BMBF

Verbmobil
Verbundvorhaben

# Bandbegrenzte Funktionen als Hilfsmittel zur Fehlerkorrektur von Grundfrequenzverläufen

Michael Lehning

TU Braunschweig

Vm Report 72
Juli 1995

Juli 1995

Michael Lehning

Institut für Nachrichtentechnik
Technische Universität Braunschweig
Schleinitzstr. 22
38092 Braunschweig

Tel.: (0531) 391 - 2453
Fax: (0531) 391 - 8218
e-mail: m.lehning@tu-bs.de

**Gehört zum Antragsabschnitt:** 14.3 Werkzeuge zur Segmentation und Etikettierung

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 N 0 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei dem Autor.

# Bandbegrenzte Funktionen als Hilfsmittel zur Fehlerkorrektur von Grundfrequenzverläufen *

Michael Lehning
Institut für Nachrichtentechnik, TU Braunschweig

## Überblick

Der vorliegende Artikel beschreibt ein Verfahren zur globalen Annäherung von Grundfrequenzverläufen durch bandbegrenzte Funktionen. Der Grundfrequenzverlauf wird dabei durch eine Fourierreihe so angenähert, daß der mittlere quadratische Fehler zwischen den gegebenen Grundfrequenzwerten und den Funktionswerten an diesen Stellen minimal wird. Aus dem Vergleich von Grundfrequenzverlauf und interpolierender Funktion können mögliche Meßfehler detektiert werden.

## 1 Einleitung

Die Grundfrequenz (F0) von gesprochener Sprache hat auf die Wahrnehmung prosodischer Eigenschaften wie Betonung und Akzentuierung einen entscheidenen Einfluß. Sie prägt neben der Sprachintensität und der zeitlichen Strukturierung (z.B. Pausen, Dehnungen) der Sprache den prosodischen Eindruck des Hörers.

Aus der Auswertung dieses prosodischen Merkmales können wichtige Informationen für die Semantik und Pragmatik einer Äußerung gewonnen werden [PAULUS1990], [SHAUGHESSY1983]. Die Nutzung dieser Information in einem automatischen System erfordert aber eine weitgehend fehlerfreie Bestimmung des Grundfrequenzverlaufes.

Für die Bestimmung des Grundfrequenzverlaufes von gesprochener Sprache gibt es zahlreiche Algorithmen [HESS1983]. Aber nur sehr wenige Verfahren berücksichtigen dabei die Eigenschaft, daß sich die Grundfrequenz von Sprache auf Grund der anatomischen Gegebenheiten des menschlichen Sprechapparates nicht beliebig schnell ändern kann.

In dem vorgestellten Verfahren wird daher vorgeschlagen, den gemessenen und fehlerbehafteten Grundfrequenzverlauf durch eine bandbegrenzte Funktion in seinem globalen Verlauf anzunähern, so daß auch größere fehlerbehaftete Bereiche detektiert werden können. Lokal operierende Verfahren (z.B. Glättung mit Median-Filtern oder lineare Interpolation [KUNDU1987]) versagen an solchen Stellen.

Als bandbegrenzte Funktion wird eine Fourierreihe gewählt. Die Vorfaktoren für die harmonischen Funktionen werden dabei so bestimmt, daß die gegebenen nicht äquidistanten F0-Werte optimal angenähert werden. Als Fehlerkriterium wird dabei die Summe der quadratischen Distanzen zwischen den gegebenen Grundfrequenzwerten und den an diesen Stel-

* Das diesem Beitrag zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministers für Forschung und Technologie unter dem Förderkennzeichen 01 IV 101 N 0 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor

len bestimmten Funktionswerten herangezogen. Für die Lösung des Problems wird auf das Verfahren der Singulärwertzerlegung (Singular Value Decomposition) zurückgegriffen [SULLIVAN1984],[WILKINSON1986].

Jedem F0-Wert kann dabei auch ein Wert für die statistische Sicherheit, mit der der F0-Wert berechnet wurde, zugeordnet werden. Die statistische Sicherheit kann entweder aus dem F0-Berechnungsverfahren selbst (bei Korrelationsverfahren z.B. der erreichte Korrelationsgrad) oder aus anderen Sprachsignalparametern (z.B. Stimmhaftigkeit) gewonnen werden.

In einem Nachverarbeitungsschritt können dann die Werte, die zu sehr vom interpolierten Verlauf abweichen, verworfen bzw. durch Alternativwerte, die auch von anderen F0-Berechnungsverfahren geliefert werden können, ersetzt werden. Dieses Verfahren erlaubt eine gute Detektion bzw. Korrektur von Ausreißern und Oktavfehlern. Dieses Vorgehen kann iterativ wiederholt werden, bis der F0-Verlauf hinreichend gut durch die bandbegrenzte Funktion angenähert wird. Auf diese Weise kann ein fehlerbereinigter F0-Verlauf bestimmt werden.

## 2 Bandbegrenzte Grundfrequenzinterpolation

Wie o.a. sollen die gemessenen Grundfrequenzwerte $s_1, \ldots, s_L$, die zu den Zeitpunkten $t_1 \ldots t_L$ gemessen wurden, durch eine bandbegrenzte Funktion angenähert werden. Neben der Glättung des geg. Verlaufes werden auch Lücken, die durch stimmlose Bereiche entstanden sind, geschlossen. Um für den Grundfrequenzverlauf eine Annäherung durch eine periodische Funktion durchführen zu können, denkt man sich den Grundfrequenzverlauf mit der Periodendauer $T$ periodisch fortgesetzt. Die Näherungswerte $\hat{s_i}$ werden aus der folgenden Funktion bestimmt:

$$\hat{s}(t) = A_0 + \sum_{n=1}^{K} [A_n \cos(2\pi \frac{n}{T} t) + B_n \sin(2\pi \frac{n}{T} t)] \qquad (1)$$

Gleichung (1) beschreibt die interpolierende Funktion als Fourierreihenentwicklung, die nach dem $K$-ten Glied abbricht und somit bandbegrenzt ist. Die $2K + 1$ Fourierkoeffizienten sind derart zu bestimmen, daß der Gesamtfehler $Err$ zwischen gemessenen und angenäherten Werten minimal wird [REINECKE1991].

$$Err = \sum_{i=1}^{L} (s_i - \hat{s}_i)^2 \longrightarrow min \qquad (2)$$

Gleichung (2) zeigt die Definition des Gesamtfehlers $Err$, wobei $s_i$ ein gegebener $F_0$-Wert zum Zeitpunkt $t_i$ ist und $\hat{s}_i$ der zugehörige Näherungswert. Der Gesamtfehler ist minimal, wenn alle $2K + 1$ partiellen Ableitungen von $Err$ nach den Koeffizienten $A_0$,...,$A_K$ und $B_1$,...,$B_K$ verschwinden [MELENTJEW1967] . Die partiellen Ableitungen bilden ein lineares Gleichungssystem zur Bestimmung der $2K + 1$ unbekannten Fourierkoeffizienten (siehe Anhang)

Die Lösung dieses Gleichungssystems ist nicht trivial, da es mitunter schlecht konditioniert sein kann [VARAH1973]. Aus diesem Grund wird auf die sogenannte Singulärwertzerlegung zurückgegriffen. Das beschriebene Verfahren beinhaltet zwei frei wählbare Parameter $T$ und $K$, durch die der Verlauf der interpolierenden Funktion beeinflußt werden kann. Mit einer Erhöhung der Oberwellenzahl ($K$) nähert sich die Funktion den gegebenen Werten besser an, was allerdings auch dazu führen kann, daß an fehlerhaften Stellen im F0-Verlauf die interpolierende Funktion diese Fehler gut annähert und somit deren Detektion verhindert.

Wird andererseits die Anzahl der Oberwellen zu klein gewählt, ist es möglich, daß der Grundfrequenzverlauf stark geglättet wird und somit F0-Werte in Bereichen starker F0-Bewegungen als Fehler gewertet werden.

## 3 Beispiele

Der Grundfrequenzverlauf des folgenden Beispielsatzes wurde mit dem korrelationsbasierten AMDF-Verfahren gemessen [ROSS1974]. Im Abstand von 10 ms wurde ein Grundfrequenz-wert ermittelt, wenn die gemittelte Energie innerhalb des Analysefensters über einer frei wählbaren Energieschwelle lag. Das Sprachmaterial stammt aus dem BMFT-Projekt Verbmobil. Bei dem Szenario handelt es sich um eine Terminabsprache ohne visuellen Kontakt zwischen zwei Personen.

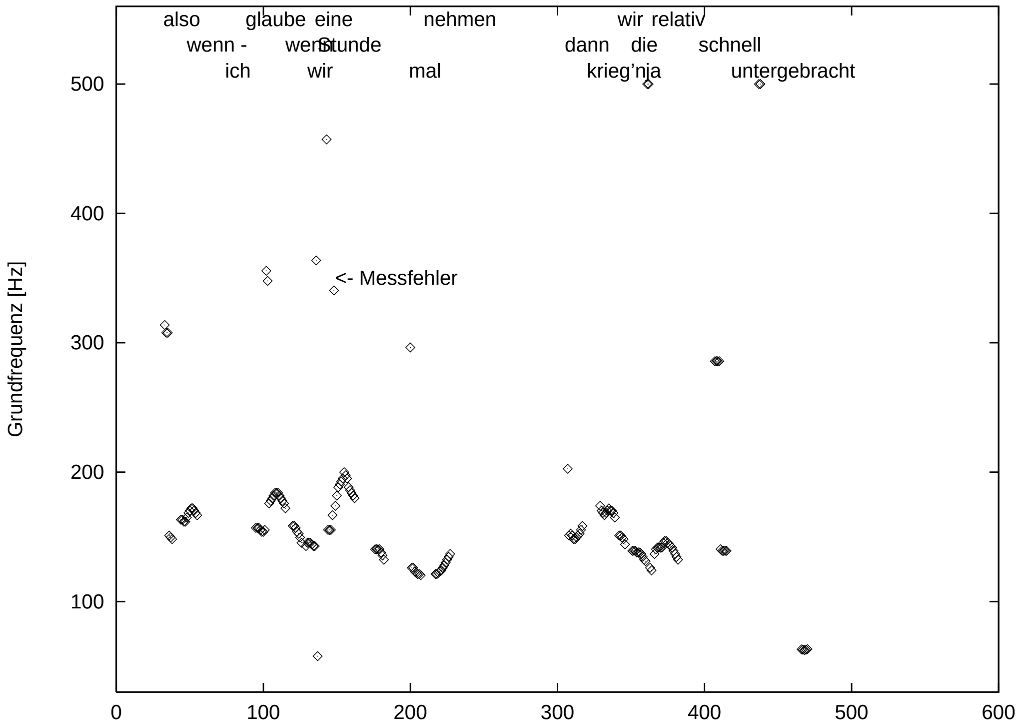


*Bild 1: Gemessener Grundfrequenzverlauf*

Man sieht, daß der dargestellte Grundfrequenzverlauf Fehler aufweist. Die Annäherung des Grundfrequenzverlaufes durch eine Fourierreihe vierter Ordnung ($K = 4$) mit einer Grundperiode von $T = 2000$ ergibt den in Bild $2$ dargestellten Verlauf $a$).

Da üblicherweise in Bereichen höherer Intensität und Stimmhaftigkeit der Grundfrequenzverlauf sicherer bestimmt werden kann als in den Bereichen niedriger Intensität, ist es sinnvoll, jedem F0-Wert noch einen Konfidenzwert $K_i = \frac{1}{\sigma_i}$ zuzuordnen, mit dem der Abweichungsfehler zwischen interpolierender Funktion und jedem einzelnen Meßwert gewichtet werden kann [PRESS1988], S. $521$ff:

$$Err = \sum_{i=1}^{L} K_i (s_i - \hat{s}_i)^2 \longrightarrow min \tag{3}$$

Als Konfidenzwert wird im folgenden Beispiel die Energie des mit $800Hz$ tiefpaßgefilterten Sprachsignals herangezogen (siehe Grundfrequenzverlauf $b$) im Bild $2$).

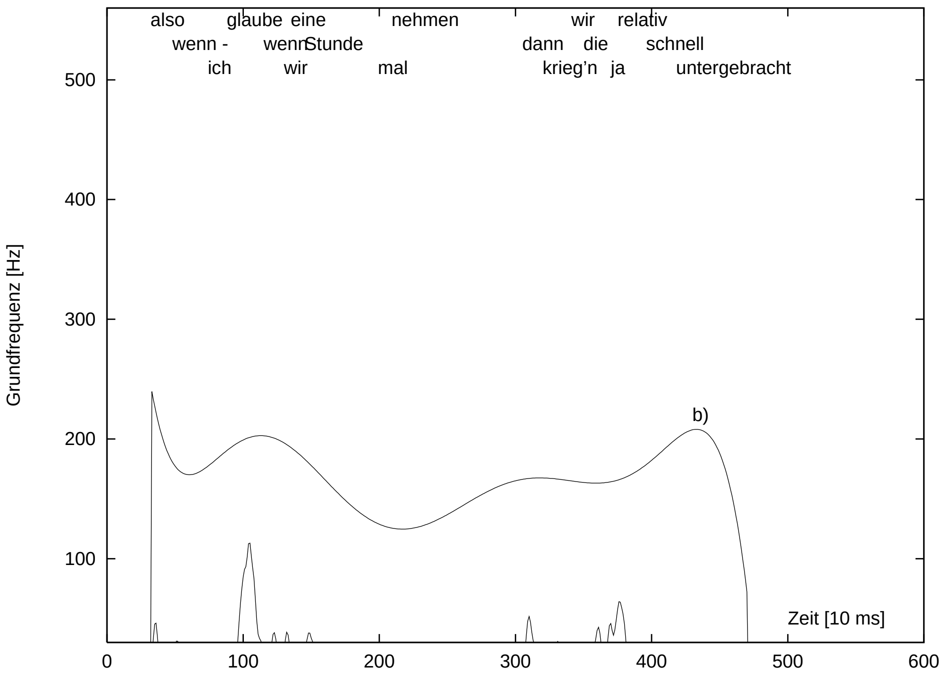


*Bild 2: Interpolierter Grundfrequenzverlauf (im unteren Bildbereich ist der Verlauf der Sprachsignalenergie vergrößert dargestellt)*

Man sieht, daß die Tendenz der interpolierenden Funktion $b)$ in Richtung der Meßfehler auszureißen, deutlich abgenommen hat.

Durch Vergleich der interpolierenden Funktion mit den Originalwerten lassen sich nun die Meßfehler beseitigen. Eine erneute Interpolation des bereinigten Verlaufes ist im folgenden Bild dargestellt:

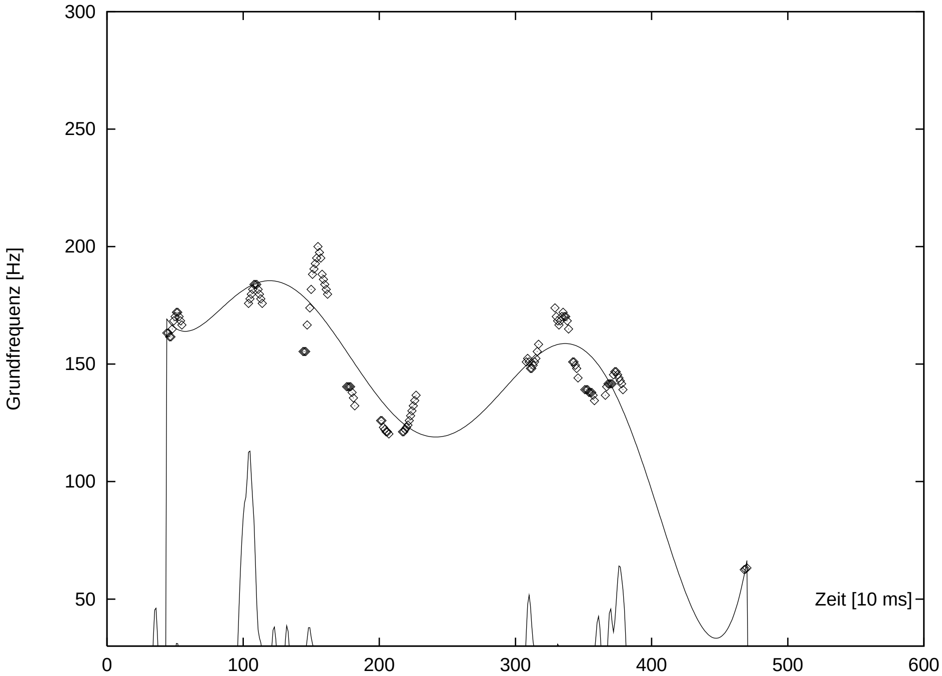


*Bild 3: Interpolation des bereinigten Grundfrequenzverlaufes (mit Energieverlauf)*

## 4 Zusammenfassung

Aus der Darstellung ist ersichtlich, daß die Detektion von Ausreißern und Blockfehlern im F0-Verlauf mit dem beschriebenen Verfahren gut möglich ist. Eine Anwendung auf andere fehlerbehaftete und bandbegrenzte Merkmalsverläufe (z.B. Formantverlauf) ist problemlos möglich, da das Verfahren kein spezifisches Wissen über den Verlauf der Grundfrequenz benutzt. Neben der Beseitigung der Meßfehler können Meßlücken durch das Verfahren geschlossen und eine parametrische Beschreibung des Verlaufes durch Auswertung der Koeffizienten der Fourierreihe gewonnen werden.

## Literatur


[HESS1983] W. Hess: *Algorithms and Devices for Pitch Determination of Speech Signals.* Springer Verlag, Berlin, 1983

[KUNDU1987] A. Kundu, S.K. Mitra:*A Computationally Efficient Approach to the Removal of Impulse Noise from Digitized Speech.* IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-35, No. 4, April 1987, 571 - 574

[PAULUS1990] E. PAULUS et al.: *Der Nutzwert prosodischer Merkmale für die automatische Spracherkennung, Proc. Elektronische Sprachsignalverarbeitung, Berlin (1990), 71-78*

[SHAUGHESSY1983] O'Shaughessy D., Allen J.:*Linguistic modality effects on fundamental frequency in speech, J. Acoust. Soc. Am. 74(4), October 1983, 1155-1171.*

[WILKINSON1986] J.H. WILKINSON, C. REINSCH: *Handbook for Automatic Computation, Vol 2: Linear Algebra*, Springer (1986)

[ROSS1974] M. Ross et al.: *Average Magnitude Difference Function Pitch Extractor*,IEEE Trans. ASSP-22, S. 353-362, 1974

[SULLIVAN1984] B.J. SULLIVAN, B. LIU: *On the Use of Singular Value Decomposition and Decimation in Discrete-Time Band-Limited Signal Extrapolation. IEEE Trans. ASSP-32(1984), 1201-1212.*

[VARAH1973] J.M. VARAH: *On the Numerical Solution of Ill-conditioned Linear Systems with Applications to Ill-posed Problems. SIAM J. Numer. Anal. Vol. 10, No. 2, April 1973*

[MELENTJEW1967] P.W. MELENTJEW, H. GRABOWSKI: *Näherungsmethoden*, VEB Fachbuchverlag Leipzig (1967)

[REINECKE1991] J. REINECKE, M. LEHNING: *Interpolation und Glättung von Sprachgrundfrequenzverläufen durch bandbegrenzte Funktionen. Proceedings DAGA 91, Bochum*

[PRESS1988] W.H. PRESS et al.: *Numerical Recipes in C — The Art of Scientific Computing.* Cambridge University Press (1988), ISBN 0-521-35465-X


## A Näherung von tabellarisch gegebenen Werten durch eine bandbegrenzte Funktion

### A.1 Theoretische Grundlagen

Es sind $L$ Werte $s_i(t_i)$ gegeben. Die $t_i$ können ungleichabständig sein. Durch diese L Werte soll nun eine Funktion $f_n(t) = F(t; v_0, v_1, \ldots, v_{n-1})$ gelegt werden. Diese Funktion besitzt demnach $n$ freie Parameter. Die Parameter sollen so bestimmt werden, daß die Summe der quadratischen Abweichungen zwischen der Funktion und den gegebenen Werten minimal wird. Die Näherung erfolgt also nach dem *Prinzip der minimalen Summe der Fehlerquadrate.* Diese Forderung läßt sich mathematisch wie folgt formulieren:

$$S(v_0, v_1, \ldots, v_{n-1}) = \sum_{i=1}^{L} [s_i - f_n(t_i)]^2 \longrightarrow min$$

## A.2 Berechnung der Lösung

Zur Erfüllung dieser Forderung ist es notwendig, daß die $n$ partiellen Ableitungen von $S(v_0, v_1, \ldots, v_{n-1})$ sämtlich null sind. Aus dieser Forderung ergibt sich somit folgendes Gleichungssystem:

$$\begin{aligned}
\sum_{i=1}^{L} [s_i - f_n(t_i)] F_{v_0}(t_i) &= 0 \\
\sum_{i=1}^{L} [s_i - f_n(t_i)] F_{v_1}(t_i) &= 0 \\
\sum_{i=1}^{L} [s_i - f_n(t_i)] F_{v_2}(t_i) &= 0 \\
&\vdots \\
\sum_{i=1}^{L} [s_i - f_n(t_i)] F_{v_{n-1}}(t_i) &= 0
\end{aligned}$$

Hierbei soll $F_{v_j}(t_i)$ der Wert der partiellen Ableitung der Funtion $f_n(t)$ nach dem Parameter $v_j$ an der Stelle $t_i$ sein.

## A.3 Anwendung des Verfahrens für die Fourierreihe

Im vorliegendem Fall sollen die vorgegebenen Werte durch eine Fourierreihe angenähert werden. Bei einer Fourierreihe, die bis zum Index $f_{max}$ entwickelt wird, ergeben sich $2f_{max} + 1$ Parameter. Im vorliegenden Fall wird die Normalform der Fourierreihe benutzt:

$$f_n(t) = A_0 + \sum_{k=1}^{f_{max}} A_k cos(\frac{2\pi}{P} kt) + \sum_{k=1}^{f_{max}} B_k sin(\frac{2\pi}{P} kt)$$

P entspricht dabei der Periodenlänge der Funktion $f_n(t)$. Die Zuordnung zu den o.g. Parametern wird nun wie folgt vorgenommen:

$$\begin{aligned}
A_0 &= v_0 \\
A_1 \ldots A_{f_{max}} &= v_1 \ldots v_{f_{max}} \\
B_1 \ldots B_{f_{max}} &= v_{f_{max}+1} \ldots v_{2f_{max}}
\end{aligned}$$

Werden die Parameter so umbenannt, ergibt sich folgende Darstellung der Fourierreihe:

$$f_n(t) = v_0 + \sum_{k=1}^{f_{max}} v_k cos(\frac{2\pi}{P} kt) + \sum_{k=f_{max}+1}^{2f_{max}} v_k sin(\frac{2\pi}{P}(k - f_{max})t)$$

Für die partiellen Ableitungen ergeben sich somit folgende Gleichungen:

$$\frac{df_n(t)}{dv_l} = \begin{cases} 1 & f\ddot{u}r \quad l = 0 \\ cos(\frac{2\pi}{P} lt) & f\ddot{u}r \quad l = 1 \ldots f_{max} \\ sin(\frac{2\pi}{P}(l - f_{max})t) & f\ddot{u}r \quad l = f_{max} + 1, \ldots, 2f_{max} \end{cases}$$

Mit diesen partiellen Ableitungen ergibt sich folgende Darstellung des oben angegebenen Gleichungssystems:

$$\begin{aligned}
\sum_{i=1}^{L} f_n(t_i) &= \sum_{i=1}^{L} s_i \\
\sum_{i=1}^{L} f_n(t_i) cos(\frac{2\pi}{P} t_i) &= \sum_{i=1}^{L} s_i cos(\frac{2\pi}{P} t_i) \\
\sum_{i=1}^{L} f_n(t_i) cos(2\frac{2\pi}{P} t_i) &= \sum_{i=1}^{L} s_i cos(2\frac{2\pi}{P} t_i) \\
&\vdots \\
\sum_{i=1}^{L} f_n(t_i) cos(f_{max}\frac{2\pi}{P} t_i) &= \sum_{i=1}^{L} s_i cos(f_{max}\frac{2\pi}{P} t_i) \\
\sum_{i=1}^{L} f_n(t_i) sin(\frac{2\pi}{P} t_i) &= \sum_{i=1}^{L} s_i sin(\frac{2\pi}{P} t_i) \\
\sum_{i=1}^{L} f_n(t_i) sin(2\frac{2\pi}{P} t_i) &= \sum_{i=1}^{L} s_i sin(2\frac{2\pi}{P} t_i) \\
&\vdots \\
\sum_{i=1}^{L} f_n(t_i) sin(f_{max}\frac{2\pi}{P} t_i) &= \sum_{i=1}^{L} s_i sin(f_{max}\frac{2\pi}{P} t_i)
\end{aligned}$$

Dieses Gleichungsystem läßt sich auch in Matrizenschreibweise angeben:

$$[A]\vec{v} = \vec{b}$$

Die Matrix $A$ teilt sich insgesamt in neun Bereiche auf:

$$A = \begin{bmatrix} I & II & III \\ IV & V & VI \\ VII & VIII & IX \end{bmatrix}$$

Die Elemente $a_{gh}$ der verschiedenen Bereiche der Matrix $A$ bestimmen sich folgendermaßen:

| | | |
|---|---|---|
| $I$ | $a_{0,0} = L$ | |
| $II$ | $a_{0,h} = \sum_{i=1}^{L} cos(h\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ h = 1, \ldots, f_{max}$ |
| $III$ | $a_{0,h} = \sum_{i=1}^{L} sin((h - f_{max})\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ h = f_{max} + 1, \ldots, 2f_{max}$ |
| $IV$ | $a_{g,0} = \sum_{i=1}^{L} cos(g\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ g = 1, \ldots, f_{max}$ |
| $V$ | $a_{g,h} = \sum_{i=1}^{L} cos(g\frac{2\pi}{P}t_i)cos(h\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ g = 1, \ldots, f_{max}$ |
| | | $und\ h = 1, \ldots, f_{max}$ |
| $VI$ | $a_{g,h} = \sum_{i=1}^{L} cos(g\frac{2\pi}{P}t_i)sin((h - f_{max})\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ g = 1, \ldots, f_{max}$ |
| | | $und\ h = f_{max} + 1, \ldots, 2f_{max}$ |
| $VII$ | $a_{g,0} = \sum_{i=1}^{L} sin((g - f_{max})\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ g = f_{max} + 1, \ldots, 2f_{max}$ |
| $VIII$ | $a_{g,h} = \sum_{i=1}^{L} sin((g - f_{max})\frac{2\pi}{P}t_i)cos(h\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ g = f_{max} + 1, \ldots, 2f_{max}$ |
| | | $und\ h = 1, \ldots, f_{max}$ |
| $IX$ | $a_{g,h} = \sum_{i=1}^{L} sin((g - f_{max})\frac{2\pi}{P}t_i)sin((h - f_{max})\frac{2\pi}{P}t_i)$ | $f\ddot{u}r\ g = f_{max} + 1, \ldots, 2f_{max}$ |
| | | $und\ h = f_{max} + 1, \ldots, 2f_{max}$ |

Die Elemente des Vektors $\vec{b}$ bestimmen sich wie folgt:

$$b[j] = \begin{cases} \sum_{i=1}^{L} s_i & f\ddot{u}r\ j = 0 \\ \sum_{i=1}^{L} s_i cos(j\frac{2\pi}{P}t_i) & f\ddot{u}r\ j = 1, \ldots, f_{max} \\ \sum_{i=1}^{L} s_i sin((j - f_{max})\frac{2\pi}{P}t_i) & f\ddot{u}r\ j = f_{max} + 1, \ldots, 2f_{max} \end{cases}$$

Dieser Report erschien als Beitrag im Tagungsband:
*Elektronische Sprachsignalverarbeitung 1994, Berlin, 166 - 173*