A Psychoacoustic "NofM"-Type Speech Coding Strategy for Cochlear Implants

Waldo Nogueira

Laboratorium für Informationstechnologie, Universität Hannover, Schneiderberg 32, 30167 Hannover, Germany
Email: nogueira@tnt.uni-hannover.de

Andreas Büchner

Department of Otolaryngology, Medical University Hanover, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany
Email: buechner@hoerzentrum-hannover.de

Thomas Lenarz

Department of Otolaryngology, Medical University Hanover, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany
Email: lenarz@hno.mh-hannover.de

Bernd Edler

Laboratorium für Informationstechnologie, Universität Hannover, Schneiderberg 32, 30167 Hannover, Germany
Email: edler@tnt.uni-hannover.de

We describe a new signal processing technique for cochlear implants using a psychoacoustic-masking model. The technique is based on the principle of a so-called "NofM" strategy. These strategies stimulate fewer channels (N) per cycle than active electrodes (NofM; $N < M$). In "NofM" strategies such as ACE or SPEAK, only the N channels with higher amplitudes are stimulated. The new strategy is based on the ACE strategy but uses a psychoacoustic-masking model in order to determine the essential components of any given audio signal. This new strategy was tested on device users in an acute study, with either 4 or 8 channels stimulated per cycle. For the first condition (4 channels), the mean improvement over the ACE strategy was 17%. For the second condition (8 channels), no significant difference was found between the two strategies.

Keywords and phrases: cochlear implant, NofM, ACE, speech coding, psychoacoustic model, masking.

1. INTRODUCTION

Cochlear implants are widely accepted as the most effective means of improving the auditory receptive abilities of people with profound hearing loss. Generally, these devices consist of a microphone, a speech processor, a transmitter, a receiver, and an electrode array which is positioned inside the cochlea. The speech processor is responsible for decomposing the input audio signal into different frequency bands or channels and delivering the most appropriate stimulation pattern to the electrodes. When signal processing strategies like continuous interleaved sampling (CIS) [1] or advanced combinational encoder (ACE) [2, 3, 4] are used, electrodes near the base of the cochlea represent high-frequency information, whereas those near to the apex transmit low-frequency information. A more detailed description of the process by which the audio signal is converted into electrical stimuli is given in [5].

Speech coding strategies play an extremely important role in maximizing the user's overall communicative potential, and different speech processing strategies have been developed over the past two decades to mimic firing patterns inside the cochlea as naturally as possible [5]. "NofM" strategies such as ACE or spectral peak (SPEAK) [4] were developed in the 1990s. These strategies separate speech signals into M subbands and derive envelope information from each band signal. N bands with the largest amplitude are then selected for stimulation (N out of M). The basic aim here is to increase the temporal resolution by neglecting the less significant spectral components and to concentrate on the more important features. These strategies have demonstrated either a significant improvement or at least
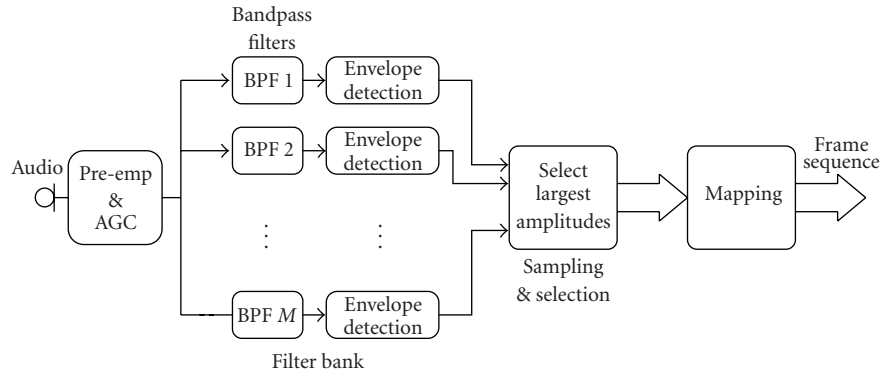
FIGURE 1: Block diagram illustrating ACE.

user preference over conventional CIS-like strategies [6, 7, 8]. However, speech recognition for cochlear implant recipients in noisy conditions—and, for some individuals, even in quiet—remains a challenge [9, 10]. To further improve speech perception in cochlear implant users, the authors decided to modify the channel selection algorithm of the ACE speech coding strategy.

This work therefore describes a new method for selecting the N bands used in "NofM" strategies. As outlined above, conventional "NofM" strategies select the N bands with the largest amplitudes from the M filter outputs of the filter bank. In the new scheme the N bands are chosen using a psychoacoustic-masking model. The basic structure of this strategy is based on the ACE strategy but incorporating the above-mentioned psychoacoustic model. This new strategy has been named the psychoacoustic advanced combination encoder (PACE). Psychoacoustic-masking models are derived from psychoacoustic measurements conducted on normal-hearing persons [11, 12, 13] and can be used to extract the most meaningful components of any given audio signal [14, 15]. Those techniques are widely used in common hi-fi data reduction algorithms, where data streams have to be reduced owing to bandwidth or capacity limitations. Well-known examples of these techniques are the adaptive transform acoustic coding (ATRAC) [16] coding system for mini-disc recorders and the MP3 [17, 18] compression algorithm for transferring music via the Internet. These algorithms are able to reduce the data to one-tenth of its original volume with no noticeable loss of sound quality.

"NofM" speech coding strategies have some similarities to the above-mentioned hi-fi data reduction or compression algorithms in that these strategies also compress the audio signals by selecting only a subset of the frequency bands. The aim in introducing a psychoacoustic model for channel selection was to achieve more natural sound reproduction in cochlear implant users.

Standardized speech intelligibility tests were conducted using both the ACE and the new PACE strategy, and the scores compared in order to test whether the use of a psychoacoustic model in the field of cochlear implant speech coding can indeed yield improved speech understanding in the users of these devices.

The paper is organized as follows. In Section 2, a review of the ACE strategy is presented. Furthermore, the psychoacoustic model and how it has been incorporated into an "NofM" strategy is described. Section 3 gives the results of the speech understanding tests with cochlear implant users and finally, in Sections 4 and 5, a discussion and the conclusions are presented respectively.

2. METHODS

2.1. Review of the ACE strategy

Several speech processing strategies have been developed over the years. These strategies can be classified into two groups: those based on feature extraction of the speech signals and those based on waveform representation. The advanced combinational encoder (ACE) [2, 3] strategy used with the Nucleus implant is an "NofM"-type strategy belonging to the second group. The spectral peak (SPEAK) [4] strategy is identical in many aspects to the ACE strategy, but different in rate. Figure 1 shows the basic block diagram illustrating the ACE strategy.

The signal from the microphone is first pre-emphasized by a filter that amplifies the high-frequency components in particular. Adaptive-gain control (AGC) is then used to limit distortion of loud sounds by reducing the amplification at the right time.

Afterwards, the signal is digitized and sent through a filter bank. ACE does not explicitly define a certain filter bank approach. The frequency bounds of the filter bank are linearly spaced below 1000 Hz, and logarithmically spaced above 1000 Hz.

An estimation of the envelope is calculated for each spectral band of the audio signal. The envelopes are obtained by computing the magnitude of the complex output. Each band pass filter is allocated to one electrode and represents one channel. For each frame of the audio signal, N electrodes are stimulated sequentially and one cycle of stimulation is completed. The number of cycles/second thus determines the rate of stimulation on a single channel, also known as channel stimulation rate.
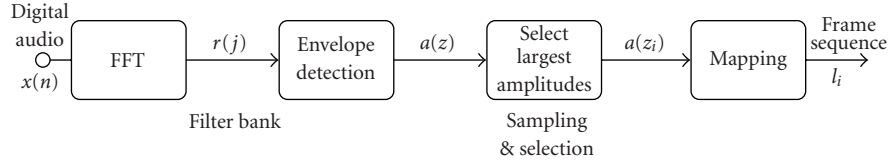
FIGURE 2: Block diagram illustrating research ACE.

TABLE 1: Number of FFT bins, center frequencies, and gains per filter band for $M = 22$.

Band number z	1	2	3	4	5	6	7	8	9	10	11
Number of bins	1	1	1	1	1	1	1	1	1	2	2
Center freqs. (Hz)	250	375	500	625	750	875	1000	1125	1250	1437	1687
Gains g_z	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.68	0.68

Band number z	12	13	14	15	16	17	18	19	20	21	22
Number of bins	2	2	3	3	4	4	5	5	6	7	8
Center freqs. (Hz)	1937	2187	2500	2875	3312	3812	4375	5000	5687	6500	7437
Gains g_z	0.68	0.68	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65	0.65

The bandwidth of a cochlear implant is limited by the number of channels (electrodes) and the overall stimulation rate. The channel stimulation rate represents the temporal resolution of the implant, while the total number of electrodes M represents the frequency resolution. However, only N out of M electrodes ($N < M$) are stimulated in each cycle, therefore a subset of filter bank output samples with the largest amplitude is selected. If N is decreased, the spectral representation of the audio signal becomes poorer, but the channel stimulation rate can be increased, giving a better temporal representation of the audio signal. Conversely, if the channel stimulation rate is decreased, N can be increased, giving a better spectral representation of the audio signal.

Finally, the last stage of the process maps the amplitudes to the corresponding electrodes, compressing the acoustic amplitudes into the subject's dynamic range between measured threshold and maximum comfortable loudness level for electrical stimulation.

2.2. Research ACE strategy used

A research ACE strategy [3] was made available by Cochlear Corporation for the purpose of deriving new speech coding strategies. However, the research ACE strategy is designed to process signals that are already digitized. For this reason, the pre-emphasis filter and adaptive-gain controls (AGC) incorporated at the analogue stage are not included in this set-up. Figure 2 shows a basic block diagram illustrating the strategy.

A digital signal sampled at 16 kHz is sent through a filter bank without either pre-amplification or adaptive-gain control. The filter bank is implemented with an FFT (fast Fourier transform). The block update rate of the FFT is adapted to the rate of stimulation on a channel (i.e., the total implant rate divided by the number of bands selected N). The FFT is performed on input blocks of 128 samples ($L = 128$) of

the previously windowed audio signal. The window used is a 128-point Hann window [19]

$$w(j) = 0.5\left(1.0 - \cos\left(\frac{2\pi j}{L}\right)\right), \quad j = 0, \dots, L - 1. \quad (1)$$

The linearly-spaced FFT bins are then combined by summing the powers to provide the required number of frequency bands M, thus obtaining the envelope in each spectral band $a(z)$ ($z = 1, \dots, M$). The real part of the jth FFT bin is denoted with $x(j)$, and the imaginary part $y(j)$. The power of the bin is

$$r^2(j) = x^2(j) + y^2(j), \quad j = 0, \dots, L - 1. \quad (2)$$

The power of the envelope of a filter band z is calculated as a weighted sum of the FFT bin powers

$$a^2(z) = \sum_{j=0}^{L/2} g_z(j) r^2(j), \quad z = 1, \dots, M, \quad (3)$$

where $g_z(j)$ are set to the gains g_z for a specific number of bins and otherwise zero. This mapping is specified by the number of bins, selected in ascending order starting at bin 2, and by the gains g_z as presented in Table 1 [3, 20].

The envelope of the filter band z is

$$a(z) = \sqrt{\sum_{j=0}^{L/2} g_z(j) r^2(j)}, \quad z = 1, \dots, M. \quad (4)$$

In the "sampling and selection" block, a subset of N ($N < M$) filter bank envelopes $a(z_i)$ with the largest amplitude are selected for stimulation.
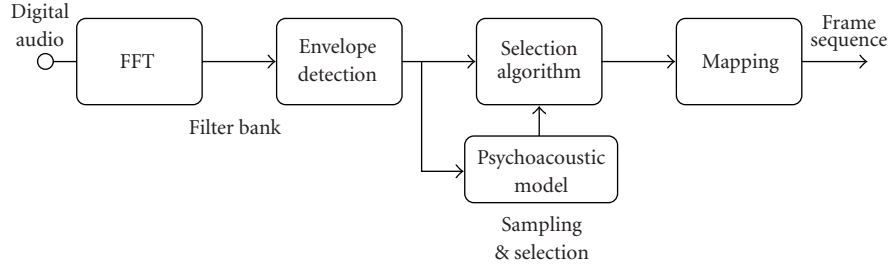
FIGURE 3: Block diagram illustrating an "NofM" strategy incorporating a psychoacoustic model for selecting the N bands. The strategy may be termed the psychoacoustic ACE strategy.

The "mapping" block, determines the current level from the envelope magnitude and the channel characteristics. This is done by using the loudness growth function (LGF) which is a logarithmically-shaped function that maps the acoustic envelope amplitude $a(z_i)$ to an electrical magnitude

$$p(z_i) = \begin{cases} \dfrac{\log\left(1 + \rho\left((a(z_i) - s)/(m - s)\right)\right)}{\log(1 + \rho)}, & s \leq a(z_i) \leq m, \\ 0, & a(z_i) < s, \\ 1, & a(z_i) \geq m. \end{cases} \quad (5)$$

The magnitude $p(z_i)$ is a fraction in the range 0 to 1 that represents the proportion of the output range (from the threshold T to the comfort level C). A description of the process by which the audio signal is converted into electrical stimuli is given in [21]. An input at the base-level s is mapped to an output at threshold level, and no output is produced for an input of lower amplitude. The parameter m is the input level at which the output saturates; inputs at this level or above result in stimuli at comfort level. If there are less than N envelopes above base level, they are mapped to the threshold level. The parameter ρ controls the steepness of the LGF, the selection of a suitable value for ρ is described in [20].

Finally, the channels z_i, are stimulated sequentially with a stimulation order from high-to-low frequencies (base-to-apex) with levels:

$$l_i = T + (C - T)p_i. \quad (6)$$

2.3. "NofM" strategy using a psychoacoustic model: the psychoacoustic ACE (PACE) strategy

Based on the general structure of the research ACE strategy (Figure 2) but incorporating a psychoacoustic model, a new approach was designed in order to select the N ($N < M$) bands in "NofM" strategies. A basic block diagram illustrating the proposed PACE strategy is presented in Figure 3.

Both the filter bank and the envelope detection process are identical to those in the research ACE strategy. A psychoacoustic-masking model—as opposed to a peak-picking algorithm—is then used to select the N bands. Consequently, the bands selected by this new approach are not necessarily those with the largest amplitudes (as is the case

in the ACE strategy) but the ones that are, in terms of hearing perception, most important to normal-hearing people. Afterwards, the bands selected are mapped to electrical impulses and sent to the electrode array following exactly the same process as in the research ACE strategy.

In the following paragraphs the psychoacoustic model and the selection algorithm will be explained.

2.3.1. Psychoacoustic model

There are different classes of psychoacoustic models, the one referred to in this manuscript being a psychoacoustic-masking model. Such models describe masking effects that take place in a healthy auditory system. Psychoacoustic models have been successfully used within the field of audio coding in order to reduce bandwidth requirements by removing the less perceptually important components of audio signals. Because "NofM" speech coding strategies only select certain spectral elements of the audio signals, it can be speculated that a psychoacoustic model may ensure more effective selection of the most relevant bands than is achieved by merely selecting the spectral maxima, as with the ACE strategy.

Psychoacoustic-masking models are based on numerous studies of human perception, including investigations on the absolute threshold of hearing and simultaneous masking. These effects have been studied by various authors [11, 12, 13, 22].

The absolute threshold of hearing is a function that gives the required sound pressure level (SPL) needed in order that a pure tone is audible in a noiseless environment. The effect of simultaneous masking occurs when one sound makes it difficult or impossible to perceive another sound of similar frequency.

A psychoacoustic model as described by Baumgarte in 1995 [15] was adapted to the features of the ACE strategy. The psychoacoustic model employed here is used to select the N most significant bands in each stimulation cycle. In the following sections we describe the steps (shown in Figure 4) that constitute the masking model. The masked threshold is calculated individually for each band selected. The overall masked threshold created by the different bands can then be approximated by nonlinear superposition of the particular masked thresholds. Figure 4 shows an example of the psychoacoustic model implemented operating on two selected bands.
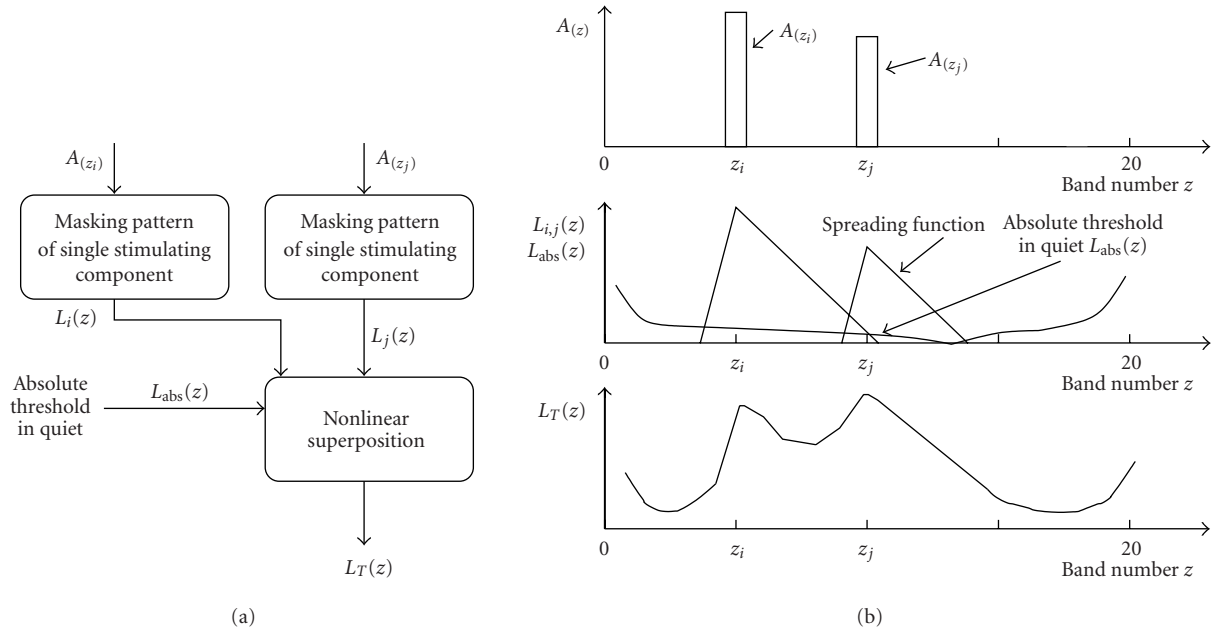
FIGURE 4: (a) Block diagram. The input comprises the envelope values of the bands chosen by the selection algorithm. The output is the overall masked threshold. (b) Associated levels over the frequency band number z.
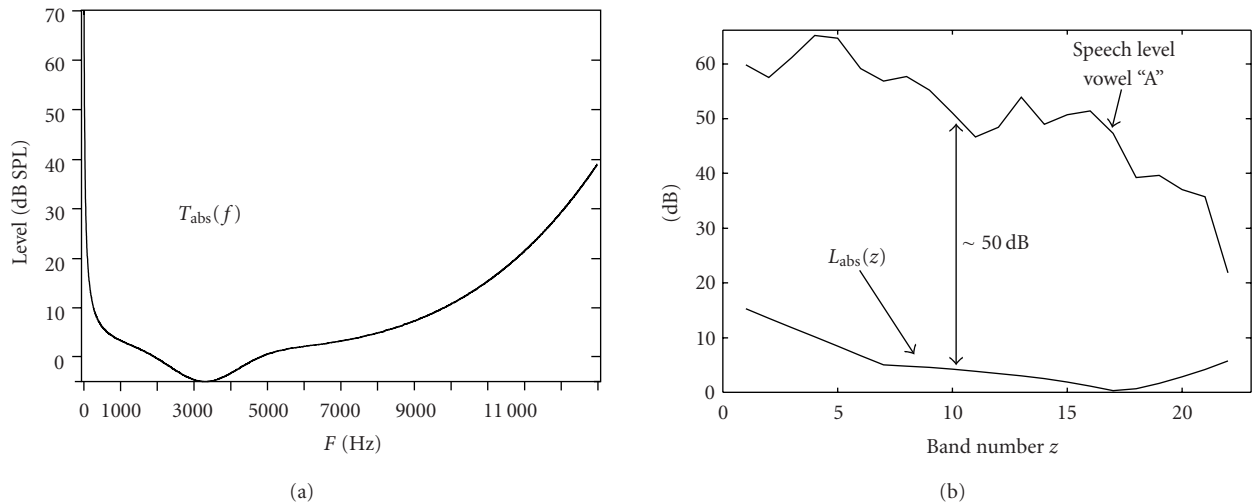

FIGURE 5: (a) Threshold in quiet over the frequency in Hz. (b) Threshold in quiet approximation over the band number z and spectral level when the vowel "A" is uttered.

2.3.1.1. Threshold in quiet

A typical absolute threshold expressed in terms of dB SPL is presented in Figure 5a [23].

The function $L_{abs(z)}$ representing the threshold in quiet in each frequency band z is obtained by choosing one representative value of the function presented in Figure 5a at the centre frequency of each frequency band (Table 1). However, as the authors have no a priori knowledge regarding playback levels (SPL) of the original audio signals, a reference had to be chosen for setting the level of the threshold in quiet. It is known that the threshold in quiet lies at around 50 dB below

"normal speech level" (i.e., between 200 Hz and 6 kHz [11]). The level of the function $L_{abs}(z)$ was therefore set at 50 dB below the level of the voiced parts from certain audio samples used as test material. Figure 5b presents the resulting $L_{abs}(z)$ and the spectral level obtained when a generic vowel "a" in the test material is uttered. The vowel "a" was stored in a "wav" file format coded with 16 bits per sample, and the standard deviation for the whole vowel was about 12 dB below the maximum possible output level. It is important to note that $T_{abs}(f)$ is expressed in terms of dB SPL and $L_{abs}(z)$ in dB (0 dB corresponds to the minimum value of the threshold in quiet mentioned before).
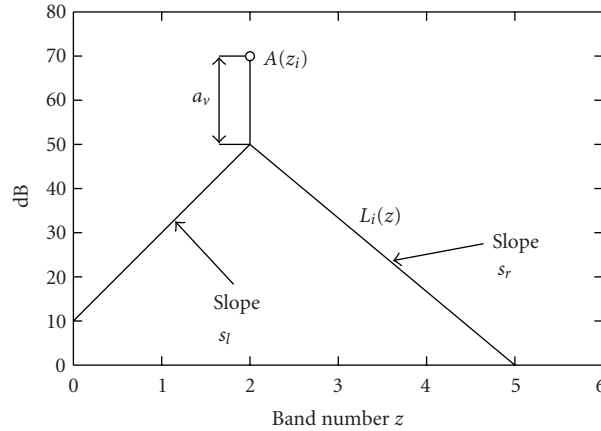
FIGURE 6: Spreading function $L_i(z)$ of one masker component $A(z_i)$ at the band z_i. The left and right slopes of the spreading function are indicated as s_l and s_r. The attenuation of the maximum relative to the masker level is denoted by a_v.

2.3.1.2. Masking pattern of single stimulating component

For each selected band, a function is calculated that models the masking effect of this band upon the others. This function familiar in the field of psychoacoustics as the so-called spreading function, expressed with the same dB units as in Figure 5b, is presented in Figure 6.

The spreading function is described by three parameters: attenuation, left slope, and right slope. The amplitude of the spreading function is defined using the attenuation parameter a_v. This parameter is defined as the difference between the amplitude of the selected band $A(z_i)$ and the maximum of the spreading function in dB units. The slopes s_l and s_r correspond to the left and right slopes, respectively, in the unit "dB/band." As presented in [15], the spreading function belonging to a band z_i with amplitude $A(z_i)$ in decibels is mathematically represented by $L_i(z)$:

$$L_i(z) = \begin{cases} A(z_i) - a_v - s_l \cdot (z_i - z), & z < z_i, \\ A(z_i) - a_v - s_r \cdot (z - z_i), & z \geq z_i, \end{cases} \quad (7)$$

where

- (i) z denotes the frequency band number at the output of the filter bank, $1 \leq z \leq M$,
- (ii) i denotes that the band selected is z_i (i.e., masker band).

In the model description of [15], z denoted the critical band rate [11, 24] or equivalently critical band number [12, 13]. Because the bandwidths of the frequency bands used in the filter bank in the ACE and PACE schemes are approximately equal to the critical bands, the frequency band number corresponds approximately to the critical band rate. Therefore, in the implementation of the masking model in the present study, it was opted to define the masking patterns as a function of the frequency band number instead of the critical band rate.

2.3.1.3. Nonlinear superposition

The sound intensities $I_{abs}(z)$ and $I_i(z)$ are calculated from the decibel levels by

$$\begin{aligned} I_{abs}(z) &= 10^{L_{abs}(z)/10}, \\ I_i(z) &= 10^{L_i(z)/10}. \end{aligned} \quad (8)$$

Threshold components should be combined in a way that reflects the characteristics of human auditory perception. Certain approaches have been based on linear addition of the threshold components [25]. However, further results proved that linear models fail in most cases where threshold components exhibit spectral overlapping [25, 26]. A nonlinear model was thus proposed to reproduce the significantly higher masking effects obtained in the overlapping threshold components by linear models [27]. Differences of the masked thresholds resulting from a linear and nonlinear superposition are discussed in [15]. Results indicate that significant improvements are possible using a nonlinear model.

A "power-law model," as described in 1995 by Baumgarte [15], was therefore used for the superposition of different masked thresholds in order to represent the nonlinear superposition. The "power-law model" is defined by the parameter α where $0 < \alpha \leq 1$. If α is 1, the superposition of thresholds is linear; if α is lower than 1, the superposition is carried out in a nonlinear mode. A description of different values of α can be also obtained from [15]. The nonlinear superposition of masking thresholds defined by $I_T(z)$ is

$$I_T(z) = \left[\left[I_{abs}(z) \right]^\alpha + \sum_i \left[I_i(z) \right]^\alpha \right]^{1/\alpha}. \quad (9)$$

The level in decibels of the superposition of the individual masking thresholds denoted by $L_T(z)$ is

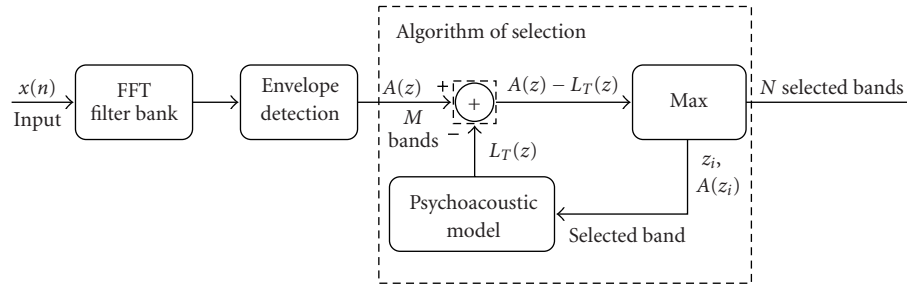$$L_T(z) = 10 \log_{10} \left(I_T(z) \right). \quad (10)$$

FIGURE 7: Selection algorithm: the audio samples are the input and the N bands selected are the output. A psychoacoustic model is used to select the bands in each iteration.

2.3.2. Selection algorithm

This algorithm is inspired by the analysis/synthesis loop [14] used in the MPEG-4 parametric audio coding tools "harmonic and individual lines plus noise" (HILN) [28]. The selection algorithm loop chooses the N bands iteratively in order of their "significance" (Figure 7).

The amplitude envelopes of the M bands $A(z)$ ($z = 1, \ldots, M$) are obtained from the filter bank. For the first iteration of the algorithm there is no masking threshold and the threshold in quiet is not considered; the first band selected is therefore the one with the largest amplitude. For this band, the psychoacoustic model calculates its associated masking threshold $L_T(z)$ ($z = 1, \ldots, M$).

In the next iteration the band z_i is selected out of the remaining $M - 1$ bands for which the following difference is largest:

$$z_i = \operatorname{argmax} \left(A(z) - L_T(z) \right), \quad z = 1, \ldots, M. \tag{11}$$

The individual masking threshold of this band $L_i(z)$ is calculated and added to the one previously determined. The masking threshold $L_T(z)$ for the actual iteration is then obtained and used to select the following band. The loop (Figure 7) is repeated until the N bands are selected. Therefore, at each step of the loop, the psychoacoustic model selects the band that is considered as most significant in terms of perception.

2.3.3. Application to the ACE strategy

The psychoacoustic model has been incorporated into a research ACE strategy made available by Cochlear Corporation as a Matlab "toolbox," designated the nucleus implant communicator (NIC). However, this ACE strategy does not incorporate the pre-emphasis and adaptive-gain control filters described in Section 2.1. The new strategy based on psychoacoustic masking has been termed the psychoacoustic ACE (PACE) strategy as explained in Section 2.3. The NIC allows the ACE and the PACE to be configured using different parameters: the rate of stimulation on a channel (channel stimulation rate), the number of electrodes or channels into which the audio signal is decomposed (M), and the number of bands selected per cycle (N). At the same time, the psychoacoustic model can be modified according to the parameters that define the spreading function (Figure 6). In the

following paragraphs we will describe the rationale for setting the parameter values that are used in the experiments.

2.3.3.1. Parameter setting for the PACE strategy

The parameter set that defines the spreading function should describe the spectral masking effects that take place in a healthy auditory system. Such effects depend strongly on the type of components that are masking and being masked [11]. However, they can be reduced to two general situations: masking of pure tones by noise and masking of pure tones by tones [11]. Furthermore, the first scenario should identify the type of masking noise, that is, whether it is broadband, narrowband, lowpass or highpass noise. For the second scenario, it should also be specified which kind of tone is having a masking effect, that is, whether it is pure tone or a set of complex tones. For each of these situations a different parameter set for the spreading function should be defined, depending on the frequencies and amplitudes of the masker and masked components. For example, in audio compression algorithms such as the MPEG1 layer 3 (MP3) [17] usually only two situations are considered [23]: noise-masking tone (NMT) and tone-masking noise (TMN). For each scenario, a different shape for the spreading function based on empirical results is defined.

The psychoacoustic model applied in this pilot study does not discriminate between tonal and noise components. Furthermore, it is difficult to specify a set of parameters for the spreading function based on empirical results as with the MP3. The parameters of the spreading function in the MP3 can be set through empirical results with normal hearing people. There are a lot of studies in this field which can be used to set the parameters of the spreading function in all the situations mentioned before. However, with cochlear implant users there is relatively little data in this field. For this reason, the results of previous studies by different authors with normal hearing people [11, 12, 13] were incorporated into a unique spreading function approximating all the masking situations discussed above. In these studies the necessity became apparent for the right slope of the spreading function to be less steep than the left slope. In consequence, the left slope of the PACE psychoacoustic model was always set to higher dB/band values than the right slope. Two configurations for the left and right slopes were chosen in order to
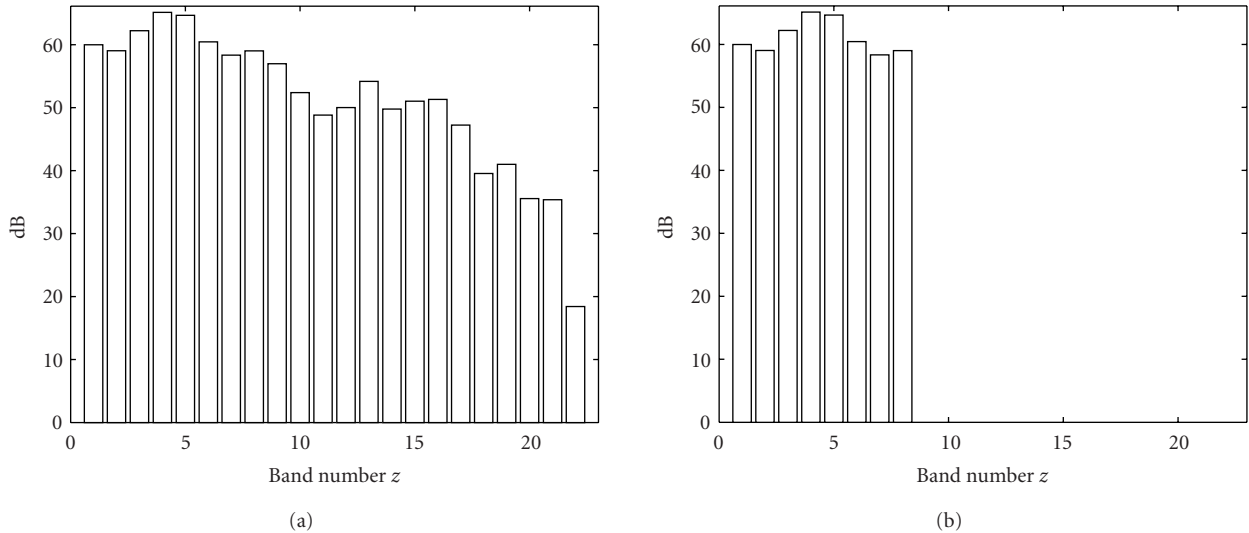
(a)

(b)

FIGURE 8: (a) Frequency band decomposition of one frame coming from a token of the vowel "a." (b) Selected bands using the ACE strategy for one frame coming from a token of the vowel "a."

test different masking effects: (left slope = 12 dB/band, right slope = 7 dB/band) and (left slope = 40 dB/band, right slope = 30 dB/band). Furthermore, outcomes from previous studies demonstrated that the value of a_v defining the attenuation of the spreading function with regard to the masker level is highly variable, ranging between 4 dB and 24 dB depending on the type of masker component [23]. For this reason, the value of a_v was set to 10 dB, which lies between the values mentioned above. The parameter α which controls the nonlinear superposition of individual masking thresholds was set to 0.25, which is in the range of values proposed in [15, 27]. Finally, the threshold in quiet was set to an appropriate level as presented in Section 2.3.1.1.

2.3.3.2. Objective analysis

The NIC software described permits a comparison between the ACE strategy and the psychoacoustic ACE strategy. Figure 8a shows the frequency decomposition of a speech token processed with both strategies. The token is the vowel introduced in Section 2.3.1.1. The filter bank used for both strategies decomposes the audio signal into 22 bands ($M = 22$). Eight of the separated-out bands are selected ($N = 8$). The bands selected differ between the two strategies, as different methods of selecting the amplitudes were used. Figure 8b gives the bands selected by the ACE strategy. Figures 9a, 9b, 10a, and 10b, respectively, illustrate the bands selected by the PACE strategy and the spreading functions used in the psychoacoustic model.

The spreading function presented in Figure 10b is steeper than that demonstrated in Figure 9b. Thus, using the psychoacoustic model based on the spreading function in Figure 9b, any frequency band will have a stronger masking effect over the adjacent frequency bands than with the psychoacoustic model based on the spreading function in

Figure 10b. The psychoacoustic models based on the spreading function shown in Figures 9b and 10b are referred to in the following sections as psychoacoustic models 1 and 2, respectively.

Looking at Figures 8, 9, and 10 it can be observed that the bands selected using a psychoacoustic model are distributed broadly across the frequency range, in contrast to the stimulation pattern obtained with the simple peak-picking "NofM" approach used in the standard ACE strategy. The ACE strategy tends to select groups of consecutive frequency bands, increasing the likelihood of channel interaction between adjacent electrodes inside the cochlea. In the PACE strategy, however, the selection of clusters is avoided owing to the masking effect that is exploited in the psychoacoustic model. This feature can be confirmed by an experiment that involves counting the number of clusters of different lengths selected by the ACE and PACE strategies during the presentation of 50 sentences from a standardized sentence test [29]. For the PACE the test material was processed twice, the first time using psychoacoustic model 1 and then using psychoacoustic model 2. The 50 sentences were processed using a channel stimulation rate of 500 Hz and selecting 8 bands in each frame for both strategies. This means that the maximum possible cluster length is 8, when all selected bands are sequenced consecutively across the frequency range as demonstrated in Figure 8b. The minimum possible cluster length is 1, which occurs when all selected bands are separated from each other by at least one channel. Table 2 presents the number of clusters of different lengths (1–8) for the ACE, PACE 1 (using psychoacoustic model 1) and PACE 2 (using psychoacoustic model 2) strategies that occur during the 50 sample sentences.

The data clearly show that ACE tends on average to produce longer clusters than PACE 1 or PACE 2. At cluster length eight, for example, the ACE strategy selects 3607 clusters,
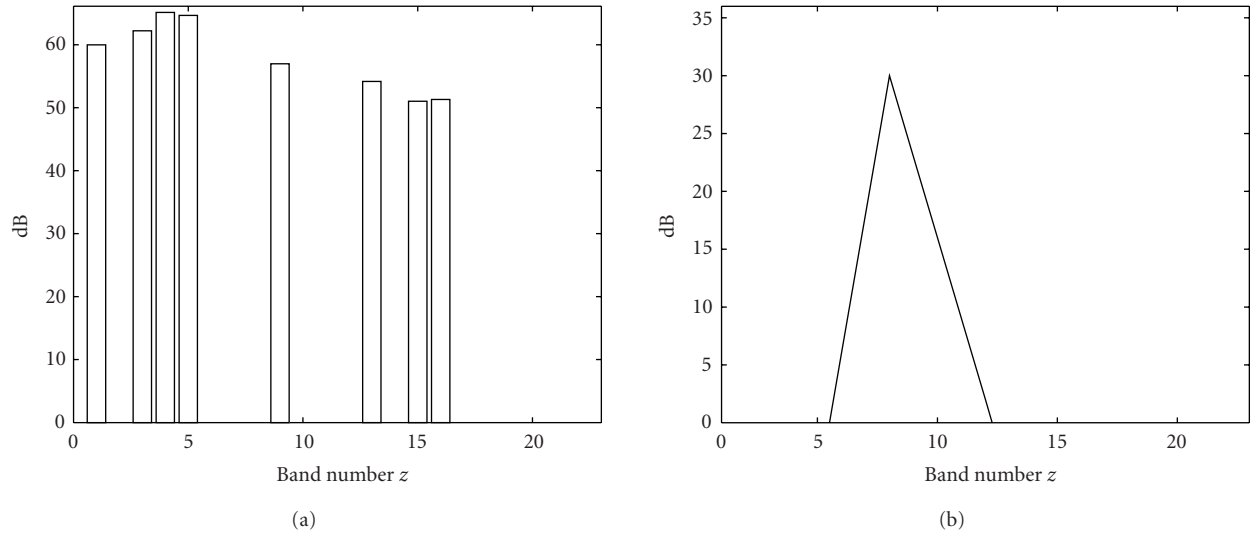
(a)

(b)

FIGURE 9: (a) Selected bands using the PACE strategy for one frame coming from a token of the vowel "a." (b) Spreading function used in the psychoacoustic model (left slope = 12 dB/band, right slope = 7 dB/band, a_v = 10 dB).
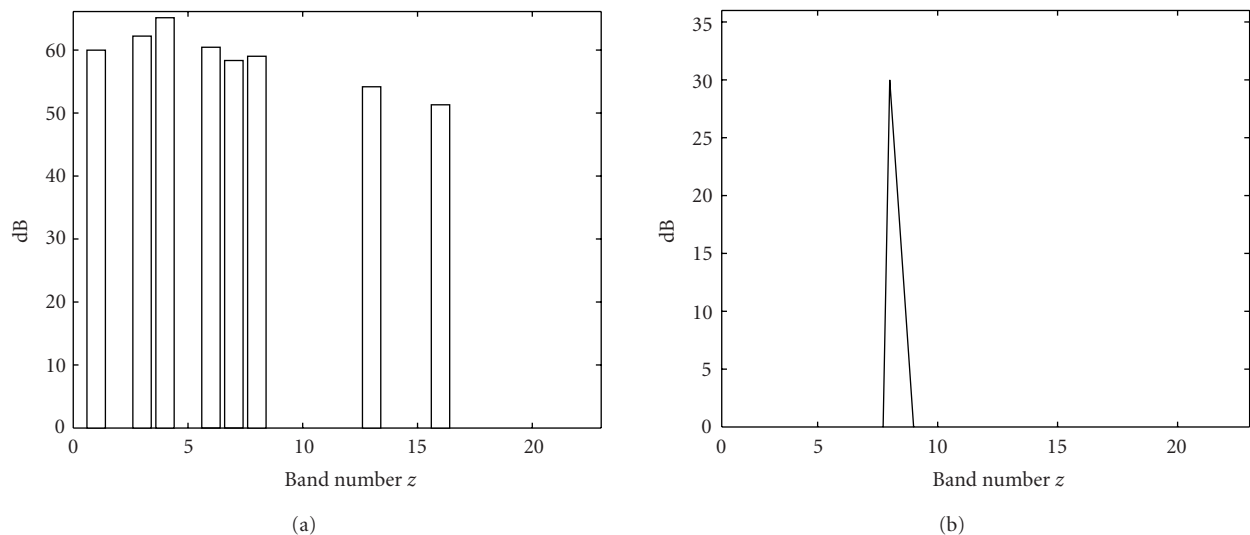

(a)

(b)

FIGURE 10: (a) Selected bands using the PACE strategy for one frame coming from a token of the vowel "a." (b) Spreading function used in the psychoacoustic model (left slope = 40 dB/band, right slope = 30 dB/band, a_v = 10 dB).

whereas the PACE strategy with the psychoacoustic model 1 selects only 33 and the PACE strategy with the psychoacoustic model 2 selects 405. The fact that the PACE 1 selects fewer clusters of 8 bands than the PACE 2 is attributable to the masking effect of the first psychoacoustic model being stronger than the second, as defined by the spreading functions of Figures 9b and 10b.

2.4. Speech intelligibility tests

2.4.1. Test environment

The strategies programmed within the NIC environment were tested with patients using a Nucleus 24 implant manufactured by Cochlear Corporation. The NIC software permits

the researcher to communicate with the Nucleus implant and to send any stimulus pattern to any of the 22 electrodes. The NIC communicates with the implant via the standard hardware also used for fitting recipients in routine clinical practice. A specially initialized clinical speech processor serves as a transmitter for the instructions from the personal computer (PC) to the subject's implant (Figure 11), so that the clinical processor does not itself perform any speech coding computations. The NIC, in conjunction with Matlab, processes the audio signals on a PC. An interface then provides the necessary functionality for a user application that takes signals, processed using the Matlab toolbox, and transmits them to the cochlear implant via the above-mentioned speech processor.

TABLE 2: Number of times that consecutive frequency bands or clusters are selected for different group lengths for the ACE and PACE strategies (using psychoacoustic model 1) and PACE (using psychoacoustic model 2).

Cluster length	Number of ACE clusters	Number of PACE 1 clusters	Number of PACE 2 clusters
1	60 564	370 161	186 338
2	34 248	107 057	114 201
3	20 557	21 449	46 124
4	15 382	3509	18 314
5	12 671	1424	8356
6	15 287	943	3129
7	17 153	566	1382
8	3607	33	405

The Nucleus 24 implant can use up to a maximum of 22 electrodes. However, only 20 electrodes were used by all of our test subjects as their speech processor in everyday use, the "ESPrit 3G," only supports 20 channels and the testees were accustomed to that configuration. For this reason, the two most basal channels were dropped from the original filter bank presented in Section 2.2 and thus could not be selected for stimulation.

2.4.2. Subjects

Eight adult users of the Nucleus 22 cochlear implant system participated in this study. The relevant details for all subjects are presented in Table 3. All test subjects used the ACE strategy in daily life and all were at least able to understand speech in quiet.

2.4.3. Study design

The test material used was the HSM (Hochmair, Schulz, Moser) sentence test [29]. Together with the Oldenburger sentence test [30], this German sentence test is well accepted among German CI centres as a measure of speech perception in cochlear implant subjects. It consists of 30 lists, each with a total of 106 words in 20 everyday sentences consisting of three to eight words. Scoring is based on "words correct." The test was created to minimize outcome variations between the lists. A study involving 16 normal-hearing subjects in noisy conditions (SNR $= -10$ dB) yielded 51.3% correctly repeated words from the lists, with a small range of only 49.8% to 52.6% [29]. The test can be administered in quiet and noise. The noise has a speech-shaped spectrum as standardized in CCITT Rec. 227 [31], and is added keeping fixed the overall output level of the test material.

In order to find suitable parameters of the spreading function in the PACE strategy, HSM test material was processed using two different parameter settings for the spreading function, as described in Section 2.3.3.1. Test signals were then delivered to the implants and the subjects reported which samples sounded clearer and more comfortable. The signals were presented in both quiet and noise. The channel stimulation rate was adapted to the needs of each user and
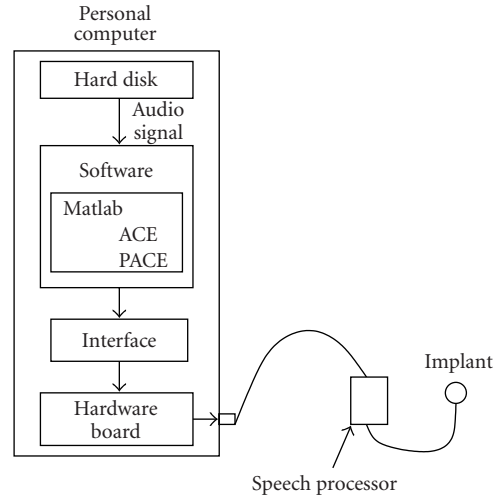

FIGURE 11: Research hardware made available by cochlear corporation.

both 4 and 8 maxima were tried. This procedure was carried out on 3 subjects over a period of several hours. All 3 subjects reported that the sound was best when using the spreading function shown in Figure 10b (psychoacoustic model 2). This particular spreading function was subsequently used for all 8 test subjects listed in Table 3.

All tests had to be conducted on an acute basis as the described research environment does not permit any chronic use, that is, take home experience. In generating the subject's program, the same psychophysical data measured in the R126 clinical fitting software were used in both the ACE and PACE programs. The parameters that define the loudness growth function (see Section 2.2): the base level of the loudness S, the saturation level M, and the steepness parameter ρ were set for all the patients to 33.86 dB, 65.35 dB, and 416.2063, respectively, which are the default parameters in the clinical fitting software [2, 20]. However, the S and M values were converted to the linear amplitudes s and m in order to be inserted in (5) according to the scaling described in Section 2.3.1. Using these values guaranteed that the level of the HSM sentence test was correctly mapped into the dynamic range defined by S and M. The threshold and maximum comfortable levels were adjusted to the needs of each patient. Before commencing actual testing, some sample sentences were processed using both the ACE and PACE strategies. The test subjects spent some minutes listening to the processed material, using both strategies, in order to become familiarized with them. At the same time, the volume was adjusted to suit the needs of the subjects by increasing or decreasing the value of the comfort and threshold levels.

For the actual testing, at least 2 lists of 20 sentences were presented in each condition, with the same number of lists used for both the ACE and PACE conditions. Sentences were presented either in quiet or in noise, depending on the subject's performance (Table 4). The lists of sentences were processed by the ACE and PACE strategies, with either 4 or 8 bands selected per frame. The order of the lists

TABLE 3: Subject demographics.

Patient id	Age	Duration of deafness (y)	Cause of deafness	Implant experience (Nucleus 24) (y)	Electrode type	Usual strategy
P1	65	0.75	Unknown	8	Straight	ACE 8 of 20 1080 pps
P2	42	0.08	Temporal bone fracture	4	Straight	ACE 8 of 20 1080 pps
P3	67	11	Otosclerosis	1	Contour	ACE 8 of 20 900 pps
P4	39	1.21	Unknown	6	Straight	ACE 8 of 20 500 pps
P5	49	2.16	Sudden hearing loss	1.5	Contour	ACE 8 of 20 900 pps
P6	64	0.66	Infection	0.6	Contour	ACE 8 of 20 900 pps
P7	68	1.3	Unknown	2	Contour	ACE 8 of 20 1200 pps
P8	55	0.08	Sudden hearing loss	0.08	Contour	ACE 8 of 20 900 pps

was randomized and the subjects had to repeat each sentence without knowing which strategy they were listening to (ACE or PACE).

As both strategies were tested on the same hardware and are based on the same psychophysical parameters, the tests permitted a fair comparison.

3. RESULTS

All subjects reported that the sound experienced using both strategies was understandable and not very different from what they were used to hearing through their everyday ACE strategy. Subjects 4 and 8 were only presented with sentences in quiet, as they were unable to understand speech in noise. Subject 1 reported that he could not perceive any difference between the two strategies. The other 7 subjects reported that the auditory sensations perceived using the new strategy were more melodious and clearer than those with the ACE, although the everyday speech-coding strategy used by their clinical speech processors was ACE. Subjects 6 and 8 had the impression that the person talking spoke more rapidly when using ACE—a common finding when cochlear implant users are having difficulties in understanding the test material.

Figures 12 and 13 present the averaged scores obtained by each subject for the different tests performed under two conditions, that is, stimulating either 4 or 8 of a total of 20 channels in each cycle. The tests were carried out in noise, with a signal-to-noise ratio of 15 dB (unless otherwise stated).

The results obtained show that all 8 subjects obtained better or equal scores using the PACE strategy when 4 electrodes were stimulated in each frame. When 8 electrodes were stimulated, only subject 7 obtained a better score using the ACE strategy than with the PACE strategy. Subject 2

achieved a better result using 4 electrodes with PACE than when using 8 electrodes with the ACE strategy. However, this may be due to a degree of variability within the test material or simply because of the subjects' diminished concentration at the end of the test session.

The scores show that the difference between the averaged groups becomes more marked when 4 electrodes are selected in each cycle instead of 8. In the former case, as fewer electrodes are stimulated, it becomes more important to select the most relevant amplitudes for each cycle. It was also observed that, when using PACE, performance using 4 electrodes matched that achieved with 8. That indicates that PACE may be able to generate the same scores as ACE while using only half as many electrodes. No significant difference could be found between the 8-channel ACE and 8-channel PACE condition. The above results are supported by the statistical analysis described below.

The program used for the analysis was SPSS V 12; the results were subjected to the Wilcoxon test [32]. Table 5 shows the outcome of the statistical analysis.

The statistical results show that the PACE strategy was found to yield a significant advantage only when 4 channels were selected for stimulation in each cycle. When 8 channels were selected, no significant difference was found between the ACE and PACE strategies.

4. DISCUSSION

The results presented suggest that a psychoacoustic model used to select the N bands in "NofM"-type strategies such as ACE can improve speech recognition by cochlear implant subjects in noise. The mean scores for the HSM sentence test were 65% using the psychoacoustic model and 57% for

TABLE 4: Test details for each patient.

Patient id	Number of lists tested for each condition	Number of electrodes selected	Channel stimulation rate	Noise or quiet conditions
P1	6	4	1080	Noise
P2	2	4	1080	Noise
P2	2	8	1080	Noise
P3	4	4	900	Noise
P4	3	4	500	Quiet
P5	2	4	900	Noise
P5	2	8	900	Noise
P6	2	4	900	Noise
P6	2	8	900	Noise
P7	2	4	1200	Noise
P7	2	8	1200	Noise
P8	2	4	900	Quiet
P8	2	8	900	Quiet

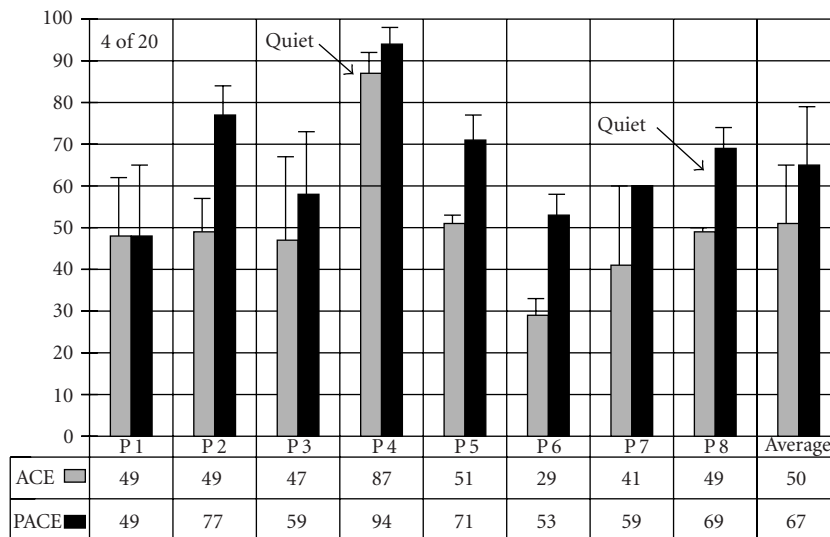|	P 1	P 2	P 3	P 4	P 5	P 6	P 7	P 8	Average
ACE	49	49	47	87	51	29	41	49	50
PACE	49	77	59	94	71	53	59	69	67

FIGURE 12: Score by subject (average and standard deviation). Unless otherwise specified, scores were obtained in noise conditions with a signal-to-noise ratio of 15 dB. The tests were performed using 4 electrodes in each cycle of stimulation.

the ACE strategy when 8 electrodes were stimulated in each cycle of stimulation. The mean score obtained with 4 electrodes stimulated was 67% using the psychoacoustic model and 50% for the ACE strategy. Results were only statistically significant under the 4-channel condition; it is, however, possible that future studies with larger sample sizes may yield significant results for the 8-channel condition as well. Interestingly, performance using PACE was virtually the same regardless of whether 4 or 8 electrodes were used. Therefore, a considerable energy saving could be made using the PACE strategy as it is able to generate the same scores as the ACE strategy while stimulating only half as many electrodes.

Another advantage is that the bands selected using a psychoacoustic model are more widely separated over the frequency domain. It can be speculated that interaction between channels could therefore be reduced. Additionally, the choice of bands is not merely a matter of selecting the largest amplitudes (as with the ACE); this means that smaller electrical currents are required, resulting in power savings.

As can be observed from the results, the difference between the PACE and ACE strategies only exists for the 4-channel condition. This may be because selecting fewer channels means that the spectrum of the audio signal is more poorly represented. This places more of a premium on selecting the right signal components. Using a psychoacoustic model to select the bands appears to be a superior approach to just selecting the channels with higher amplitudes—at least, that is, if the number of selected channels is small. As more channels are selected (8-channel condition), the spectrum of the audio signal is better represented and the selection of the most important components becomes less relevant.
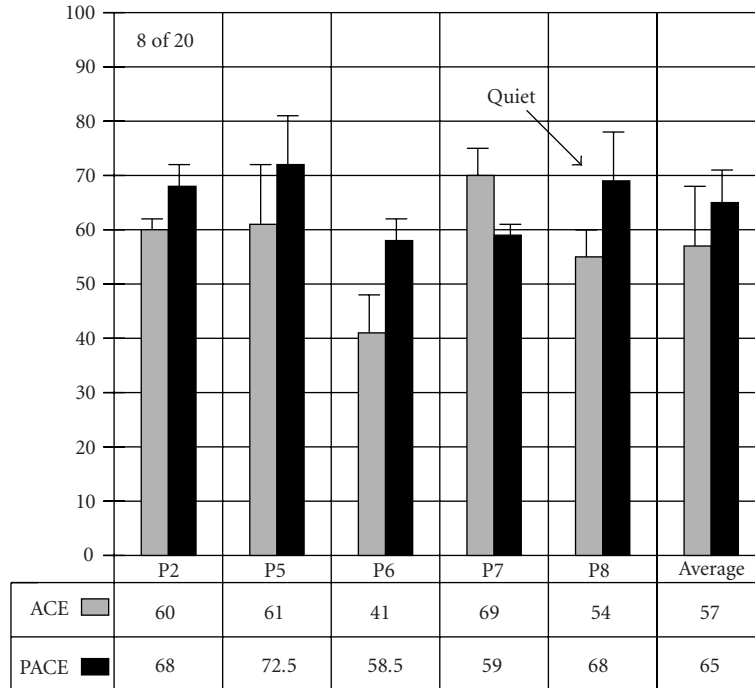
FIGURE 13: Score by subject (average and standard deviation). Unless otherwise specified, scores were obtained in noise conditions with a signal-to-noise ratio of 15 dB. The tests were performed with 8 electrodes in each cycle of stimulation.

TABLE 5: Statistical analysis.

Condition 1	Condition 2	Test	p-value
4-channel ACE	4-channel PACE	Wilcoxon	Significant $p = 0.017$
8-channel ACE	8-channel PACE	Wilcoxon	Not significant $p = 0.138$
4-channel PACE	8-channel PACE	Wilcoxon	Not significant $p = 0.786$
4-channel ACE	8-channel ACE	Wilcoxon	Significant $p = 0.043$

The choice of the parameters that define the spreading function requires more thorough investigation in the future. The spreading function determines how much one channel masks the adjacent frequency bands. As this is not a long-term study and subjects' attention span during speech perception tests is limited, only two different parameter sets are investigated in this paper. The spreading function determined by the first parameter set presented a stronger masking effect than the spreading function determined by the second parameter set. First experiments with cochlear implant subjects revealed that stronger masking effect results in poorer speech perception. One explanation for this might be that important speech cues are being left out by the wider masking curves, which then become inaudible to the subject. Nevertheless, the results obtained thus far are encouraging and indicate the usefulness of a psychoacoustic-masking model in the field of cochlear implants. As the optimal parameter set might vary among subjects, further studies are planned to determine the optimal parameter set for the psychoacoustic-masking model. There are also plans to incorporate masking effects whose occurrence may be due

to overlapping of the electrical fields inside the cochlea. The excitation of a subset of neurons that are being stimulated by adjacent electrodes can be determined by measurements using the neural response telemetry (NRT) capabilities of the Nucleus 24 implant [33]. The data derived from such tests can be used to determine the degree of channel interaction [34] and this knowledge could be additionally exploited in a future version of our masking model. There is, however, currently only relatively limited data on electrical masking in cochlear implant subjects, and this influenced the authors' decision to initially concentrate on a psychoacoustic-masking model for which fundamental knowledge was already available.

It should be reiterated that our research ACE strategy and the new PACE strategy used for the tests do not make use of a pre-emphasis filter. The ACE and PACE strategies process signals fed directly from a computer hard disk, so that the analogue front end of the speech processor containing both pre-emphasis and AGC functionality is bypassed. The high-frequency gain usually leads to the ACE strategy selecting higher-frequency bands than when a pre-emphasis

filter is absent, and high-frequency components are important for speech understanding. The PACE strategy may already account partially for the lack of pre-emphasis by introducing the absolute threshold in quiet function where the higher-frequency parts of a white-noise signal are more above threshold than the low-frequency parts. For this reason the effect of the pre-emphasis may work differently for the PACE strategy than for the ACE strategy.

Another important aspect is the complexity of the new PACE strategy. As presented in Section 3, this strategy uses the same block structure as the ACE strategy but incorporates a psychoacoustic model to select the bands. This allowed the major blocks of the ACE strategy to be adopted for the PACE strategy. Our implementation of PACE on a personal computer was not specifically optimized in terms of computational efficiency. However, it is worth mentioning that the PACE strategy has been already implemented in a commercial speech processor for chronic investigations not posing any challenge in terms of computational demands.

The selection of the appropriate signal components is obviously of great importance. The introduction of simple "NofM" approaches in the 1990s already represented a significant improvement over conventional CIS-like strategies by stimulating fewer electrodes per frame but increasing the channel rate in each channel [6, 7, 35]. However, the stimulation rate may not be the only factor contributing to better hearing with "NofM"-type strategies, as researchers have also observed that these strategies have advantages over CIS-like speech coding using comparable stimulation rates [6, 7, 8]. The close relationship between "NofM"-type strategies and psychoacoustic masking has been already been mentioned in [35].

Advances in the field of speech coding mean that understanding in quiet is no longer a major problem for most recipients, although hearing in noisy conditions is still severely limited [36, 37]. Nevertheless, technical progress in this field has in the recent past led to remarkable performance enhancements in device users. Moreover, intelligent new speech coding strategies such as transient emphasis spectral maxima (TESM), which emphasize certain cues in the audio signal, have demonstrated improvement in terms of speech perception [38]. However, the electrode-nerve interface that is intended to substitute for the hair cells inside the cochlea is clearly not remotely as sophisticated as a fully functional cochlea. With today's systems we are attempting to mimic thousands of nerve fibres using crude electrode arrays that contain 8 to 22 electrode contacts at most. Bearing these limitations in mind, it becomes apparent that the way in which these few electrodes are selected and stimulated plays a key role in helping cochlear implant subjects understand speech in difficult hearing situations.

5. CONCLUSIONS

The results of the PACE strategy, as described above, suggest that psychoacoustic masking is also applicable to cochlear implant recipients. The idea behind the PACE strategy was to present to users of such devices only those signal components that are most clearly perceived by normal-hearing people. In so doing, the limited resolution of the cochlear implant and the electrode-nerve interface can be used more effectively. Results obtained with device users showed significant improvement in speech perception when 4 electrodes were selected using the PACE strategy. No significant improvement was found when 8 electrodes were selected.

One important final comment: it can be expected that the adoption of a psychoacoustic model in speech processors for chronic use may result in even higher scores using the new PACE strategy. The implementation of a psychoacoustic model increases the complexity of simpler "NofM" approaches. However, its implementation is clearly viable in commercial speech processor for cochlear implants. We are currently setting up a long-term study on the PACE strategy which will be conducted using a commercially available speech processor, thus utilizing the usual analogue front end (with AGC and pre-emphasis filter) and giving users take-home experience.

REFERENCES

[1] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, no. 6332, pp. 236–238, 1991.

[2] "ACE Speech Coding Strategy," *Nucleus Technical Reference Manual*, Z43470 Issue 3, Cochlear Corporation, Lane Cove, New South Wales, Australia, December 2002.

[3] "ACE and CIS DSP Strategies," *Software Requirements Specification*, N95287F Issue 1, Cochlear Corporation, Lane Cove, New South Wales, Australia, October 2002.

[4] P. M. Seligman and H. J. McDermott, "Architecture of the spectra 22 speech processor," *Annals of Otology, Rhinology and Laryngology*, vol. 104, suppl. 166, pp. 139–141, 1995.

[5] P. C. Loizou, "Signal-processing techniques for cochlear implants," *IEEE Eng. Med. Biol. Mag.*, vol. 18, no. 3, pp. 34–46, 1999.

[6] J. Kiefer, S. Hohl, E. Sturzebecher, T. Pfennigdorff, and W. Gstoettner, "Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system," *Audiology*, vol. 40, no. 1, pp. 32–42, 2001.

[7] M. W. Skinner, L. K. Holden, L. A. Whitford, K. L. Plant, C. Psarros, and T. A. Holden, "Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," *Ear & Hearing*, vol. 23, no. 3, pp. 207–223, 2002.

[8] H. N. Kim, Y. J. Shim, M. H. Chung, and Y. H. Lee, "Benefit of ACE compared to CIS and SPEAK coding strategies," *Advances in Oto-Rhino-Laryngology*, vol. 57, pp. 408–411, 2000.

[9] M. F. Dorman, P. C. Loizou, and J. Fitzke, "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," *Ear & Hearing*, vol. 19, no. 6, pp. 481–484, 1998.

[10] J. Kiefer, J. Müller, T. Pfennigdorff, et al., "Speech understanding in quiet and in noise with the CIS speech coding strategy (MED-EL Combi-40) compared to the Multipeak and Spectral Peak strategies (Nucleus)," *Journal for Oto-Rhino-Laryngology and Its Related Specialties*, vol. 58, pp. 127–135, 1996.

[11] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer, Berlin, Germany, 1990.

[12] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, London, UK, 3rd edition, 1989.

[13] B. C. J. Moore, D. A. Vickers, C. J. Plack, and A. J. Oxenham, "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," *Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2761–2778, 1999.

[14] B. Edler, H. Purnhagen, and C. Ferekidis, "ASAC—analysis/synthesis codec for very low bit rates," *100th Audio Engineering Society Convention*, Copenhagen, Denmark, preprint 4179, May 1996.

[15] F. Baumgarte, C. Ferekidis, and H. Fuchs, "A nonlinear psychoacoustic model applied to the ISO MPEG layer 3 coder," *99th Audio Engineering Society Convention*, preprint 4087, New York, NY, USA, October 1995.

[16] K. Tsutsui, H. Suzuki, O. Shimoyoshi, M. Sonohara, K. Akagiri, and R. Heddle, "ATRAC: adaptive transform acoustic coding for minidisc," *93rd Audio Engineering Society Convention*, preprint 3456, San Francisco, Calif, USA, October 1992.

[17] ISO/IEC, JTC1/SC29/WG11 MPEG, "Information technology—coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s-part 3: audio," IS11172-3, 1992, ("MPEG-1").

[18] Website of the Fraunhofer institute on MPEG Audio: http://www.iis.fraunhofer.de/amm/techinf/layer3/index.html.

[19] F. J. Harris, "On the use of windows for harmonic analysis with the discrete Fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[20] "Nucleus MATLAB Toolbox 2.11," *Software User Manual*, N95246F Issue 1, Cochlear Corporation, October 2002.

[21] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Processing Mag.*, vol. 15, no. 5, pp. 101–130, 1998.

[22] B. C. J. Moore and B. R. Glasberg, "A model of loudness perception applied to cochlear hearing loss," *Auditory Neuroscience*, vol. 3, pp. 289–311, 1997.

[23] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, 2000.

[24] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, vol. 6, no. 2, pp. 314–323, 1988.

[25] D. M. Green, "Additivity of masking," *Journal of the Acoustical Society of America*, vol. 41, no. 6, pp. 1517–1525, 1967.

[26] R. A. Lutfi, "Additivity of simultaneous masking," *Journal of the Acoustical Society of America*, vol. 73, no. 1, pp. 262–267, 1983.

[27] R. A. Lutfi, "A power-law transformation predicting masking by sounds with complex spectra," *Journal of the Acoustical Society of America*, vol. 77, no. 6, pp. 2128–2136, 1985.

[28] H. Purnhagen and N. Meine, "HILN-the MPEG-4 parametric audio coding tools," in *Proc. IEEE International Symposium on Circuits and Systems (ISCAS '00)*, vol. 3, pp. 201–204, Geneva, Switzerland, May 2000.

[29] I. Hochmair-Desoyer, E. Schulz, L. Moser, and M. Schmidt, "The HSM sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users," *American Journal of Otology*, vol. 18, suppl. 6, pp. 83–83, 1997.

[30] K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests [Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test]," *Zeitschrift für Audiologie*, vol. 38, no. 1, pp. 4–15, 1999.

[31] H. Fastl, "A masking noise for speech intelligibility tests," *Proceedings TC Hearing, Acoustical Society of Japan*, vol. 93, pp. 70–70, 1993.

[32] W. W. Daniel, *Applied Nonparametric Statistics*, PWS-Kent, Boston, Mass, USA, 2nd edition, 1990.

[33] J. K. Shallop, G. W. Facer, and A. Peterson, "Neural response telemetry with the nucleus CI24M cochlear implant," *Laryngoscope*, vol. 109, no. 11, pp. 1755–1759, 1999.

[34] P. J. Abbas, M. L. Hugher, C. J. Brown, C. A. Miller, and H. South, "Channel interaction in cochlear implant users evaluated using the electrically evoked compound action potential," *Audiology & Neuro-Otology*, vol. 9, no. 4, pp. 203–213, 2004.

[35] A. Büchner, R. D. Battmer, P. Boyle, and T. Lenarz, "Clarion Research Interface used to implement an n of m strategy in patients implanted with the CLARION cochlear implant," in *Proc. 6th International Cochlear Implant Conference*, Miami Beach, Fla, USA, February 2000.

[36] B. L. Fetterman and E. H. Domico, "Speech recognition in background noise of cochlear implant patients," *Otolaryngology—Head and Neck Surgery*, vol. 126, no. 3, pp. 257–263, 2002.

[37] M. F. Dorman, P. C. Loizou, J. Fitzke, and Z. Tu, "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," *Journal of the Acoustical Society of America*, vol. 104, no. 6, pp. 3583–3585, 1998.

[38] A. E. Vandali, "Emphasis of short-duration accoustic speech cues for cochlear implants users," *Journal of the Acoustical Society of America*, vol. 109, no. 5, pp. 2049–2061, 2001.
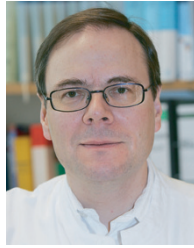
Waldo Nogueira received the Telecommunication Engineering degree from the Polytechnic University of Catalonia (UPC), Barcelona, Spain, in 2003. He wrote his Diploma thesis as an integrated exchange student at the University of Hanover, Germany. Since then, he has been working toward the Ph.D. degree at the Information Technology Laboratory, University of Hanover. His research interests are audio and speech signal processing, psychoacoustics, and audio coding.

Andreas Büchner received his M.S. degree in informatics in 1995 at the University of Hildesheim, Germany, and his Ph.D. degree (human biology) at the Medical University of Hannover in 2002. After developing signal processing and pattern recognition algorithms for medical imaging systems at the University of Hildesheim, he became a Research Scientist in the field of audiology at the Department of Otolaryngology at the Medical University of Hannover (MHH) in 1995. Since 2003, he has been the Scientific Director of the Hannover Hearing Center, subsidiary of the Department of Otolaryngology, MHH. At MHH, he quickly focused on the development and improvement of signal processing strategies for cochlear implants and their introduction into the clinical field. Other research interests are combination of acoustic and electric hearing (electroacoustic stimulation), tinnitus treatment, evaluation, and interpretation of electrically evoked action potentials of the auditory nerve.

Thomas Lenarz pursued his medical studies in otolaryngology during his residency at the University of Tuebingen, Erlangen, London and Heidelberg between 1975 and 1985. He then became Assistant Professor at the Department of Otolaryngology, University of Heidelberg. Since 1993, he has been Chairman and Professor of the Department of Otolaryngology at the Medical University of Hannover, Germany. The department has the largest number of cochlear implant patients in the world. Fields of research are cochlear implants, electrophysiological measurements, research on implantable hearing aids, neonatal hearing screening, pharmacology of the auditory system, otology, and skull base surgery including neuronavigation. He has membership in national/international societies of otolaryngology, audiology, and skull base surgery. He is member of the Editorial Board of the American Journal of Otology, Otology & Neurotology, Laryngo-Rhino-Otology. He has more than 200 publications in national and international journals.

Bernd Edler was a Research Assistant from 1979 to 1985 at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung of the University of Hannover. His major research activities were in the field of filter banks and their applications in audio coding. He developed the window switching and aliasing reduction techniques for the filter bank used in the ISO/MPEG-1/2 Layer 3 audio coding standard, which is usually better known as "mp3." Since 1993, he has been the Head of the Systems Technologies Group at the Information Technology Laboratory of the University of Hannover. His current research fields cover very low-bit-rate audio coding and models for auditory perception. He was actively participating in the development of the MPEG-4 audio standard. From 1998 to 1999, he was a Research Visitor at Lucent Technologies' Bell Laboratories, Murray Hill, NJ.