Signal Analysis by using Adaptive Filterbanks in Cochlear Implants

Amparo Albalate*, Waldo Nogueira§, Bernd Edler§ and Andreas Büchner‡
* Department of Information Technology, University of Ulm, 89081 Ulm, Germany
Email: amparo.albalate@uni-ulm.de
§ Information Technology Laboratory, University of Hannover, 30167 Hannover, Germany
Email: {nogueira,edler}@tnt.uni-hannover.de
‡ Medical University of Hannover, 30625 Hannover, Germany, Email: buechner@hoerzentrum-hannover.de

***Abstract*— Current speech processing in cochlear implants use a filterbank to analyse audio signals into several frequency bands, each associated with one electrode. Because the processing is performed on input signal blocks of fixed sizes, the filterbank provides a unique time-frequency resolution to represent the various signal features. However, different components of audio signals may require different time-frequency resolutions for an accurate representation and perception. In this paper we investigate the influence on speech intelligibility in cochlear implants users when filterbanks with different time-frequency resolutions are used. In order to represent all signal features accurately, an *adaptive filterbank* has been developed that accepts input blocks of different sizes. The different resolutions required are achieved by adequately switching between block sizes depending on the input signal characteristics. The filterbank was incorporated into the commercial Advanced Combinational Encoder (ACE) and acutely tested on six cochlear implant recipients.**

I. Introduction

Cochlear implants (CI) are widely accepted as prosthetic devices that improve the hearing ability of people with profound hearing loss. In essence, the implant electronics consist of five main units: a microphone, a transmitter, a receiver, a speech processor and an electrode array surgically placed into the cochlea. Among them, the speech processor is responsible for analysing input audio signals and delivering the most appropriate patterns to the electrodes.

Knowledge of the mechanisms used by CI recipients to extract acoustical information is of great relevance for the adequate coding of speech signals. Two basic cues are responsible for providing pitch perception in cochlear implants: place pitch and temporal pitch [1], [2].

Place pitch is extracted from the particular section in the cochlea that is being stimulated in each cycle. This mechanism is exploited in current multichannel strategies by means of a filterbank used to analyse input signals into several frequency bands or channels. Each channel is then mapped to a corresponding electrode or *place* in the cochlea. On the other hand, *temporal pitch* is related to the temporal fluctuations in the envelopes of each spectral band.

Perception of place pitch and temporal pitch by CI users is constrained by the limited number of electrodes and the limited Channel Stimulation Rate (CSR). Such limitations are most effectively addressed by "NofM" strategies, such as the Advanced Combinational Encoder (ACE) or the Psychoacoustical ACE [3]. Here the signals are decomposed into M channels, but only the N most meaningful bands are selected for each cycle stimulation. Thereby the CSR can be increased without significant loss of spectral information.

Typical reachable CSRs are in the order of 0.5 ms, which should be appropriate to accurately represent all signal features. However, block lengths used to analyse signals are substantially larger. Thus the filterbank time resolution may not be sufficient to represent short duration components, such as the attack of plosive phonemes in speech [4]. It has already been shown in [5], [6] that the identification of plosive consonants, which are associated with most of the consonant confusion errors, can be improved by emphasising short-duration cues at the consonant onset.

Higher time resolution would also be desirable to improve the perception of the fundamental frequency (F0) in speech. Although the place pitch mechanism is considered to be the predominant cue for frequency discrimination in tonal components, it may not hold for the whole range of F0 values. This is due to the relatively broad bands used in the signal analysis and the rejection of lowest frequencies for stimulation. Consequently, in order to emphasize the periodical structure of tonal components through temporal pitch extraction, an accurate coding of F0 time patterns is required [7].

Given the above statements, the approach described in this paper is intended to enhance temporal pitch extraction of the fundamental frequency and short-duration cues in speech (also referred to as *transients* in the following). Thereby we presume that better speech intelligibility can be achieved in cochlear implants. Besides, while higher time resolution is required in such cases, frequency resolution should be also preserved for the rest of components. In order to adapt the output resolution to the different signal features, we propose an *adaptive filterbank* that has been incorporated for testing into the Advanced Combinational Encoder.

The paper is organised as follows: in Section 2 an overview of the ACE strategy is presented. Section 3 describes the new adaptive filterbank in detail. Intelligibility test procedures and results are provided in Sections 4 and 5, respectively. Finally, we draw conclusions and show the future directions of the proposed algorithm in Section 6.

II. The Advanced Combinational Encoder (ACE)

The Advanced Combinational Encoder used by the Nucleus Implant of Cochlear corporation can be observed in Figure 1.

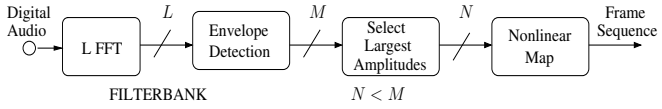


Fig. 1. ACE block diagram

The input signal, sampled at $f_s = 16kHz$, is analysed on input windowed frames of 8 ms or a number L of 128 samples. The spectral decomposition of each frame is obtained through an M-band filterbank, where M is the number of implanted electrodes (typically 20 or 22). In the following, 20 bands are considered because only 20 electrodes were configured on the implants of the patients tested. The filterbank is implemented by means of an L-point FFT that provides $L = 128$ spectral coefficients. Due to the FFT symmetry property for real input samples, the second half of the spectrum is discarded without loss of information. The first 64 bins are then used by the envelope detector, which rearranges them in a non-uniform way closer to the critical band partition by the human cochlea [9].

In "NofM" strategies, the N envelopes with the most meaningful information ($N < M$) are extracted for each temporal stimulation according to certain criteria. In the case of ACE, the N spectral maxima become the selected bands for stimulation.

Finally, a non-linear mapping function is applied to each envelope in order to map acoustical amplitudes into electrical amplitudes and fit the dynamic range of the patient using the implant. The resulting amplitudes, modulated through electrical pulses, are then delivered to the corresponding electrodes.

III. Design of the ACE Strategy with Adaptive Filterbanks

The new strategy was designed at the Information Technology Laboratory at the University of Hannover [8]. The algorithm consists of migrating the filterbank employed in the Advanced Combinational Encoder towards an adaptive filterbank (Figure 2). Based on a previous analysis of the actual signal frame, the new filterbank is able to adapt the output resolution for an accurate representation of all signal components.

As described in Section 2, the input signal is fragmented in blocks of 8 ms or 128 samples, $x[n]$. These blocks are referred to as the *long frames*. However, in contrast to the ACE strategy, the long frames are now subdivided in three shorter 4-ms frames with a 50% overlapping ratio. These shorter blocks are referred to as the *past*, *current* and *future sub-frames*, $x_p[n]$, $x_c[n]$ and $x_f[n]$, respectively, since they represent three different time instants of the long frame. The long frame and short sub-frames are next windowed through a 128-hanning window and 64-hanning windows, respectively.

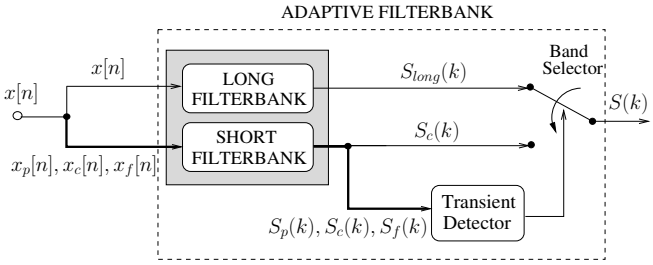


Fig. 2. Block diagram of the adaptive filterbank

Afterwards, the spectral decomposition of the long and short frames is calculated in parallel by the ACE filterbank (hereinafter the *long filterbank*), and an additional *short filterbank* that accepts the 4 ms windowed subframes. An analysis of the envelope temporal evolution for rapid changes is performed in each k_{th} band by the *transient detector*, enabling the selection of a proper resolution. Because the long filterbank provides better frequency resolution, it is normally selected. In the case of transient detection, higher time resolution is demanded and the short filterbank is selected instead. This is done independently for each k_{th} spectral band, in such a way that the final spectrum S is obtained as a combination of bands from either filterbank.

A. Long and short filterbanks

In the new approach, the long filterbank corresponds identically to the ACE filterbank structure. A 128-point FFT computes the spectral decomposition of long frames into 128 symmetrical spectral bins, $X(l)$. The envelope detector maps the first 64 bins into 20 bands, according to equation (1).

$$S_{long}(k) = g_L(k) \cdot \sqrt{\sum_{l_i(k)}^{l_i(k)+N_b(k)} E(l)} \qquad (1)$$
$$k = 1, \cdots, 20 \qquad l = 1, \cdots, 64$$

Where larger values of the band index, k, denote increasing frequencies. $S_{long}(k)$ denotes the long frames' k_{th} band envelope, $E(l)$ refers to the energy of the l_{th} FFT output bin, ($E(l) = X(l) \cdot X(l)^*$), $N_b(k)$ is the number of consecutive FFT bins that compose the k_{th} band, and $l_i(k)$ denotes the index of the initial FFT bin in that band. Finally, the weights $g_L(k)$ are gain factors needed to equalize different effects introduced by the filterbank. For further details on numerical values of the parameters $N_b(k)$, $l_i(k)$ and $g_L(k)$, see [3].

Likewise, the short filterbank is used in parallel to analyse the short subframes. A 64-point FFT computes the frames' spectral decomposition into 64 symmetrical coefficients. In order to apply the same mapping of bins into bands like in the long filterbank, the spectrum of bins must be first linearly upsampled by a factor of 2. The output envelopes, $S_p(k)$, $S_c(k)$ and $S_f(k)$, are obtained for the *past*, *current* and *future* sub-frames respectively, in a similar way as described for the long filterbank:

$$S_c(k) = \frac{1}{\sqrt{2}} \cdot g_L(k) \cdot \sqrt{\sum_{l'_i(k)}^{l'_i(k)+N_b(k)} E_c(l')} \qquad (2)$$
$$k = 1, \cdots, 20 \qquad l' = 1, \cdots, 64$$

Equation (2) shows the particular envelopes calculation for the current subframe. $E_c(l')$ refers to the l' interpolated bin energy. The different index notation l' indicates that the coefficients are now considered after interpolating the original FFT bins. The envelopes $S_p(k)$ and $S_f(k)$ can be identically obtained by adequately replacing $E_c(l')$ in (2) with the corresponding interpolated-bins energies for the past and future frames, $E_p(l')$ and $E_f(l')$, respectively.

An additional gain factor of $1/\sqrt{2}$ has been introduced to regulate different envelopes' amplitudes in short and long filterbanks caused by the different number of FFT points while placing some emphasis on transients. Consider, for example, the two limit cases of an ideal stationary component (a signal with constant amplitude in time domain) and an ideal transient component (Dirac's delta). In the first case, the FFT output only contains a single non-zero bin (DC coefficient). The bin' amplitude computed using (normalized) 64 and 128-point FFTs remains inalterable. In the second case, the bin' amplitudes are doubled when halving the number L of frame samples. Thus the mentioned gain factor places different emphasis on components analysed by the short filterbank, which increases with the higher transient behaviour of such components.

B. Transient detector

This block analyses the temporal envelope evolution in each spectral band for rapid changes that may correspond to transient components. The signal variation is first estimated according to equation (3).

$$Sv(k) = \frac{(S_f(k))^\alpha - (S_c(k))^\alpha}{(S_c(k))^\alpha - (S_p(k))^\alpha} \qquad (3)$$

The factor α has been introduced to reject possible high fluctuations of very low energies corresponding to noise backgrounds. A value of 4 selected for this parameter proved optimal for the transient detector performance.

A transient is detected in the k_{th} band if the signal variation exceeds a predetermined threshold Sv_{thresh}, in which case the transient onset begins in the *future sub-frame*. The chosen threshold level was low enough so that signal fluctuations related to fundamental period peaks in voiced sounds could also be captured and emphasised.

C. Band Selector

The final output bands $S(k)$ are then obtained by switching between corresponding envelopes from the long or short filterbanks according to expression (4).

$$S(k) = \begin{cases} S_c(k), & Sv(k) > Sv_{thresh} \\ S_{long}(k), & Sv(k) \leq Sv_{thresh} \end{cases} \qquad (4)$$

Although the transient onset is located in the future subframe, in practice the switching is performed towards $S_c(k)$ during three consecutive frames. This is done to prevent preceding and succeeding long windows from capturing the transient.

An example of the band selection is illustrated in Figure 3. An utterance */aka/* produced by a male speaker with a fundamental frequency, $F0 \sim 120Hz$ (Figure 3a) has been processed using an adaptive filterbank. Figures 3b and 3c plot the envelope temporal evolution in two different frequency bands at the output of the long, short and adaptive filterbanks, $S_{long}(k)$, $S_c(k)$ and $S(k)$. Figure 3b shows band envelopes associated to electrode number 1 ($k = 1$) in several frames of the second */a/* phoneme. Coinciding with the fundamental period in the vowel, transients are detected and hence the short filterbank is selected ($S(1) = S_c(1)$). Elsewhere, the switching is performed towards the long filterbank output ($S(1) = S_{long}(1)$). Figure 3c represents envelopes in the $k = 12$ band for some frames of the */k/* consonant. Transients are found exactly at the consonant onset (frame 161). Again, the short filterbank output is selected and higher time resolution and emphasis can be observed.

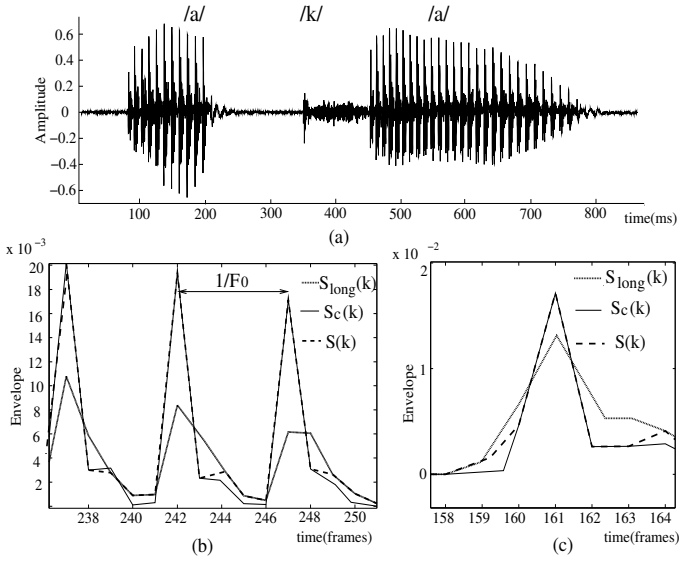


Fig. 3. Example of output bands selection in the adaptive filterbank: a) input utterance /aka/ produced by a male speaker (in quiet). b) Envelopes obtained at the output of the long, short and adaptive filterbanks in several frames in the second /a/ phoneme, and mapped to electrode Number 1 in the electrode array (central frequency, $f_c = 250Hz$); b) Envelopes in several frames located at the /k/ phoneme and mapped to electrode Number 12 ($f_c = 2126Hz$).

IV. INTELLIGIBILITY TESTS

The new filterbank has been incorporated into a research ACE strategy made available by Cochlear Corporation. The interface used is termed NIC (Nucleus Implant Communicator), specially designed for research purposes to allow the researcher to communicate with the inside-the-ear part of the Nucleus implant and send any stimulus pattern to the electrodes.

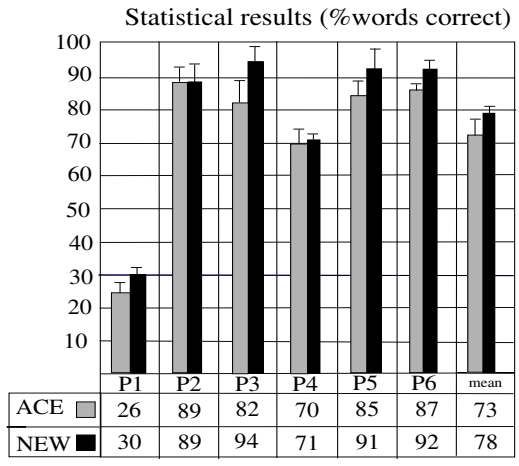
First, the NIC processes the audio signals on a personal computer (PC). A specially initialised clinical processor used for the fitting of patients in routine clinical practice serves as a transmitter of the instructions from the PC to the subject's implant so that the functionality of the processor in the implant is bypassed.

The subjects were six postlingually deafened adults, patients of the Medical University of Hannover, who wore a Nucleus 24 implant with the ACE strategy on a daily basis. The test used for evaluation was the HSM (Hochmair, Schultz, Moyer) sentence test [10], composed of 30 lists with 20 everyday sentences and a total of 106 words each. Scoring was based on correct words.

The test material was processed by the ACE strategy and the new approach with an adaptive filterbank. Signals had been previously processed through a pre-emphasis filter that models the frequency response of the microphone used in commercial implants. In training sessions, lists were presented to the subjects in quiet conditions, so that patients could familiarize with the test material. For test sessions, the material was processed in noise with a Signal-to-Noise Ratio (SNR) of 15 dB. Patients had to repeat the lists of words they heard, randomly selected from the set of HSM lists, without knowing which strategy was being selected at each moment. Two different lists were tested on each patient for each strategy. Scores could be assigned because the researcher had the corresponding word transcriptions for the lists selected.

V. Results

Figure 4 shows the statistical speech intelligibility performance results of the tests on patients.



	P1	P2	P3	P4	P5	P6	mean
ACE	26	89	82	70	85	87	73
NEW	30	89	94	71	91	92	78

Fig. 4. Results in patients. Average and standard deviation over the total number of lists tested on each patient in noisy conditions (SNR=15 dB). Last bars indicate the group average over all tested patients.

Speech intelligibility performance obtained by the new strategy using an adaptive filterbank was similar to the Advanced Combinational Encoder for all subjects tested. Furthermore, a group mean improvement of 5% (non significant) in speech intelligibility is shown in the new strategy. However, a trend can be observed as to some improvements in noisy conditions through the use of adaptive filterbanks. The observed results can be attributed to the better resolution and emphasis provided for the perception of short-time components and F0 time patterns, consistently with the initial motivation for the new algorithm presented.

VI. Conclusions and Future directions

In this paper we have investigated the influence of a more accurate time representation and emphasis of short-time and F0-tonal components of audio signals on the overall speech performance achieved by a cochlear implant. With this aim the commercial Advanced Combinational Encoder (ACE) has been compared to a new approach incorporating an adaptive filterbank which is able to alternate different resolutions to represent the various signal features.

The outcome of intelligibility tests conducted on six cochlear implant recipients showed a 5% (non significant) group mean improvement on speech performance when using the adaptive filterbank. Such results are encouraging, provided that the present study was an initial pilot test on a few number of patients whose daily strategy was the ACE.

Further research based on the presented work could be directed towards the spectral enhancement of the components detected as transients. Several means to provide higher frequency resolution in those cases can be analysed. In addition, spectral contrast can be also enhanced with small modifications of the current algorithm to reject transients detected in bands that do not correspond to spectral peaks. Thereby potential improvements in the identification of wowels and plosive consonants may be achieved [6], [11], thus contributing to improving speech intelligibility in cochlear implant users.

References

[1] P. C. Loizou, *Signal Processing Techniques for Cochlear Implants*, IEEE Engineering in Medicine and Biology magazine, pp. 34-46, 2006.

[2] C. M. Mackay, *Place, temporal cues in pitch perception: are they truly independent?*, Acoustic Research Letters Online, Acoustical Society of America, Vol. 1, pp.25-30, July 2000.

[3] W. Nogueira, et al., *A Psychoacoustic NofM type speech coding strategy for cochlear implants*, Eurasip Journal on Applied Signal Processing, Special Issue on DSP in Hearing Aids and Cochlear Implants, 2005.

[4] W. Nogueira, et al., *Wavelet Packet Filterbank for Speech Processing Strategies in Cochlear Implants*, 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, May 2006, Tolouse, France.

[5] A. E. Vandali, *Emphasis of short-duration acoustic cues for cochlear implant users*, Journal of the Acoustic Society of America, Vol.109, pp. 2049-61, May 2001.

[6] M. Dorman, P. C. Loizou, *Improving consonant intelligibility for Ineraid patients fit with Continuous Interleaved Samples (CIS) processors by enhancing contrast among channel outputs*, Journal of the Acoustic Society of America, 1996.

[7] W. Nogueira, et al., *Fundamental frequency coding in NofM strategies for cochlear implants*, Preprint 6615, 118th AES Convention, Barcelona, Spain, 2005.

[8] A. Albalate, *Signal Analysis by using Adaptive Filterbanks in Cochlear Implants*, Diplomarbeit, Universität Hannover, 2005.

[9] E. Zwicker, H.Fastl, *Psychoacoustics. Facts and Models*, Springer,1999.

[10] I. Hochmair-Desoyer, et al., *The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users*, J.Otol, Vol.18, Suppl.6, pp. 83, November 1997.

[11] P. C. Loizou, O.Poroy, *Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners*, Journal of the Acoustical Society of America, Vol.110, pp. 1619-27, 2001.