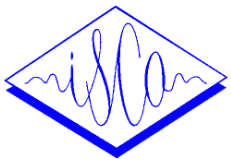

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

A CONCEPT FOR A COCKTAIL-PARTY-PROCESSOR

Markus Bodden

Lehrstuhl für allgemeine Elektrotechnik und Akustik (Prof. Dr.-Ing. J. Blauert),
Ruhr-Universität Bochum, FRG

0. ABSTRACT

This paper proposes a speech enhancement method that has no restrictions with regard to sound-field conditions or signal characteristics. It is specially designed to model the Cocktail-Party-Effect, that is, to suppress interfering speech signals, which is the most challenging problem concerning speech enhancement. This important aim is achieved by involving the knowledge of binaural signal processing of the human auditory system. Combining tools that simulate binaural signal processing with a powerful, adapted noise cancelling algorithm leads to a new concept for a unique system. The individual steps of the processing scheme are described and promising, preliminary results are presented.

1. INTRODUCTION

The human auditory system is characterized by features which are extremely important for communication situations: it is able to do both, noise cancellation and dereverberation. Especially the ability to provide speech intelligibility in cases where concurrent speech signals are present is of great interest.There are different fields where technical equivalents for these abilities could be utilized. Let us first consider hearing-impaired persons who have to come to terms with reduced intelligibility even using conventional hearing aids. Also the growing importance of speech recognition technology aiming at the realization of speech recognizers that work in real, noisy environments opens a new field and need for applications of speech enhancement technology.

Most of the well-known noise cancellation methods are based on pure signal-theoretical approaches (*Lim*, [4]). Their general applications are limited by restrictions to special, well-defined sound field conditions (e.g., stationary noise sources) or unpracticable, big arrangements of microphone arrays. In particular these tools offer poor results in cases where speech is the interfering signal.

The aim must be to evaluate a speech enhancement method that has no principle restrictions with regard to sound field conditions and that enables us to increase speech intelligibility for all kinds of interfering noise.

2. THE CONCEPT

By means of binaural processing the human auditory system is able to extract signals from spatially separated sound sources. Therefore it seems to be a powerful tool for speech enhancement including the ability to deal with additional interfering speech signals. The most obvious method to realize a Cocktail-Party-Processor will consequently be to profit from the knowledge concerning the signal processing in the human auditory system. However, approaches to simulate the Cocktail-Party-Effect in all details suffer from the lack of knowledge concerning processing of the higher stages of the auditory system. Nevertheless, the state-of-the-art knowledge about binaural processing can be used as a basis for a speech enhancement system. By means of combining them with an adapted powerful signal-theoretical method it is possible to set up a complete system which is already quite close to the human ideal.

With respect to the possibility of modelling the knowledge concerning the processing of the human auditory system can be summarized as follows:

- the influence of the outer-ear transfer functions are coded in the ear-signals and can thereby be included by dummy-head recording.
- the middle-ear can easily be modelled (or even neglected for this application).
- the processing of the cochlea can be modelled in a detailed (*Michel*, [5]), or in a very simplified, but satisfactory manner.
- spatial informations are used for processing (*Blauert*, [1]).
- further processing is essentially vague. Research points to aspects like attention (we 'concentrate' on one speaker) or morphology (we try to group parts of the excitation-patterns to matching events, e.g., when rising slopes occur simultaneously in several frequency bands) play an important role.

The overall structure of the resulting system is shown in *Fig. 1*, the individual processing steps are described in the following chapters.

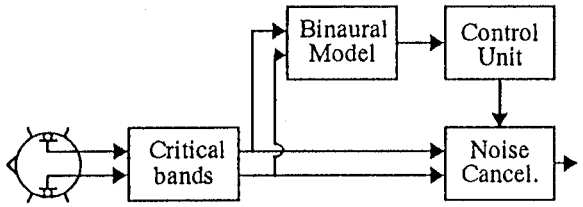

Fig. 1: Structure of the Cocktail-Party-Processor

2.1 The Binaural Model

The directional information is coded in the signals arriving at the eardrums. This information can be extracted by using a binaural model. *Lindemann* [3] developed a model on the basis of interaural cross-correlation-functions extended by a mechanism of contralateral inhibition. The model was improved and an algorithm to adapt to outer-ear-transfer functions was added by *Gaik* [2], resulting in a model that is able to process head-related signals. The model calculates time-, frequency- and direction-dependent excitation patterns passed to the control unit. An example for these patterns is shown in *Fig. 2* (1. speaker 45 degrees, 2. speaker -45 degrees).

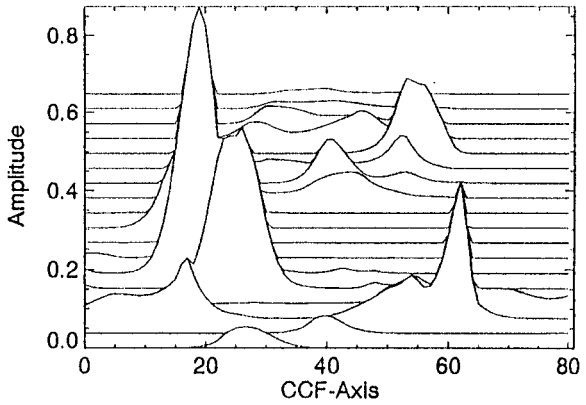

Fig. 2: Binaural excitation pattern in frequency bands 2 - 19, 2 speakers

2.2 The Control Unit

This unit extracts parameters from the excitation patterns of the model to control a powerful noise cancelling algorithm that is more signal-theoretical based, but adapted to auditory processing. The lack of knowledge in respect to the higher stages of the human auditory system excludes a further processing strictly according to the human ideal.

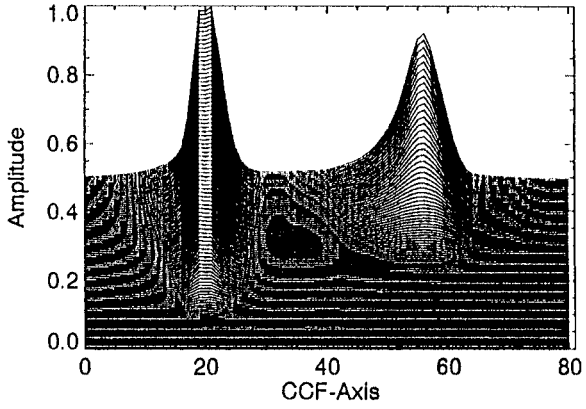

Fig. 3: Two-dimensionally averaged excitation patterns (first 100 ms)

The control unit

- estimates the number of sound sources and their lateralization. The patterns are averaged over two dimensions, frequency and time (time-constant 10-1000

ms). The number of maxima yields the number of sound sources and their interaural delays correspond to their lateralization. *Fig. 3* shows the averaged excitation pattern corresponding to the situation of *Fig. 2*.

- evaluates properties of the signals corresponding to the different incident angles. The knowledge of the lateralization of the sound sources enables the system to estimate the energies of the signals in each frequency band selectively for each detected direction.

Fig. 4 shows a survey of the processing steps of the control unit.

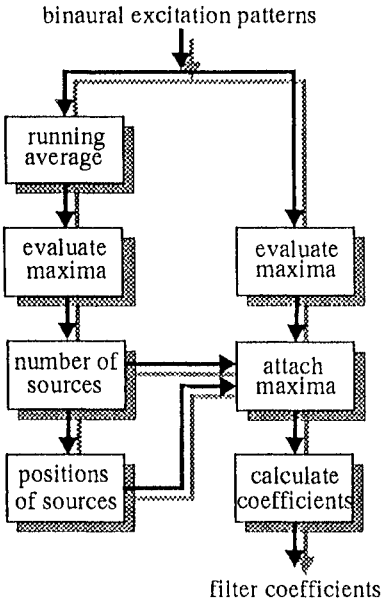

Fig. 4: Analysis of the binaural excitation patterns

2.3 The Noise Cancelling Algorithm

The noise cancelling algorithm consists of a modified optimal filter (Wiener filter) which is adapted to auditory processing. The theory of Wiener filtering describes the transfer-function of a predictionary filter to get an optimal estimate $\hat{S}$ for the desired signal S from the disturbed signal X = S + N:

$$H(f) = \frac{C_{SS}(f)}{C_{SS}(f) + C_{NN}(f)} \qquad (1)$$

with $C_{SS}(f)$:Cross power density spectrum of S
$C_{NN}(f)$:Cross power density spectrum of N

Considering the information available in this system an important simplification has to be introduced to solve *eq (1)*: we neglect the frequency-dependance of the transfer function for bandwidths corresponding to critical bands. This means that the transfer function H(f) can be described as a set of weighting factors g_i. This seems to be a pure theoretical approach, but there exists a correspondance to the processing of the human auditory system: it is barely able to separate signals within one frequency-band. The factors g_i can then be calculated from the rms-values of the signals :

$$g_i = \frac{E_S^2}{E_S^2 + E_N^2} \qquad (2)$$

with E_S : RMS value of S
E_N : RMS value of N

The time-variance of the signals is taken into account by updating these factors g in time intervals of the length T. A value of about 10 ms has proven to be a reasonable length. The flow chart of the algorithm is depicted in *Fig. 5*. Processing effort has been reduced to one multiplication per frequency band.

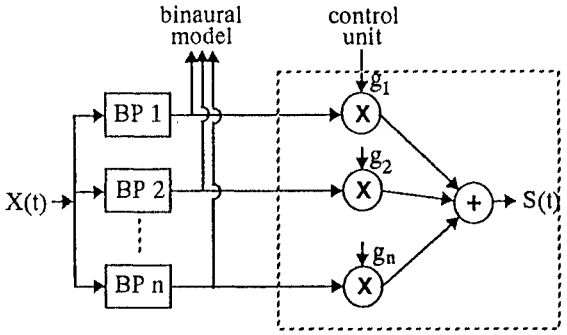

Fig. 5: Flow chart of the noise-cancellation-algorithm

3. PRELIMINARY RESULTS

In a first the step the noise cancelling algorithm has separetly been tested to verify whether the simplified processing provides satisfactory results. Therefore two known speech signals have monaurally been mixed and the optimal factors g_i have been calculated from these signals. *Fig. 6* shows the improvement of the signal-to-noise ratio.

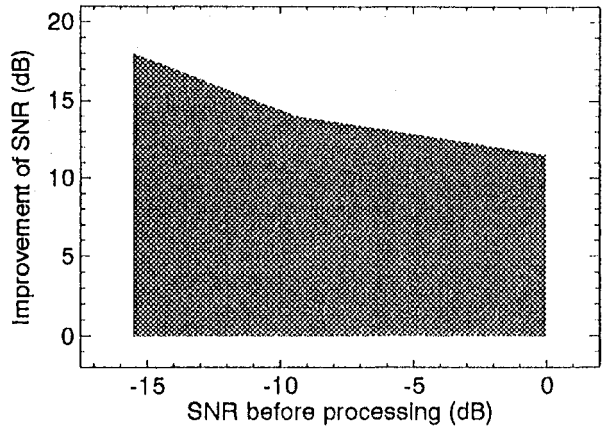

Fig. 6: Improvement of SNR for optimal weighting of critical bands

The results show an improvement in signal-to-noise ratio ranging from 11 to 18 dB for SNR's of 0 to -16 dB before processing. In order to avoid time-consuming intelligibility tests the Articulation-Index (AI) presented by *Pavlovic* [6] was chosen to judge improvements in intelligibility. Results are shown in *Fig. 7*. Articulation Index increased about an absolut value of 0.5.

The total system still is in the stage of developement, so that only preliminary results can be described. The independant components of the system have been tested for a two-speaker-situation (*Fig. 2, 3*), and the results indicate the ability to separate the speech signals.

4. SUMMARY

The knowledge of binaural processing offers a new promising approach to resolve the problem of evaluating efficient speech enhancement methods. Following the processing of the human auditory system means a considerable step forward. Collecting and adapting knowledge of hearing research and signal theory leads to a unified system which models the cocktail-party-effect. Further work has to be done to refine several aspects of the system, but the overall structure has proven to offer an appropriate and satisfactory solution for the speech enhancement problem.

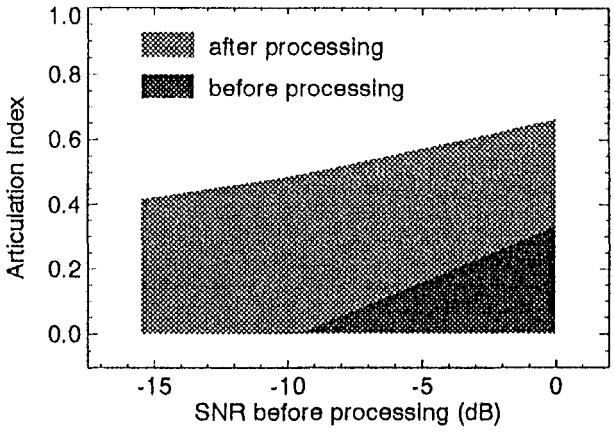

Fig. 7: Improvements of Articulation-Index for optimal weighting of critical bands

5. LITERATURE

[1] *Blauert, J. (1984)*, Spatial Hearing, MIT-Press, Cambridge, Mass.

[2] *Gaik, W. (1990)*,"Untersuchungen zur binauralen Verarbeitung kopfbezogener Signale", Fortschr.-Ber. VDI Reihe 17 Nr. 63. Düsseldorf: VDI-Verlag

[3] *Lindemann, W. (1983)*, "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization of stationary signals", JASA 80, 1608-1622.

[4] *Lim, J.S. (1983)*, "Speech Enhancement", Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[5] *Michel, D. (1988)*, "A model for peripheral auditory preprocessing", in: Cochlear Mechanics, Structure, Functions and Models, Plenum Press, New York, 425-436.

[6] *Pavlovic, C.V. (1987)*, "Derivation of primary parameters and procedures for use in speech intelligibility predictors", JASA 82(2), 413-422.