PARAMETRIC BASED TRANSFORMATION OF SPEECH SIGNALS

F K Fink, U Hartmann, K Hermansen

DSP-Group, Dept. of Communication Technology, Institute for Electronic Systems
Fr. Bajers Vej 7, DK-9220 Aalborg, AALBORG UNIVERSITY, DENMARK

ABSTRACT

A concept for parametric transformation of speech signals - PARTRAN - is presented in this paper. Using the concept we can present the speech information of interest in a frequency range at choice. For profoundly deaf people having only a minor frequency range available for reception of speech information this is an interesting feature as they do not benefit sufficiently from standard wideband hearing aids. Our concept includes 1) parametric modeling of the speech production system, 2) transformation of the speech production model to match the available frequency range, 3) resynthesis of the speech using the transformed model. Supplementary, this concept is believed to reduce wideband background noise which is a problem for hearing disabled as well as for people with normal hearing ability. The technique is well suited for real time implementation (VLSI), primarily caused by numeric robustness of the algorithms.

RÉSUMÉ

Cet article présente le nouveau concept de transformation paramétrique - PARTRAN - des signaux de parole. A l'aide de ce concept, le signal de parole utile peut être délivré dans une gamme de fréquence au choix. Il s'agit là d'une caractéristique intéressante pour les personnes malentendantes ne percevant qu'une gamme de fréquence restreinte du signal de parole, ces personnes ne tirant pas pleinement profit des appareils auditifs classiques à large bande. Notre concept se décompose en: 1) modélisation paramétrique du système de production de parole, 2) transformation du signal de parole produit afin de l'ajuster à la gamme de fréquence désirée, 3) reconstitution du signal à l'aide du modèle transformé. D'autre part, ce concept est à même de diminuer le bruit de fond à large bande, ce dernier étant un problème autant pour les personnes malentendantes que pour les personnes entendant normalement. Cette technique est bien adaptée à une implémentation temps-réel (VLSI), essentiellement en raison de la robustesse des algorithmes de calcul numérique.

1 INTRODUCTION

Commonly severe hearing losses are located in the high-frequency range, i.e. in the frequency range of most importance for speech perception. On this background several attempts have been made in the past decades to develop methods for transposing high frequency parts of speech into the lower frequency range available for most hearing impaired people. Attempts have been made to transform frequency ranges above 1000 Hz down below 1000 Hz using continuosly distributed frequencies or singletones representing different high-frequency bands. Results have never been convincing [1],[2].

This paper presents a new concept, a Parametric Transformation - PARTRAN, making it possible to transpose the speech signal in a very flexible way aiming a substantial support in distinguishing speech sounds.

2 BACKGROUND

Background noise is one of the main reasons hearing impaired people do not benefit sufficiently from usual hearing aids. Also for people with normal hearing ability noisy environments can cause inconvenience and reduced speech intelligibility. Harmonics of the pitch frequency in between the formants can be considered as "background noise" masking important cues of the speech signal. In some situations when these harmonics are comparable in effect to (minor) formants this problem can be severe. Decreasing the content of pitch harmonics will then to some extend decrease the contamination of wide band background noise. Therefore, our concept will be of interest even to people with normal hearing ability in noisy environments, aiming at increased distinction of speech sounds.

Transformation of the speech signal must not cause any

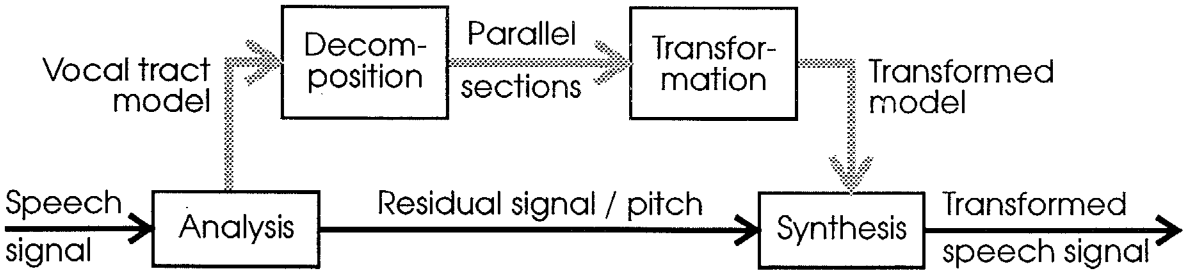

Figure 1: PARTRAN concept for speech transformation.

possible confusion of the listener. Therefore it is very important that high-frequency acoustic cues are not presented in frequency bands already occupied by low-frequency parts of the speech signal.

On this background we can conclude that the transformation must include:

1. Removal of harmonics of the pitch between the formants as well as removal of background noise.
2. Entire speech frequency range must be part of the transformation.
3. Different degrees of transformation of different frequency ranges according to individual needs.

This work enable speech understanding based on use of a reduced frequency range preserving all frequency information in a compressed form.

3 SIGNAL PROCESSING

The signal processing can be divided into four parts as shown i figure 1:

- parametric analysis of the speech signal,
- decomposition,
- transformation of the parametric model
- speech synthesis based on of the transformed model.

3.1 Speech Analysis

Initially the speech signal is analysed using traditional LPC modeling. This analysis separates the speech signal into

- a set of model coefficients describing the position of the vocal tract and

- a residual signal representing the pitch signal.

In this work, a 12th order LPC model is estimated every 2.5 msec based on 300 samples from the speech signal sampled at 8 kHz. The residual signal is used as exitation (pitch) in the synthesis process ensuring the transformed speech being "speech like".

3.2 Pseudodecomposition

The transformation of the speech spectrum is based on the assumption that the regions of maxima of the speech power density spectrum describes the primary cues necessary for speech understanding. Therefore the LPC-model is split into a number of separate second order sections each representing such a maximum - a resonans frequency of the vocal tract. Ideally these sections should be the true decomposition of the estimated LPC - model, but instead we introduce what is called a pseudodecomposition.

We estimate the high-Q and high-power poles from the LPC-spectrum directly by determining the resonans frequency and the 3dB bandwith respectively. Figure 2 shows an example of LPC-poles and figure 3 the poles and zeros resulting from the pseudodecomposition. From the estimated LPC-poles we form a number of second order sections each appended with 2 zeros, one in (1,0) and one in (-1,0) in the z-plane. As a consequence of the pseudodecomposition also some zeros are present in between the dominant poles (figure 3). The decomposed second order sections characterised by the three elements - resonans frequency, bandwidth and power - are the building blocks in forming the transformed speech spectrum.

As a consequence of the pseudodecomposition we can not expect the correct composed spectrum, zeros are introduced as shown in figure 3. However, experiments have shown negligible unwanted effects in the resulting power spectrum. Indeed it is seen from figure 4 that the zeros introduced in the decomposition supports the wanted effect of removal of pitch harmonics and background noise. Obviously we can exploit the flexibility given in the decomposition concept. A flexibility we use in the speech transformation in order to adjust individually the resonans frequency, power and bandwith.

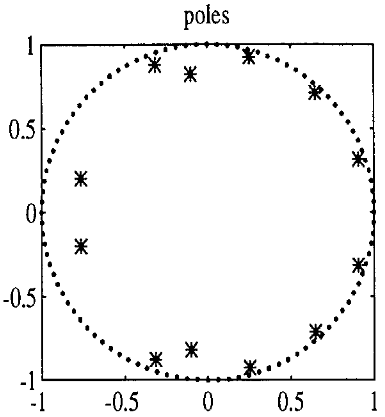

Figure 2:True poles for a 12 order LPC model.

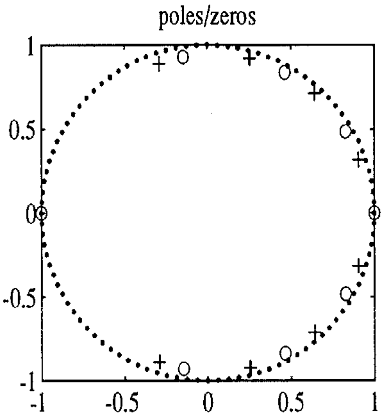

Figure 3: Poles and zeros resulting from pseudodecomposition.

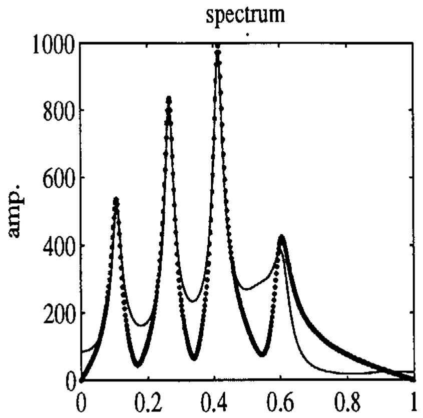

Figure 4:LPC - spectrum (solid) and pseudodecomposed spectrum.

3.3 Transformation

Each of the second order sections of interest is described by the following parameters: center frequency f_0, bandwidth BW and power P each of which is transformed separately. This is a very important option of the PARTRAN concept.

First the resonance frequencies of the parallel sections are transformed individuelly by a piecewise linear transformation. This transformation can be defined by specifying frequency ranges that are unchanged, compressed more or less or even omitted.

Second the bandwidth BW is transformed. We have used two possibilities: Multiplying each bandwidth by a specified number or - most promising - transforming each BW into a constant preassigned value. Instead of transforming the bandwidth one could choose transforming the Q-value. We propose to use the same transformation for Q as for BW.

Third transformation option is to change the power P in a second order section. We have used the identity function.

In this presentation we have choosen a transformation that demonstrates some of the possibilities. The transformation leaves all frequensies from 0 to 500 Hz unchanged, squeezes all frequency information from 500 Hz to 2.5 kHz into a narrow range from 500 Hz to 1.5 kHz and deletes all information above 2.5 kHz. In another work [4] we have shown results from squeezing all the frequency range 0-4 kHz into a narrow band centered about 1 kHz.

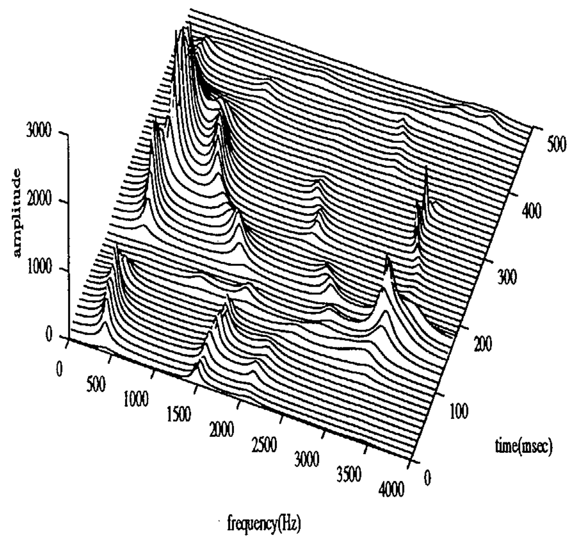

Figure 5: Original LPC spectra of the speech signal.

Figure 5 shows an example of the original LPC power spectra of the speech signal. Transforming center frequencies as described above, setting each bandwidth to 80 Hz and keeping the power in each second order section unchanged results in a set of transformed parametric models with the power spectra shown in figure 6.

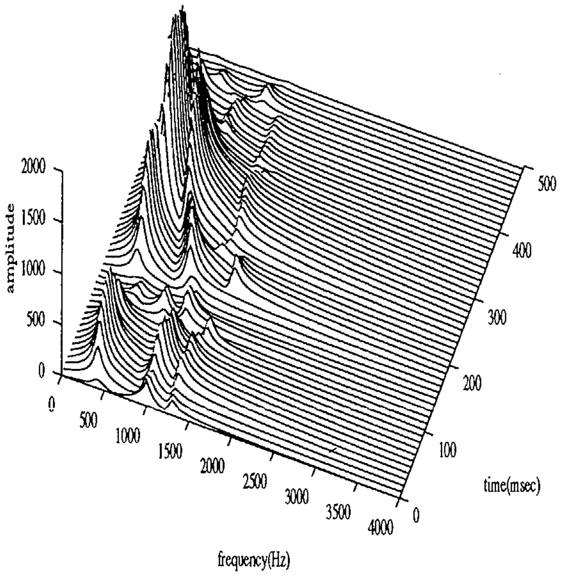

Figure 6: Spectra of the transformed parametric models.

3.4 Resynthesis

Given the transformed parallel filter sections a single transfer function (MA type) is composed. This can be seen as a model of a transformed vocal tract and used as such for speech synthesis. The 'pitch' input for this speech synthesis is based on the residual signal from the initial speech analysis. We found the best input signal to be a modified residual signal, still 'pitch'-like but more 'white' than the original residual signal. Transients caused by shifting the filter coefficients seems to be negligible, primarily because shifting rate is high (every 20 sample).

4 RESULTS

This paper reports on results from evaluating the PARTRAN concept using the described transformation. We have shown how each sound element described via center frequncy f_0, bandwidth BW and power P can be transformed individually. Our experiments shows that the concept works well. The flexibility of the PARTRAN concept allows us to combine the transformation of sound elements. One possibility is to reduce the number of sound elements by combining two elements to one preserving the impression of frequency contents and power.

The speech signal is sampled at 8 kHz. Based on a framelength of 300 samples (37.5 msec), a 12th order LPC model is estimated every 2.5 msec. The power density spectra are represented by 12 LPC coefficients. As a result of the parametric modeling the pitch and its harmonics are inhibited in the power density spectrum and the major peaks (formants) are easily identified. Filtering the speech signal by the inverse of the LPC model, the residual signal i.e. the part of the speech signal not included in the model, is found. By successful modelling the residual signal will be "near white", which is essential for the syntesis part.

The decomposition, transformation and recomposition is evaluated by listening and inspection of spectrograms. Listening test show good "speech-like" performance. An advantage of the PARTRAN concept is the preservation of power and intonation of the speech signal. Unwanted side effects caused by lack of prerequisitions such as small correlation between neighbouring second order sections seems to be of minor significance in practice.

5 CONCLUSION

We have developed a new concept, the PARTRAN concept. The pseudodecomposition and transformation concept seems to be very attractive seen in the light of its flexibility, modularity and numerical robustness. Furthermore the technique is wellsuited for VLSI - implementation.

REFERENCES

[1] Braida, L.D., Durlach, N.I., Lippmann, R.P., Hicks, B.L., Rabinowitz, C.M., Reed, C.M: "Hearing aids - A rewiev of past research on linear amplification, amplitude compression and frequency lowering", ASHA Monographs No. 19, 1979.

[2] Risberg, A.: "Speech Coding in Aids for the Deaf: an overview of research from 1924 to 1982", QPSR 4/1982, Speech Transmission Lab., KTH, Stockholm, Sweden

[3] Kuwabara, H., Takagi, T.: "Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method", Speech Communication 10, 1991.

[4] Hartmann, U., Hermansen, K., Fink, F.K.: "Feature extraction for Profoundly Deaf People", EUROSPEECH'93, Berlin, Germany, 1993.