine Bandbreite von 0.1 Bark ergibt. (Ein Bark ist die psychoakustische Maßeinheit der Frequenzgruppenbreite /4/.)

Die Glättungszeitkonstante T_G ist ebenfalls frequenzabhängig und in ihrem Verlauf proportional zur Wahrnehmungsgrenze der Rauhigkeit amplitudenmodulierter Sinustöne /3/. Bis 3kHz verläuft T_G reziprok zu a mit T_G = 0.2/a und bleibt für höhere Frequenzen konstant. Die Berechnung von (1) und (2) erfolgt rekursiv und zu jedem Abtastzeitpunkt des Zeitsignals /1/.

Jedem relativen Maximum in G(f,t) über der Frequenz wird ein Teilton mit der Frequenz f_i und dem Pegel L_i zugeordnet, sofern es genügend ausgeprägt ist. Dazu muß der Abstand des Maximums zu den benachbarten Minima mindestens 3dB betragen.

Das Ensemble der so ermittelten Teiltöne, dessen Zusammensetzung sich mit der Zeit im allgemeinen fortwährend ändert, wird als Teiltonzeitmuster bezeichnet:

$$TTZM(t): \{f_1(t), L_1(t);..;f_i(t), L_i(t);..;f_{m(t)}(t), L_{m(t)}(t)\} \quad . \quad (3)$$

Außer den Frequenzen und Pegeln der Teiltöne ändert sich auch deren Anzahl in Abhängigkeit vom analysierten Signal über der Zeit; dies wird in (3) durch m(t) ausgedrückt. Das Teiltonzeitmuster beschreibt somit den Verlauf der Frequenzen und Pegel von m(t) Teiltönen.

Aufgrund der Glättung (2) und der endlichen Bandbreiten in (1) kann sich G(f,t) nicht beliebig schnell ändern /2/ und damit auch nicht die Frequenzen und Pegel der Teiltöne. Dies kann zur Auswertung von G(f,t) bzw. TTZM(t) in Zeitintervallen T_A, die wesentlich größer sind als das Abtastintervall T, ausgenutzt werden. Bei den in dieser Arbeit beschriebenen Messungen erfolgt die Bestimmung des TTZM aus dem geglätteten Leistungsspektrum im Zeitintervall T_A=5ms.

Die angenäherte Resynthese eines Zeitsignals erfolgt durch die Überlagerung der Zeitfunktionen der einzelnen Teiltöne in Abschnitten der Dauer T_A /2/. Da die Teiltöne voneinander unabhängige Beiträge zum resynthetisierten Zeitsignal liefern, ist es möglich, sie vor der Resynthese zu verändern oder sie ganz wegzulassen. Um Phasensprünge des resynthetisierten Zeitsignals von einem zum nächsten Syntheseabschnitt weitgehend zu vermeiden, werden die Zeitfunktionen der Teiltöne aufeinanderfolgender TTZM aneinandergebunden, wenn der Frequenzunterschied kleiner als 0,15Bark ist.

3. Datenreduktion

Um ein Zeitsignal durch sein TTZM zu beschreiben, sind folgende Parameter notwendig: Anzahl m der Teiltöne, Frequenzen der Teiltöne und Pegel der Teiltöne. Bei einer digitalen Übertragung oder Speicherung des TTZM müssen diese Parameter kodiert werden. Die notwendigen Wortlängen n_f für die Frequenz und n_L für den Pegel hängen von der Art der Quantisierung ab. Die momentane Datenrate des TTZM erhält man allgemein mit:

$$d=m_k*(n_f+n_L)/T_A \quad ;k=1,2,3... \quad . \quad (4)$$

Bei den im nächsten Abschnitt beschriebenen Verständlichkeitsmessungen wurde der Einfluß einer Begrenzung vom m(t) auf die Verständlichkeit untersucht. Die Frequenzen und Pegel der Teiltöne wurden mit n_f=16bit und n_L=8bit so genau kodiert, daß der Einfluß der Quantisierung auf die Qualität des resynthetisierten Signals vernachlässigbar ist. Ebenso ist das Auswerteintervall T_A mit 5ms genügend klein, um zeitliche Änderungen von G(f,t) zu erfassen.

Die Datenreduktion wird im wesentlichen durch Begrenzung der Anzahl der Teiltöne auf maximal N Teiltöne erzielt. Enthält das TTZM in einem Abtastintervall mehr als N Teiltöne, so wird eine Auswahl getroffen, wobei der Teiltonpegel als Kriterium dient. Dazu werden die Teiltöne eines Abtastintervalls nach fallendem Pegel geordnet und nur die ersten N Teiltöne zur Resynthese ausgewählt.

In Tabelle 1 sind die verwendeten Werte für N und die damit erzielten maximalen Datenraten angegeben, mit Ausnahme von Messung 2, bei der der Durchschnittswert angegeben ist. Dies entspricht einer durchschnittlichen Teiltonanzahl von N=12,65. Als Spitzenwert treten bis zu 60 Teiltöne auf.

Tabelle 1: Daten- und Fehlerraten der einzelnen Messungen

Messung	Testschalle	Datenrate kBit/s	Fehlerrate in % Gesamt	A	E	I	O	U
1	Digitalisierte Originalzeitf.	153	4	0	2,0	0,4	0,2	1,4
2	Resynthese, TTZM komplett	60,7	4,2	0	1,6	0,1	0,7	1,7
3	Resynthese, TTZM mit N=10	$\leq$48	4	0	1,8	0,1	0,5	1,5
4	Resynthese, TTZM mit N=5	$\leq$24	7,5	0	2,8	0,1	1,5	2,6
5	Resynthese, TTZM mit N=3	$\leq$14,4	23	0,1	11	6,8	0,8	5,0

In Fig. 1 und 2 sind die Teiltonzeitmuster zweier Einzelvokale mit nach rechts abnehmender Teiltonanzahl als "Maxigramme" dargestellt. Das Maxigramm zeigt den Verlauf der Teiltonfrequenzen über der Zeit, ohne jedoch eine Information über den Teiltonpegel zu enthalten. Die Frequenzachse ist entsprechend der Barkskala eingeteilt. Im Maxigramm kann man die Lage der Formanten und den Verlauf der Grundfrequenz gut erkennen. Durch die Teiltonauswahl über den Pegel verschwinden zuerst die Teiltöne, die nur wenig zu einem Formanten beitragen. Mit abnehmender Anzahl der Teiltöne verschwinden jedoch auch die Teiltöne, die schwache Formanten bilden, wie in Fig. 2b und c deutlich sichtbar.

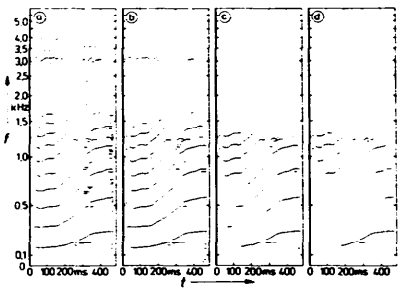

Fig. 1: Maxigramm des Vokals 'A' einer Frauenstimme.

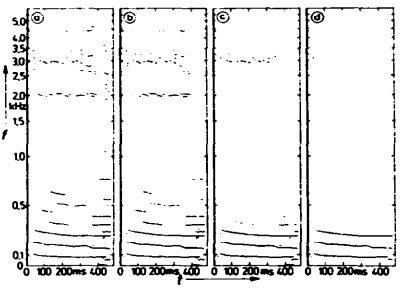

Fig. 2: Maxigramm des Vokals 'I' einer Männerstimme.

Gemeinsame Daten: a) gesamtes TTZM, b) zehn Teiltöne, c) fünf Teiltöne, d) drei Teiltöne

4. Verständlichkeitsmessungen

Als Testschalle der Verständlichkeitsmessungen wurden die Einzelvokale *A, E, I, O, U* von vier männlichen und vier weiblichen Sprechern verwendet, so daß insgesamt 40 Testschalle zur Verfügung standen. Die Vokale wurden in einer schallgedämmten Meßkabine auf Tonband aufgenommen und hatten eine Dauer zwischen 300 und 700ms. Die Sprecher wurden angewiesen, die Vokale einzeln und deutlich

zu sprechen; Anforderungen an die Grundfrequenz oder Gleichmäßigkeit der Artikulation wurden nicht gestellt.

An den Messungen nahmen insgesamt neun normalhörende Versuchspersonen im Alter zwischen 25 und 33 Jahren teil; zwei davon waren zugleich Sprecher. Fünf Versuchspersonen hatten nur geringe oder keine Erfahrung in Hörversuchen. Bei einer Messung wurde jeder der 40 Testschalle zweimal dargeboten, so daß eine Versuchsperson insgesamt 80 Antworten geben mußte. Die Darbietung erfolgte in zufälliger Reihenfolge diotisch über Kopfhörer in einer schallgedämmten Meßkabine. Bei den Messungen 4 und 5 wurde die Zufallsliste gegenüber den Messungen 1 bis 3 geändert. Nach Darbietung eines Testschalles mußte die Versuchsperson innerhalb von drei Sekunden mit der Angabe einer der genannten Vokale auf einem Protokollblatt antworten. Andere Antworten oder Enthaltungen wurden nicht zugelassen.

In Tabelle 1 sind die relativen Fehlerraten (Gesamtfehler bzw. Fehler pro Vokal) aufgeführt. Zwischen den Messungen 1 bis 3 zeigt sich keine signifikante Änderung der Fehlerrate. Dies entspricht auch dem Höreindruck bezüglich der Sprachqualität und der Wiedergabe sprechertypischer Merkmale. Die zunehmende Fehlerrate bei den Messungen 4 und 5 ist vor allem auf das Fehlen der Teiltöne, die den zweiten Formanten bilden (Fig. 2d), zurückzuführen. Dies zeigt sich deutlich im starken Ansteigen der Fehlerraten von E und I. Dagegen wird der Vokal A in seiner Verständlichkeit durch die Reduzierung der Teiltonanzahl nicht oder nur gering beeinträchtigt.

5. Zusammenfassung und Ausblick

Die natürlichen Einzelvokale A, E, I, O, U von acht Sprechern wurden mit Hilfe der FTT-Spektralanalyse in Teiltonzeitmuster übergeführt. Durch Resynthese wurden daraus, unter Weglassen von Teiltönen, Zeitsignale zurückgewonnen. Die Verständlichkeit der so datenreduzierten Vokale in Abhängigkeit von der Anzahl der Teiltöne wurde mit neun Versuchspersonen bestimmt. Dabei zeigte sich, daß bei Verwendung von zehn Teiltönen keine Änderung der Fehlerrate gegenüber den Originalsignalen auftritt. Bei Verwendung von weniger als fünf Teiltönen steigt der Fehler deutlich an. Hier kann eventuell durch andere Auswahlkriterien eine Verbesserung erreicht werden. Die Datenraten lassen sich durch entsprechende Kodierung der Teiltonfrequenzen und -pegel weiter verringern. Dadurch erhält man Datenraten von 8 - 16kbit/s bei sehr guter Sprachqualität.

6. Literatur

/1/ Terhardt, E.: Fourier transformation of time signals, conceptual revision. Acustica, 57 (1985), 242-256

/2/ Heinbach, W.: Untersuchung einer gehörbezogenen Spektralanalyse mittels Resynthese. In: "Fortschritte der Akustik" (DAGA '86), Bad Honnef 1986, 453-456

/3/ Terhardt, E.: Über die durch amplitudenmodulierte Sinustöne hervorgerufene Hörempfindung. Acustica 20 (1968), 210-214

/4/ Zwicker, E.: "Psychoakustik", Springer Heidelberg/New York (1982).

Diese Arbeit wurde im Sonderforschungsbereich 204 "Gehör", München, gefördert durch die Deutsche Forschungsgemeinschaft, durchgeführt.