

Research Article

Signal Processing Strategies for Cochlear Implants Using Current Steering

Waldo Nogueira, Leonid Litvak, Bernd Edler, Jörn Ostermann, and Andreas Büchner

Laboratorium für Informationstechnologie, Leibniz Universität Hannover, Schneiderberg 32, 30167 Hannover, Germany

Correspondence should be addressed to Waldo Nogueira, waldon@abionics.fr

Received 29 November 2008; Revised 19 April 2009; Accepted 22 September 2009

Recommended by Torsten Dau

In contemporary cochlear implant systems, the audio signal is decomposed into different frequency bands, each assigned to one electrode. Thus, pitch perception is limited by the number of physical electrodes implanted into the cochlea and by the wide bandwidth assigned to each electrode. The Harmony HiResolution bionic ear (Advanced Bionics LLC, Valencia, CA, USA) has the capability of creating virtual spectral channels through simultaneous delivery of current to pairs of adjacent electrodes. By steering the locus of stimulation to sites between the electrodes, additional pitch percepts can be generated. Two new sound processing strategies based on current steering have been designed, SpecRes and SineEx. In a chronic trial, speech intelligibility, pitch perception, and subjective appreciation of sound were compared between the two current steering strategies and standard HiRes strategy in 9 adult Harmony users. There was considerable variability in benefit, and the mean results show similar performance with all three strategies.

Copyright © 2009 Waldo Nogueira et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Cochlear implants are an accepted and effective treatment for restoring hearing sensation to people with severe-to-profound hearing loss. Contemporary cochlear implants consist of a microphone, a sound processor, a transmitter, a receiver, and an electrode array that is positioned inside the cochlea. The sound processor is responsible for decomposing the input audio signal into different frequency bands and delivering information about each frequency band to the appropriate electrode in a base-to-apex tonotopic pattern. The bandwidths of the frequency bands are approximately equal to the critical bands, where low-frequency bands have higher frequency resolution than high-frequency bands. The actual stimulation to each electrode consists of nonoverlapping biphasic charge-balanced pulses that are modulated by the lowpass-filtered output of each analysis filter.

Most contemporary cochlear implants deliver interleaved pulses to the electrodes so that no electrodes are stimulated simultaneously. If electrodes are stimulated simultaneously, thereby overlapping in time, their electrical fields add and create undesirable interactions. Interleaved stimulation

partially eliminates these undesired interactions. Research shows that strategies using nonsimultaneous stimulation achieve better performance than strategies using simultaneous stimulation of all electrodes [1].

Most cochlear implant users have limited pitch resolution. There are two mechanisms that can underlie pitch perception in cochlear implant recipients, temporal/rate pitch and place pitch [2]. Rate pitch is related to the temporal pattern of stimulation. The higher the frequency of the stimulating pulses, the higher the perceived pitch. Typically, most patients do not perceive pitch changes when the stimulation rate exceeds 300 pulses per second [3]. Nonetheless, temporal pitch cues have shown to provide some fundamental frequency discrimination [4] and limited melody recognition [2]. The fundamental frequency is important for speaker recognition and speech intelligibility. For speakers of tone languages (e.g., Cantonese or Mandarin), differences in fundamental frequency within a phonemic segment determine the lexical meaning of a word. It is not surprising, then, that cochlear implant users in countries with tone languages may not derive the same benefit as individuals who speak nontonal languages [5].

Speech intelligibility in noise environments might be limited for cochlear implant users because of the poor perception of temporal cues. It has been shown that normal hearing listeners benefit from temporal cues to improve speech intelligibility in noise environments [6].

The place pitch mechanism is related to the spatial pattern of stimulation. Stimulation of electrodes located towards the base of the cochlea produces higher pitch sensations than stimulation of electrodes located towards the apex. The resolution of pitch derived from a place mechanism is limited by the few number of electrodes and the current spread produced in the cochlea when each electrode is activated. Pitch or spectral resolution is important when the listening environment becomes challenging in order to separate speech from noise or to distinguish multiple talkers [7]. The ability to differentiate place-pitch information also contributes to the perception of the fundamental frequency [4]. Increased spectral resolution also is required to perceive fundamental pitch and to identify melodies and instruments [8]. As many as 100 bands of spectral resolution are required for music perception in normal hearing subjects [7].

Newer sound-processing strategies like HiRes are designed to increase the spectral and temporal resolution provided by a cochlear implant in order to improve the hearing abilities of cochlear implant recipients. HiRes analyzes the acoustic signal with high temporal resolution and delivers high stimulation rates [9]. However, spectral resolution is still not optimal because of the limited number of electrodes. Therefore, a challenge for new signal processing strategies is to improve the representation of frequency information given the limited number of fixed electrodes. Recently, researchers have demonstrated a way to enhance place pitch perception through simultaneous stimulation of electrode pairs [3, 10–12]. This causes a summation of the electrical field producing a peak of the overall field located in the middle of both electrodes. It has been reported that additional pitch sensations can be created by adjusting the proportion of current delivered simultaneously to two electrodes [13]. This technique is known as current steering [7]. As the implant can represent information with finer spectral resolution, it becomes necessary to improve the spectral analysis of the audio signal performed by classical strategies like HiRes.

In addition to simultaneous stimulation of electrodes, multiple intermediate pitch percepts also can be created using by sequential stimulation of adjacent electrodes in quick succession [14]. Electrical models of the human cochlea and psychoacoustic experiments have shown that simultaneous stimulation generally is able to produce a single, gradually shifting intermediate pitch. On the other hand, sequential stimulation often produces two regions of excitation. Thus, sequential stimulation often requires an increase in the total amount of current needed to reach comfortable loudness, and may lead to the perception of two pitches or a broader pitch as the electrical field separates into two regions [15].

The main goal of this work was to improve speech and music perception in cochlear implant recipients through the development of new signal processing strategies that

take advantage of the current-steering capabilities of the Advanced Bionics device. These new strategies were designed to improve the spectral analysis of the audio signal and to deliver the signal with greater place precision using current steering. The challenge was to implement the experimental strategies in commercial speech processors so that they could be evaluated by actual implanted subjects. Thus a significant effort was put into executing the real-time applications in commercial low power processors. After implementation, the strategies were assessed using standardized tests of pitch perception and speech intelligibility and through subjective ratings of music appreciation and speech quality.

The paper is organized as follows. Section 2 describes the commercial HiRes and two research strategies using current steering. Section 3 details the methods for evaluating speech intelligibility and frequency discrimination in cochlear implant recipients using the new strategies. Sections 4, 5, and 6 present the results, discussion, and conclusions.

2. Methods

2.1. The High Resolution Strategy (HiRes). The HiRes strategy is implemented in the Auria and Harmony sound processors from Advanced Bionics LLC. These devices can be used with the Harmony implant (CII and the HiRes90k). In HiRes, an audio signal sampled at 17400 Hz is pre-emphasized by the microphone and then digitized. Adaptive gain control (AGC) is performed digitally using a dual-loop AGC [16]. Afterwards the signal is broken up into frequency bands using infinite impulse response (IIR) sixth-order Butterworth filters. The center frequencies of the filters are logarithmically spaced between 350 Hz and 5500 Hz. The last filter is a high-pass filter whose bandwidth extends up to the Nyquist frequency. The bandwidth covered by the filters will be referred to as subbands or frequency bands. In HiRes, each frequency band is associated with one electrode.

In HiRes, the subband outputs of the filter bank are used to derive the information that is sent to the electrodes. Specifically, the filter outputs are half-wave rectified and averaged. Half-wave rectification is accomplished by setting to 0 the negative amplitudes at the output of each filter band. The outputs of the half-wave rectifier are averaged for the duration T_s of a stimulation cycle. Finally, the “Mapping” block maps the acoustic values obtained for each frequency band into current amplitudes that are used to modulate biphasic pulses. A logarithmic compression function is used to ensure that the envelope outputs fit the patient’s dynamic range. This function is defined for each frequency band or electrode z ($z = 1, \dots, M$) and is of the form presented in the following equation:

$$Y_z(X_{\text{Filt}_z}) = \frac{(\text{MCL}(z) - \text{THL}(z))}{\text{IDR}} \times (X_{\text{Filt}_z} - m_{\text{satdB}} + 12 + \text{IDR}) + \text{THL}(z) \quad (1)$$

$$z = 1, \dots, M,$$

where Y_z is the (compressed) electrical amplitude, X_{Filt_z} is the acoustic amplitude (output of the averager) in dB and IDR is

the input dynamic range set by the clinician. A typical value for the IDR is 60 dB. The mapping function used in HiRes maps the MCL at 12 dB below the saturation level m_{satdB} . The saturation level in HiRes is set to $20 \log_{10}(2^{15} - 1)$.

In each stimulation cycle, HiRes stimulates all M implant electrodes sequentially to partially avoid channel interactions. The number of electrodes for the HiRes90k implant is $M = 16$, and all electrodes are stimulated at the same fixed rate. The maximum channel stimulation rate (CSR) used in the HiRes90k is 2899 Hz.

2.2. The Spectral Resolution Strategy (SpecRes). The spectral resolution (SpecRes) strategy is a research version of the commercial HiRes with Fidelity 120 strategy and, like HiRes can be used with the Harmony implant. This strategy was designed to increase the frequency resolution so as to optimize use of the current steering technique. In [10], it was shown that cochlear implant subjects are able to perceive several distinct pitches between two electrodes when they are stimulated simultaneously. In HiRes each center frequency and bandwidth of a filter band is associated with one electrode.

However, when more stimulation sites are created using current steering, a more accurate spectral analysis of the incoming sound is required. For this reason, the filter bank used in HiRes is not adequate and a new signal processing strategy that enables higher spectral resolution analysis is required. Figure 1 shows the main processing blocks of the new strategy designed by Advanced Bionics LLC.

In SpecRes, the signal from the microphone is first pre-emphasized and digitized at $F_s = 17400$ Hz as in HiRes. Next the front-end implements the same adaptive-gain control (AGC) as used in HiRes. The resulting signal is sent through a filter bank based on a Fast Fourier Transform (FFT). The length of the FFT is set to $L = 256$ samples; this value gives a good compromise between spectral resolution (related to place pitch) and temporal resolution (related to temporal pitch). The longer the FFT, the higher the frequency resolution and thus, the lower the temporal resolution.

The linearly spaced FFT bins then are grouped into analysis bands. An analysis band is defined as spectral information contained in a range allocated to two electrodes. For each analysis band, the Hilbert envelope is computed from FFT bins. In order to improve the spectral resolution of the audio signal analysis, an interpolation based on a spectral peak locator [17] inside each analysis band is performed. The spectral peaks are an estimation of the most important frequencies. The frequency estimated by the spectral peak locator is used by the frequency weight map and the carrier synthesis. The carrier synthesis generates a pulse train with the frequency determined by the spectral peak locator in order to deliver temporal pitch information. The frequency weight map converts the frequency determined by the spectral peak locator into a current weighting proportion that is applied to the electrode pair associated with the analysis band.

All this information is combined and nonlinearly mapped to convert the acoustical amplitudes into electrical

current amplitudes. For each stimulation cycle, pairs of electrodes associated with one analysis band are stimulated simultaneously, but the pairs of channels are stimulated sequentially in order to reduce undesired channel interaction. Furthermore, the order of stimulation is selected to maximize the distance between consecutive analysis bands being stimulated. This approach reduces further channel interaction between stimulation sites. The next section presents each block of SpecRes in detail.

2.2.1. FFT and Hilbert Envelope. The FFT is performed on input blocks of $L = 256$ samples of the previously windowed audio signal:

$$x_w(l) = x(l)w(l), \quad l = 0, \dots, L-1, \quad (2)$$

where $x(l)$ is the input signal and $w(l)$ is a 256-blackman hanning window:

$$w(l) = \frac{1}{2} \left(0.42 - 0.5 \cos\left(\frac{2\pi l}{L}\right) + 0.08 \cos\left(\frac{4\pi l}{L}\right) \right) + \frac{1}{2} \left(0.5 - 0.5 \cos\left(\frac{2\pi l}{L}\right) \right) \quad l = 0, \dots, L-1. \quad (3)$$

The FFT of the windowed input signal can be decomposed into its real and imaginary components as follows:

$$X(n) = \text{FFT}(x_w(l)) \\ = \text{Re}\{X(n)\} + j \text{Im}\{X(n)\}, \quad n = 0, \dots, L-1, \quad (4)$$

where

$$\text{Re}\{X(n)\} \triangleq X_r(n) = \frac{1}{L} \sum_{l=0}^{L-1} x_w(l) \cos\left(2\pi \frac{n}{L} l\right), \\ \text{Im}\{X(n)\} \triangleq X_i(n) = \frac{1}{L} \sum_{l=0}^{L-1} x_w(l) \sin\left(2\pi \frac{n}{L} l\right). \quad (5)$$

The linearly spaced FFT bins are then combined to provide the required number of analysis bands N . Because the number of electrodes in Harmony implant is $M = 16$ electrodes, the total number of analysis bands is $N = M - 1 = 15$. Table 1 presents the number of FFT bins assigned to each analysis band and its associated center frequency.

The Hilbert envelope is computed for each analysis band. The Hilbert envelope for the analysis band z is denoted by H_{Ez} and is computed from the FFT bins as follows:

$$H_{r_z}(\tau) = \sum_{n=n_{\text{start}_z}}^{n_{\text{end}_z}-1} X_r(n) \cos\left(\frac{2\pi n \tau}{L}\right) - X_i(n) \sin\left(\frac{2\pi n \tau}{L}\right), \\ H_{i_z}(\tau) = \sum_{n=n_{\text{start}_z}}^{n_{\text{end}_z}-1} X_r(n) \sin\left(\frac{2\pi n \tau}{L}\right) - X_i(n) \cos\left(\frac{2\pi n \tau}{L}\right), \quad (6)$$

where H_{r_z} and H_{i_z} are the real and imaginary parts of the Hilbert transform, τ is the delay within the window and $n_{\text{end}_z} = n_{\text{start}_z} + N_z$.

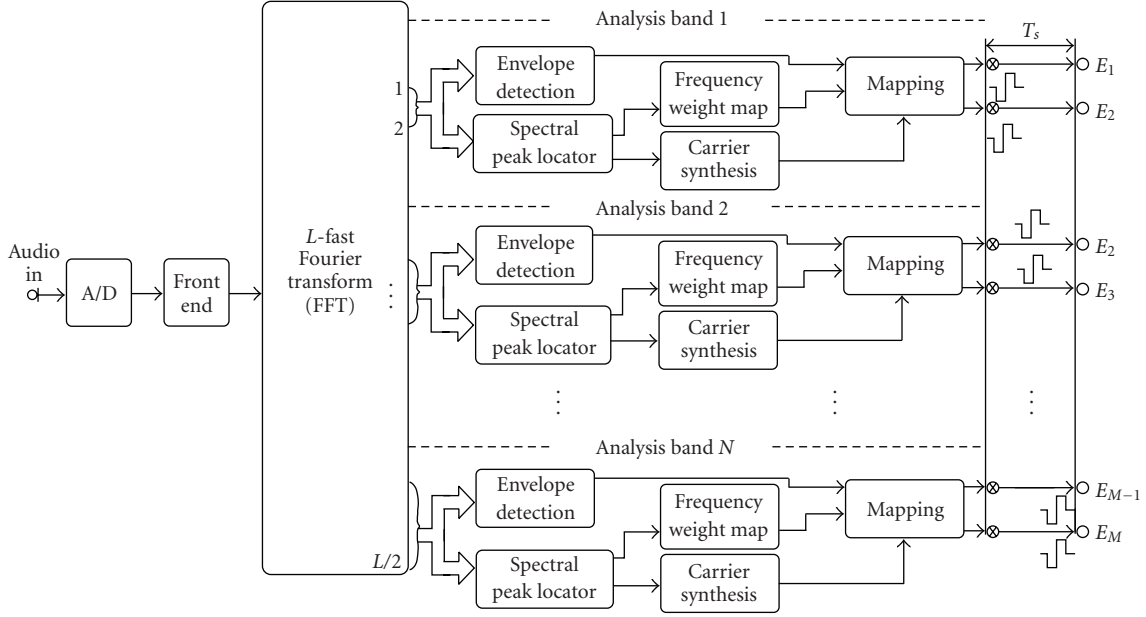


FIGURE 1: Block diagram illustrating SpecRes.

TABLE 1: Number of FFT bins related to each analysis band and its associated center frequencies in Hz. The FFT bins have been grouped in order to match the center frequencies of the standard HiRes filterbank used in clinical routine practice.

Analysis band z	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Number of bins N_z	2	2	1	2	2	2	3	4	4	5	6	7	8	10	55
Start bin n_{start_z}	5	7	9	10	12	14	16	19	23	27	32	38	45	53	63
Center freqs f_{center} (Hz)	408	544	646	748	884	1020	1190	1427	1700	2005	2379	2821	3330	3942	6491

Specifically, for $\tau = L/2$, the Hilbert transform is calculated in the middle of the analysis window:

$$H_{r_z} = \sum_{n=n_{\text{start}_z}}^{n_{\text{end}_z}} X_r(n)(-1)^n, \quad (7)$$

$$H_{i_z} = \sum_{n=n_{\text{start}_z}}^{n_{\text{end}_z}} X_i(n)(-1)^n.$$

the Hilbert envelope $HE(\tau)$ is obtained from the Hilbert transform as follows:

$$HE(\tau) = \sqrt{H_{r_z}(\tau)^2 + H_{i_z}(\tau)^2}. \quad (8)$$

To implement stimulation at different positions between two electrodes, each analysis channel can create multiple virtual channels by varying the proportion of current delivered to adjacent electrodes simultaneously. The weighting applied to each electrode is controlled by the spectral peak locator and the frequency weight map.

2.2.2. Spectral Peak Locator. Peak location is determined within each analysis band z . For a pure tone within a channel,

spectral peak location should estimate the frequency of the tone. The frequency resolution obtained with the FFT is half a bin. A bin represents a frequency interval of F_s/L Hz. The maximum resolution that can be achieved is therefore 67.96 Hz. However, it has been shown in [12] that patients are able to perceive a maximum of around 30 distinct pitch percepts between pairs of the most apical electrodes. Because the bandwidth associated with the most apical electrode pair is around 300 Hz and the maximum resolution is 30 pitch percepts, the spectral resolution required for the analysis should be around 10 Hz. This resolution is accomplished by using a spectral peak locator. Spectral peak location is computed in two steps. The first step is to determine the FFT bin within an analysis band with the most energy. The power $e(n)$ in each bin equals the sum of the squared real and the imaginary parts of that bin:

$$e(n) = X_r^2(n) + X_i^2(n). \quad (9)$$

The second step consists of fitting a parabola around the bin n_{max_z} containing maximum energy in an analysis band z , that is, $e(n_{\text{max}_z}) \geq e(n)$ for all $n \neq n_{\text{max}_z}$ in that analysis band. To describe the parabolic interpolation strategy, a coordinate

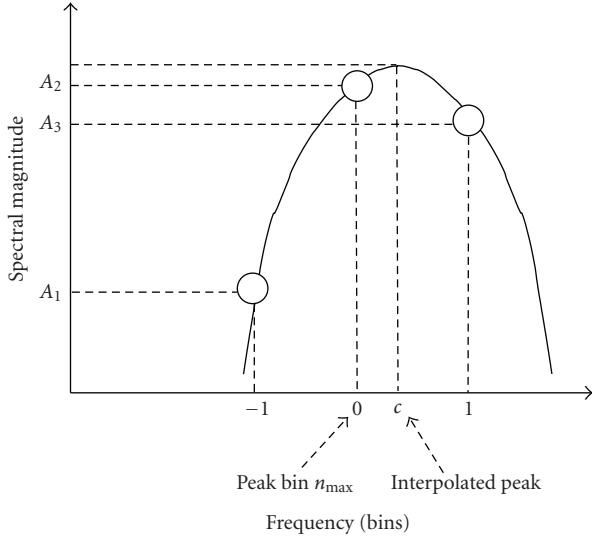


FIGURE 2: Parabolic fitting between three FFT bins.

system centered at n_{\max} is defined. $e(n_{\max} - 1)$ and $e(n_{\max} + 1)$ represent the energy of the two adjacent bins. By taking the energies in dB, we have

$$\begin{aligned} A_1 &= 20 \log_{10}(e(n_{\max_z} - 1)), \\ A_2 &= 20 \log_{10}(e(n_{\max_z})), \\ A_3 &= 20 \log_{10}(e(n_{\max_z} + 1)). \end{aligned} \quad (10)$$

The optimal location is computed by fitting a generic parabola

$$y(f) = a(f - c)^2 + b, \quad (11)$$

to the amplitude of the bin n_{\max} and the amplitude of the two adjacent bins and taking its maximum. a , b , and c are variables and f indicates frequency in Hz.

Figure 2 illustrates the parabolic interpolation [17, 18]. The center point or vertex c gives the interpolated peak location (in bins). The parabola is evaluated at the three bins nearest to the center point c :

$$\begin{aligned} y(-1) &= A_1, \\ y(0) &= A_2, \\ y(1) &= A_3. \end{aligned} \quad (12)$$

The three samples can be substituted in the parabola defined in (11). This yields the frequency difference in FFT bins:

$$c = \frac{1}{2} \frac{A_1 - A_3}{A_1 - 2A_2 + A_3} \in \left[-\frac{1}{2}, \frac{1}{2}\right], \quad (13)$$

and the estimate of the peak location (in bins) is

$$n_{\max_z}^* = n_{\max_z} + c. \quad (14)$$

If the maximum bin within the channel is not the local maximum, this can only occur near the boundary of the channel, the spectral peak locator is placed at the boundary of the channel.

2.2.3. Frequency-Weight-Map. The purpose of the frequency-weight-map is to translate the spectral peak into cochlear location. For each analysis band z two weights are calculated w_{z_1} and w_{z_2} that will be applied to the two electrodes forming that analysis band. This can be achieved using the cochlear frequency-position function [19]

$$f = A(10^{ax}), \quad (15)$$

f represents the frequency in Hz and x the position (in mm) along the cochlea. A and a were set to 350 Hz and 0.07, respectively, considering the known dimensions of the CII and HiRes90k [20]. The locations associated to the electrodes were calculated by substitution of its corresponding frequencies in the above equation. The location of each electrode is denoted by x_z ($z = 1, \dots, M$).

The peak frequencies are also translated to positions using (15). The location corresponding to a peak frequency in the analysis band z is denoted by x_{z_p} . To translate a cochlear location to weights that will be applied to individual currents of each electrode, the peak location is subtracted from the location of the first electrode x_z in a pair (x_z, x_{z+1}). The weight applied to the second electrode x_{z+1} (higher frequency) of the pair is calculated using the following equation:

$$w_{z_2} = \frac{x_{z_p} - x_z}{d_z}, \quad (16)$$

and the weight applied to first electrode x_z of the pair is

$$w_{z_1} = \frac{x_{z+1} - x_{z_p}}{d_z}, \quad (17)$$

where d_z is the distance (in mm) between the two electrodes forming an analysis band, that is,

$$d_z = |x_{z+1} - x_z|. \quad (18)$$

2.2.4. Carrier Synthesis. The carrier synthesis attempts to compensate for the low temporal resolution given by the FFT-based approach. The goal is to enhance temporal pitch perception by representing the temporal structure of the frequency corresponding to the spectral peak in each analysis band. Note that the electrodes are stimulated with a current determined by the HE at a constant rate determined by the CSR. The carrier synthesis modulates the Hilbert envelope of each analysis band with a frequency coinciding with the frequency of the spectral peak.

Furthermore, the modulation depth (relative amount of oscillation from peak to valley) is reduced with increasing frequency as shown in Figure 3.

The carrier synthesis defines the phase variable $ph_{h,z}$ for each analysis band z and frame h , where $0 \leq ph_{h,z} \leq \text{CSR} - 1$. During each frame h , $ph_{h,z}$ is increased by the minimum of the estimated frequency f_{\max_z} and CSR:

$$ph_{h,z} = (ph_{h-1,z} + \min(f_{\max_z}, \text{CSR})) \bmod (\text{CSR}), \quad (19)$$

where $f_{\max_z} = n_{\max_z}^* (F_s/L)$, h indicates the actual frame, and mod indicates the modulo operator.

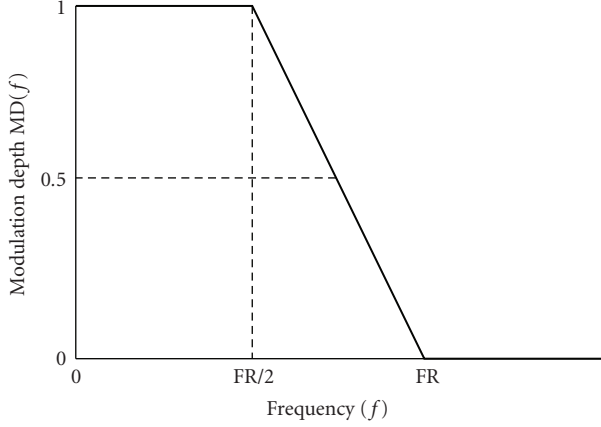


FIGURE 3: Modulation depth as a function of frequency. FR is a constant of the algorithm equal to 2320 Hz which is the maximum channel stimulation rate that can be delivered with the implant using the current steering technique.

The parameter s is defined for each analysis band z as follows:

$$s_z = \begin{cases} 1, & ph_{h,z} \leq \frac{CSR}{2}, \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Then, the final carrier for each analysis band z is defined as

$$c_z = 1 - s_z MD(f_{\max_z}), \quad (21)$$

where $MD(f_{\max_z})$ is the modulation depth function defined in Figure 3.

2.2.5. Mapping. The final step of the SpecRes strategy is to convert the envelope, weight, and carrier into the current magnitude to apply to each electrode pair associated with each analysis band. The mapping function is defined as in HiRes (1). For the two electrodes in the pair that comprise the analysis band; the current delivered is given by

$$I_z = Y_z(\max(\text{HE}_z)) w_{z_1} c_z, \quad (22)$$

$$I_{z+1} = Y_{z+1}(\max(\text{HE}_z)) w_{z_2} c_z, \quad (23)$$

where $z = 1, \dots, M - 1$.

In the above equation, Y_z and Y_{z+1} are the mapping functions for the two electrodes forming an analysis band, w_{z_1} and w_{z_2} are the weights, $\max(\text{HE}_z)$ is the largest Hilbert envelope value that was computed since the previous mapping operation for the analysis band z , and c_z is the carrier.

2.3. The Sinusoid Extraction Strategy (SineEx). The new sinusoid extraction (SineEx) strategy is based on the general structure of the SpecRes strategy but incorporates a robust method for estimating spectral components of audio signals with high accuracy. A block diagram illustrating SineEx is shown in Figure 4.

The front-end, the filterbank, the envelope detector, and the mapping are identical to those used in SpecRes strategy. However, in contrast to the spectral-peak-picking algorithm performed by SpecRes, a frequency estimator that uses an iterative analysis/synthesis algorithm selects the most important spectral components in a given frame of the audio signal. The analysis/synthesis algorithm models the frequency spectrum as a sum of sinusoids. Only the perceptually most important sinusoids are selected using a psychoacoustic masking model.

The analysis/synthesis loop first defines a source model to represent the audio signal. The model's parameters are adjusted to best match the audio signal. Because of the few number of analysis bands in the Harmony system ($N = 15$), only a small number of parameters of the source model can be estimated. Therefore, the most complex task in SineEx is determining the few parameters that describe the input signal. The selection of the most relevant components is controlled by a psychoacoustic masking model in the analysis/synthesis loop. The model simulates the effect of simultaneous masking that occurs at the level of the basilar membrane in normal hearing.

The model estimates which sinusoids are masked the least to drive the stimulation to the electrodes. The idea behind this model is to deliver only those signal components that are most clearly perceived by normal-hearing listeners to the cochlear implant. A psychoacoustic masking model used to control the selection of sinusoids in an analysis/synthesis loop has been shown to provide improved sound quality with respect to other methods in normal hearing [21].

For example, other applications of this technique, where stimulation was restricted to the number of physical electrodes, demonstrated that the interaction between channels could be reduced by selecting fewer electrodes for stimulation. Therefore, because current steering will allow stimulation of significantly more cochlear sites compared to nonsimultaneous stimulation strategies, the masking model may contribute even further to the reduction of channel interaction and therefore improve sound perception. In [22] a psychoacoustic masking model was also used to select the perceptually most important components for cochlear implants. One aspect assumed in [22] was that the negative effects of channel interaction on speech understanding could be reduced by selecting less bands for stimulation.

The parameters extracted for the source model are then used by the frequency weight map and the carrier synthesis to code place pitch through current steering and to code temporal pitch by modulating the Hilbert envelopes, just as in SpecRes. Note that a high-accuracy estimation of frequency components is required in order to take advantage of the potential frequency resolution that can be delivered using current steering.

For parametric representations of sound signals, as in SineEx, the definition of the source model, the method used to select the model's parameters, and the accuracy in the extraction of these parameters play a very important role in increasing sound perception performance [21]. The next sections present the source model and the algorithm

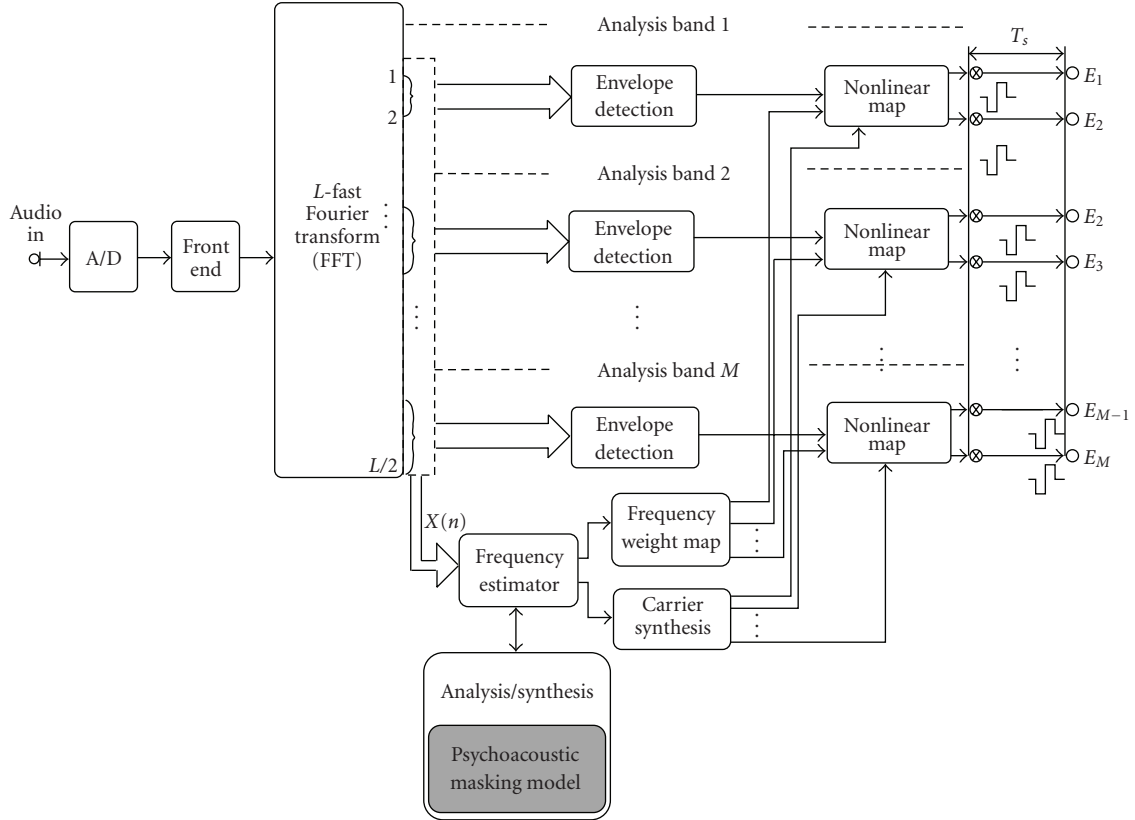


FIGURE 4: Block diagram illustrating SineEx.

used to estimate the model's parameters based on an analysis/synthesis procedure.

2.3.1. Source Model. Advanced models of the audio source are advantageous for modeling audio signals with the fewest number of parameters. To develop the SineEx strategy, the source model had to be related to the current-steering capabilities of the implant. In SineEx, the source model decomposes the input signal into sinusoidal components. A source model based on sinusoids provides an accurate estimation of the spectral components that can be delivered through current steering. Individual sinusoids are described by their frequencies, amplitudes, and phases. The incoming sound $x(l)$ is modeled as a summation of N sinusoids as follows:

$$x(l) \approx \tilde{x}(l) = \sum_{i=1}^N c_i e^{j(2\pi m_i l/L + \phi_i)}, \quad (24)$$

where $x(l)$ is the input signal, $\tilde{x}(l)$ is the model of the signal, c_i is the amplitude, m_i is the frequency, and ϕ_i is the phase of the i th sinusoid.

2.3.2. Parameter Estimation for the Source Model. The parameters of individual sinusoids are extracted iteratively in an analysis/synthesis loop [23]. The algorithm uses a dictionary of complex exponentials $s_m(l) = e^{j2\pi m l/L(l-(L-1)/2)}$ ($l = 1, \dots, L$) with P elements ($m = 1, \dots, P$) [24] as source

model. The analysis/synthesis loop is started with the windowed segment of the input signal $x(l)$ as first residual $r_1(l)$:

$$r_1(l) = x(l)w(l), \quad l = 0, \dots, L-1, \quad (25)$$

where $x(l)$ is the input audio signal and $w(l)$ is the same blackman-hanning window as in SpecRes (3).

The window $w(l)$ is also applied to the dictionary elements:

$$g_m(l) = w(l)s_m(l) = w(l)e^{j2\pi m l/L(l-(L-1)/2)}. \quad (26)$$

It is assumed that $g_m(l)$ has unity norm, that is, $\|g_m(l)\| = 1$ for $l = 0, \dots, L-1$.

For the next stage, since $x(l)$ and $r_1(l)$ are real values, the next residual can be calculated as follows:

$$r_{i+1}(l) = r_i(l) - c_i g_{m_i}(l) - c_i^* g_{m_i}^*(l). \quad (27)$$

The estimation consists of determining the optimal element $g_{m_i}(l)$ and a corresponding weight c_i that minimizes the norm of the residual:

$$\min \|r_{i+1}(l)\|. \quad (28)$$

For a given m the optimal real and imaginary component of c_i ($c_i = a_i + jb_i$) according to (28) can be found by setting the partial derivatives of $\|r_{i+1}(l)\|$ with respect to a_i and b_i to 0:

$$\begin{aligned} \frac{\Delta r_{i+1}(l)}{\Delta a_i} &= 0, \\ \frac{\Delta r_{i+1}(l)}{\Delta b_i} &= 0. \end{aligned} \quad (29)$$

This leads the following equation system:

$$\begin{pmatrix} \sum_l \operatorname{Re}\{g_m(l)\} \operatorname{Re}\{g_m(l)\} & \sum_l \operatorname{Re}\{g_m(l)\} \operatorname{Im}\{g_m(l)\} \\ \sum_l \operatorname{Re}\{g_m(l)\} \operatorname{Im}\{g_m(l)\} & \sum_l \operatorname{Im}\{g_m(l)\} \operatorname{Im}\{g_m(l)\} \end{pmatrix} \times \begin{pmatrix} 2a \\ -2b \end{pmatrix} = \begin{pmatrix} \sum_l \operatorname{Re}\{g_m(l)\} r_i(l) \\ \sum_l \operatorname{Re}\{g_m(l)\} r_i(l) \end{pmatrix}. \quad (30)$$

As the window used is symmetric $w(l) = w(-l)$, $\operatorname{Re}\{g_m(l)\}$, and $\operatorname{Im}\{g_m(l)\}$ become orthogonal, that is, the scalar product between them is 0:

$$\sum_l \operatorname{Re}\{g_m(l)\} \operatorname{Im}\{g_m(l)\} = 0, \quad \forall l, \quad (31)$$

and the previous Equations can be simplified as follows:

$$\begin{aligned} a &= \frac{1}{2} \frac{\sum_l \operatorname{Re}\{g_m(l)\} r_i(l)}{\sum_l \operatorname{Re}\{g_m(l)\} \operatorname{Re}\{g_m(l)\}}, \\ b &= \frac{-1}{2} \frac{\sum_l \operatorname{Im}\{g_m(l)\} r_i(l)}{\sum_l \operatorname{Im}\{g_m(l)\} \operatorname{Im}\{g_m(l)\}}. \end{aligned} \quad (32)$$

The element g_{m_i} of the dictionary selected for the i th iteration is obtained by minimizing $\|r_{i+1}(l)\|$. This is equivalent to maximizing c_i as can be observed in (27). Therefore, the element selected g_{m_i} corresponds to the one having the largest scalar product with the signal $r_i(l)$ for $l = 0, \dots, L-1$.

Finally, the amplitude c_i , frequency f_{\max_i} , and phase ϕ_i for the i th sinusoid are

$$\begin{aligned} c_i &= \sqrt{a_i^2 + b_i^2}, \\ f_{\max_i} &= n_{\max_i} \frac{2\pi}{L}, \\ \phi_i &= \arctan\left(\frac{b_i}{a_i}\right). \end{aligned} \quad (33)$$

2.3.3. Analysis/Synthesis Loop Implementation. The analysis/synthesis algorithm can be efficiently implemented in the frequency domain [25]. The frequency domain implementation was used to incorporate the algorithm into the Harmony system. A block diagram illustrating the implementation is presented in Figure 5.

The iterative procedure uses as input the FFT spectrum of an audio signal $X(n)$. The magnitude spectrum $|X(n)|$

then is calculated. It is assumed that in the i th iteration $i-1$ sinusoids already have been extracted and a signal $S_{i-1}(n)$ containing all sinusoids has been synthesized. The magnitude spectrum $|S_{i-1}(n)|$ is calculated.

The synthesized spectrum is subtracted from the original spectrum and then weighted by the magnitude masking threshold $I_{w_{i-1}}(n)$ caused by the sinusoids already synthesized. The detection of the maximum ratio $E_{n_{\max}}$ is calculated as follows:

$$\begin{aligned} E_{n_{\max_i}} &= \max\left(0, \frac{|X(n)| - |S_{i-1}(n)|}{|I_{w_{i-1}}(n)|}\right), \quad n = 0, \dots, L-1, \\ n_{\max_i} &= \arg \max\left(0, \frac{|X(n)| - |S_{i-1}(n)|}{|I_{w_{i-1}}(n)|}\right), \quad n = 0, \dots, L-1, \end{aligned} \quad (34)$$

where $I_{w_i}(n)$ is the psychoacoustic masking model at the i th iteration of the analysis/synthesis loop. The frequency n_{\max_i} is used as a coarse frequency estimate of each sinusoid. Its accuracy corresponds to the FFT frequency resolution.

The spectral resolution of the frequency estimated is improved using a high accuracy parameter estimation on the neighboring frequencies of n_{\max_i} . The high accuracy estimator implements (30) iteratively in the frequency domain. The algorithm takes first, the positive part of the spectrum $X(n)$, that is, the analytical signal of $x(l)$. As the algorithm is implemented in the frequency domain, the dictionary elements $g_m(l)$ are transformed into the frequency domain. If $G_0(n)$ denotes the Fast Fourier Transform of $g_0(n) = w(l)$, the frequency domain representation of the other dictionary elements can be derived by simple displacement of the frequency axis $G_m(n) = G_0(n-m)$. For this reason, $G_0(n)$ is also referred to as ‘‘prototype.’’ Note that as the window $w(l)$ is known (3), the frequency resolution of the prototype can be increased just by increasing the length of the FFT used to transform $g_0(n)$. Because most of the energy of the prototype $G_0(l)$ concentrates in a small number of samples around the frequency $n=0$, a small section of the prototype is stored. By reducing the length of the prototype, the complexity of the algorithm drops significantly in comparison to the time domain implementation presented in Section 2.3.2.

The solution to (30) is solved iteratively as follows. In the first iteration ($r=1$), the prototype is centered on the $n_{\max_{i,r}} = n_{\max_i}$ coarse frequency. A displacement variable δ_r is set to $1/2r$, where r indicates the iteration index. The correlation is calculated at $n_{\max_{i,r}} - \delta_r$, $n_{\max_{i,r}}$, and $n_{\max_{i,r}} + \delta_r$. The position leading to maximum correlation at these three locations is denoted by $n_{\max_{i,r+1}}$. For the next iteration ($r+1$) the value δ_{r+1} is halved ($\delta_{r+1} = 1/2(r+1)$) and the prototype is centered on $n_{\max_{i,r+1}}$. The correlation is calculated at $n_{\max_{i,r+1}} - \delta_{r+1}$, $n_{\max_{i,r+1}}$, and $n_{\max_{i,r+1}} + \delta_{r+1}$ and the maximum correlation is picked up. This procedure is repeated several times, and the final iteration gives the estimated frequency denoted by $n_{\max_i}^*$.

2.3.4. Psychoacoustic Masking Model. The analysis/synthesis loop of [25] is extended by a simple psychoacoustic model for the selection of the most relevant sinusoids. The model

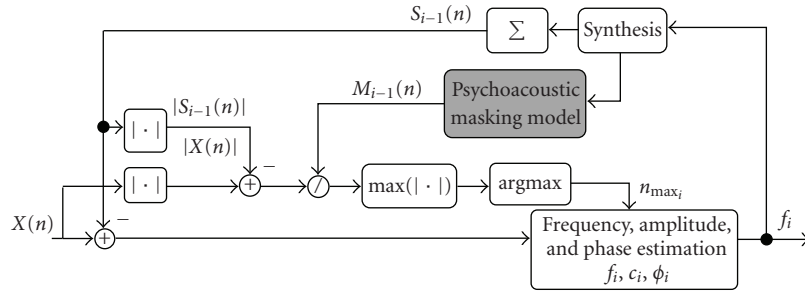


FIGURE 5: Frequency domain implementation of the analysis/synthesis loop including a psychoacoustic masking model for extraction and parameter estimation of individual sinusoids.

is a simplified implementation of the masking model used in [22]. The effect of masking is modeled using a spreading masking function $L(z)$. This function has been modeled using a triangular shape with left slope s_l , right slope s_r , and peak offset a_v , as follows:

$$L_i(z) = \begin{cases} \text{HE}_{\text{dB}_i} - a_v - s_l \cdot (z_i - z), & z < z_i, \\ \text{HE}_{\text{dB}_i} - a_v - s_r \cdot (z - z_i), & z \geq z_i. \end{cases} \quad (35)$$

The amplitude of the spreading function is derived from the Hilbert Envelope in decibels $\text{HE}_{\text{dB}_i} = 20 \log_{10}(\text{HE}(z))$ associated to the analysis band containing the sinusoid extracted at the iteration i of the analysis/synthesis loop. The sound intensity $I_i(z)$ is calculated as

$$I_i(z) = 10^{L_i(z)/20}, \quad z = 1, \dots, M. \quad (36)$$

The superposition of thresholds is simplified as a linear addition of thresholds (37) in order to reduce the number of calculations

$$I_{T_i}(z) = \sum_{k=0}^i I_k(z), \quad z = 1, \dots, M. \quad (37)$$

The spreading function has been defined in the nonlinear frequency domain, that is, in the analysis band domain z . As the sinusoids are extracted in the uniformly spaced frequency domain of the L -FFT, the masking threshold must be unwarped from the analysis band domain into the uniformly spaced frequency domain. The unwarping is accomplished by linearly interpolating the spreading function without considering that the two scales have different energy densities as follows:

$$I_{w_i}(n) = I_{T_i}(z - 1) + (n - n_{\text{center}}(z - 1)) \times \frac{I_{T_i}(z) - I_{T_i}(z - 1)}{n_{\text{center}}(z) - n_{\text{center}}(z - 1)}, \quad (38)$$

$$z = 1, \dots, M, \quad i = 1, \dots, N,$$

where M denotes the number of analysis bands, N gives the number of sinusoids selected, and $n_{\text{center}}(z)$ is the center frequency for the analysis band z in bins (see Table 1):

$$n_{\text{center}}(z) = \frac{n_{\text{start}_{z+1}} - n_{\text{start}_z}}{2}. \quad (39)$$

In normal hearing, simultaneous masking occurs at the level of the basilar membrane. The parameters that define the spread of masking can be estimated empirically with normal hearing listeners. Simultaneous masking effects can be used in cochlear implant processing to reduce the amount of data that is sent through the electrode nerve interface [22]. However, because simultaneous masking data is not readily available from cochlear implant users, the data from normal hearing listeners were incorporated into SineEx. The choice of the parameters that define the spread of masking require more investigation, and probably should be adapted in the future based upon the electrical spread of masking for each individual.

The parameters that define the spreading function were configured to match the masking effect produced by tonal components [26, 27] in normal hearing listeners, since the maskers are the sinusoids extracted by the analysis/synthesis loop. The left slope was set to $s_l = 40$ dB/band, the right slope to $s_r = 30$ dB/band, and the attenuation to $a_v = 15$ dB.

SineEx is an N-of-M strategy because only those bands containing a sinusoid are selected for stimulation. The analysis/synthesis loop chooses N sinusoids iteratively in order of their “significance.” The number of virtual channels activated in a stimulation cycle is controlled by increasing or decreasing the number of extracted sinusoids N . It should be noted that the sinusoids are extracted over the entire spectrum and are not restricted to each analysis band as in SpecRes. Therefore, in some cases, more than one sinusoid may be assigned to the same analysis band and electrode pair. In those situations, only the most significant sinusoid is selected for stimulation because only one virtual channel can be created in each analysis band during one stimulation cycle.

2.4. Objective Analysis: HiRes, SpecRes, and SineEx. Objective experiments have been performed to test the three strategies: HiRes, SpecRes, and SineEx. The strategies have been evaluated analyzing the stimulation patterns produced by each strategy for synthetic and natural signals. The stimulation patterns represent the current level applied to each location l_{exc} along the electrode array in each time interval or frame h . The total number of locations L_{sect} is set to 16000 in

this analysis. The number of locations associated with each electrode n_{loc} is

$$n_{loc} = \frac{L_{sect}}{M}, \quad (40)$$

M indicates the number of electrodes. The location of each electrode is $l_{el_z} = (z - 1)n_{loc}$, $z = 1, \dots, M$. The stimulation pattern is obtained as follows. First the total current produced by two electrodes at the frame h is calculated

$$Y_{T_z}(h) = Y_z(h) + Y_{z+1}(h), \quad z = 1, \dots, M - 1, \quad (41)$$

where $Y_z(h)$ and $Y_{z+1}(h)$ denote the current applied to the first and second electrode pairs forming an analysis channel (22). Then, the location of excitation is obtained as follows:

$$\tilde{l}_{exc} = l_{el_z} \frac{Y_z(h)}{Y_{T_z}(h)} + l_{el_{z+1}} \frac{Y_{z+1}(h)}{Y_{T_z}(h)}, \quad (42)$$

where l_{el_z} and $l_{el_{z+1}}$ denote the location of the first and the second electrode in a pair forming an analysis channel. Note that for sequential nonsimultaneous stimulation strategies $Y_{z+1}(h)$ is set to 0 and therefore, the location of excitation \tilde{l}_{exc} coincides with the location of the electrode l_{el_z} . For sequential stimulation strategies $z = 1, \dots, M$. Finally, \tilde{l}_{exc} is rounded to the first integer, that is, $l_{exc} = \lfloor \tilde{l}_{exc} \rfloor$ and the excitation pattern S_{exc} at frame h and location l_{exc} is expressed as

$$S_{exc}(l_{exc}, h) = Y_{T_z}(h). \quad (43)$$

The first signal used to analyze the strategies was a sweep tone of constant amplitude and varying frequency from 300 Hz to 8700 Hz during 1 second. The spectrogram of this signal is shown in Figure 6(a). The sweep tone has been processed with HiRes, SpecRes, and SineEx and the stimulation patterns produced by each strategy are presented in Figures 6(b), 6(c), and 6(d), respectively.

In HiRes, the location of excitation always coincides with the position of the electrodes. However, in SpecRes and SineEx, the location of excitation can be steered between two electrodes using simultaneous stimulation.

Moreover, it should be remarked that the frequency estimation performed by SineEx is more distinct than with SpecRes. It can be observed from Figure 6(d) that during the whole signal almost only two neighboring electrodes (1 virtual channel) are being selected for stimulation. This fact causes that only one virtual channel is used to represent the unique frequency presented at the input. In the case of SpecRes (Figure 6(c)), it is shown that more than one virtual channel is generated to represent a unique sinusoid in the input signal. This is caused by the simple modeling approach performed by SpecRes to represent sinusoids. This fact should cause smearing in pitch perception because different virtual channels are combined to represent a unique frequency. White Gaussian noise was added to the same sweep signal with at total SNR of 10 dB. The stimulation patterns obtained in noise are presented in Figures 7(b), 7(c), and 7(d). Figure 7(b) shows the stimulation pattern

generated by HiRes for the noisy sweep tone. It can be observed that HiRes mixes both, the noise and the sweep tone, in terms of place of excitation, as the location of excitation coincides with the electrodes. This fact should cause difficulties to separate the tone from the noise. Figures 7(c) and 7(d) present the stimulation patterns when processing the noisy sweep tone with SpecRes and SineEx, respectively. It can be observed that when noise is added, SpecRes stimulates more times the electrodes than SineEx. As white Gaussian noise is added, frequency components are distributed along the whole frequency domain. SpecRes selects peaks of the spectrum without performing any model assumption of the input signal, therefore noise components are treated as if they were pure tone components. This fact should lead to the perception of tonal signal when in reality the signal is noisy. SineEx, however, is able to estimate and track the frequency of the sweep tone as it matches the sinusoidal model. In contrast, the added white Gaussian noise does not match the sinusoidal model and those parts of the spectrum containing noise components are not selected for stimulation. On the one hand, this test presents the potential robustness of SineEx in noise situations to represent tonal or sine-like components. On the other hand, the experiment shows the limitations of SineEx to model noisy-like signals like some consonants.

A natural speech signal consisting of a speech token, where “asa” is uttered by a male voice, has also been processed with HiRes, SineEx, and SpecRes. Figures 8(b), 8(c), and 8(d) present the stimulation patterns obtained for each strategy.

In HiRes, the location of excitation coincides with the position of the electrodes. This fact causes a limitation to code accurately formant frequencies because the spectral resolution with HiRes is limited by the number of implanted electrodes. It is known that formants play a key role in speech recognition. The poor representation of formants with HiRes can be observed comparing the stimulation pattern generated by HiRes (Figure 8(b)) and the spectrogram presented in Figure 8(a). Using SpecRes, the formants can be represented with improved spectral resolution compared to HiRes as the location of excitation can be varied between two electrodes (Figure 8(c)). However, the lower accuracy of the method used by SpecRes to extract the most meaningful frequencies, based on a peak detector, makes the formants less distinguishable than with SineEx (Figure 8(d)). SpecRes selects frequency components without making a model assumption of the incoming sound; therefore noise and frequency components are mixed causing possible confusions between them. In SineEx, both “a” vowels can be properly represented as a sum of sinusoids. However, the consonant “s” which is a noise-like component is not properly represented using just a sinusoidal model.

SineEx and SpecRes combine the current steering technique with a method to improve temporal coding, by adding the temporal structure of the frequency extracted in each analysis band. This temporal enhancement was incorporated to SineEx and SpecRes in order to compensate for the lower temporal resolution of the 256-FFT used by these strategies in comparison to the IIR filterbank used by Hires. For this

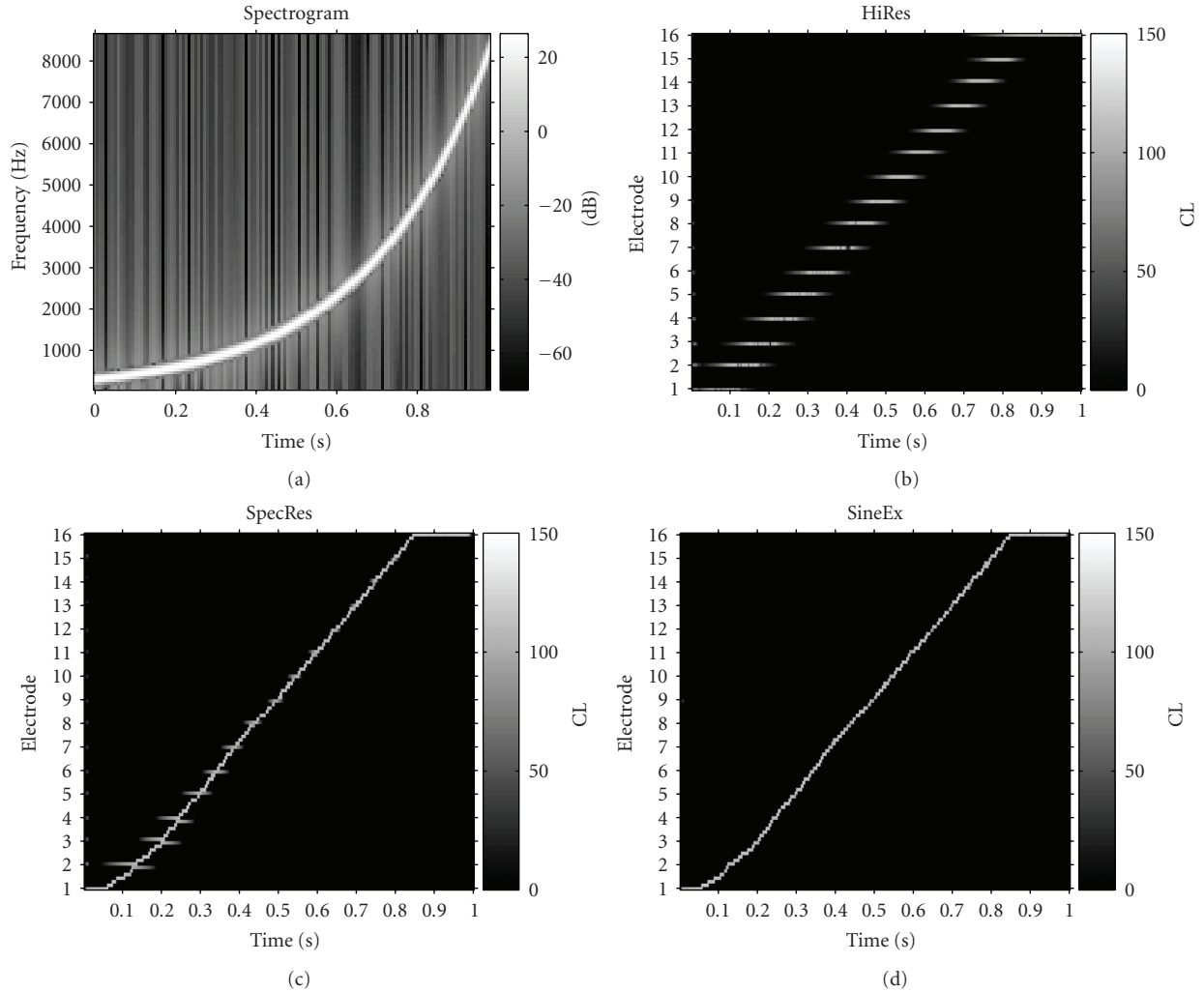


FIGURE 6: Stimulation patterns obtained with (b) HiRes, (c) SpecRes, and (d) SineEx in quiet when the input signal is a sweep tone of 1 millisecond of constant amplitude (70 dB) and frequency varying from 300 Hz until 8700 kHz shown in (a). The horizontal axis represents time in seconds, and the vertical axis represents the electrode location. The level applied in current level (CL) is coded with the colors given in the color bars. The location of excitation is obtained as presented in Section 2.4.

reason, we assume that a hypothetical improvement of pitch perception provided by SineEx or SpecRes might be caused by the current steering technique rather than by the temporal enhancement technique.

With one final comment from the objective analysis, as SineEx generally selects less frequencies than SpecRes, this strategy has the potential to reduce interaction between channels and significantly reduce power consumption in comparison to SpecRes. This feature can be confirmed by an experiment that involves counting the number of channels being stimulated by HiRes, SpecRes, and SineEx during the presentation of 50 sentences from a standardized sentence test [28]. The CSR was set to 2320 stimulations/second for all three strategies. Table 2 presents the total number of channels stimulated by each strategy.

As it can be observed from Table 2, the number of stimulations by SpecRes doubles the number of stimulations performed by HiRes. However, as SpecRes divides the current

TABLE 2: Number of stimulations for 50 sentences of the HSM sentence test [28] with HiRes, SpecRes, and SineEx.

	HiRes	SpecRes	SineEx
Number of stimulations	464,895	1,087,790	536,878

between two electrodes, both strategies would lead to a similar power consumption. In SineEx however, less channels are stimulated and this could lead to an improvement in power consumption.

3. Study Design

HiRes, SpecRes, and SineEx were incorporated into the research platform Speech Processor Application Framework (SPAF) designed by Advanced Bionics. Using this Platform,

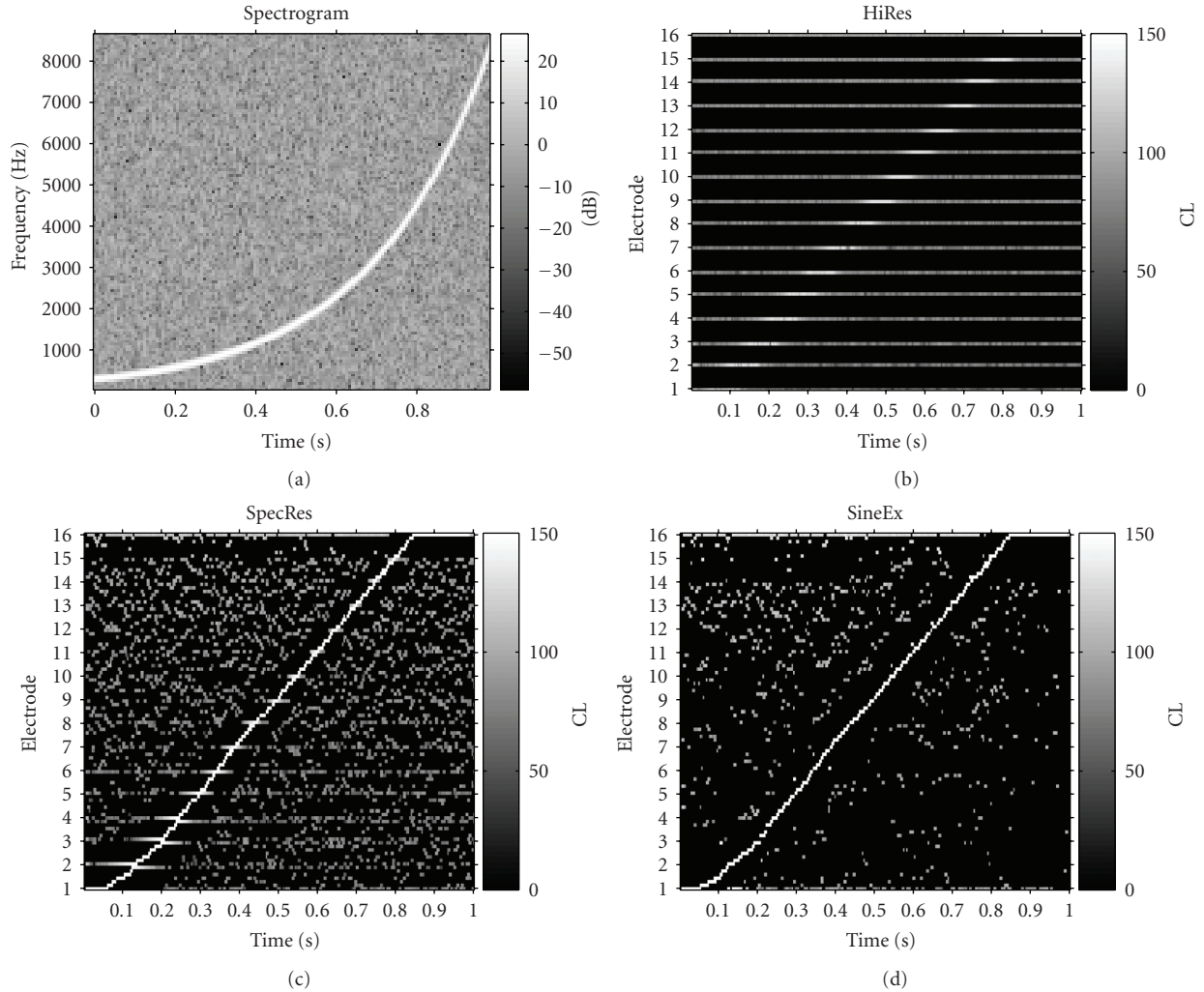


FIGURE 7: Stimulation patterns obtained with (b) HiRes, (c) SpecRes, and (d) SineEx in noise (SNR = 10 dB) when the input signal is a sweep tone of 1 millisecond of constant amplitude (70 dB) and frequency varying from 300 Hz until 8700 kHz added with white Gaussian noise (SNR = 10 dB) shown in (a). The horizontal axis represents time in seconds, and the vertical axis represents the electrode location. The level applied in current level (CL) is coded by the colors given in the color bars. The location of excitation is obtained as presented in Section 2.4.

a chronic trial was conducted at the hearing center of the Medical University of Hannover with 9 Harmony implant users. The SPAF and the three strategies were implemented in the Advanced Bionics bodyworn Platinum series processor (PSP). The aim of the study was to further investigate the benefits of virtual channels or current steering after a familiarization period. Subjects were tested with all three strategies (HiRes, SpecRes, and SineEx). The study was divided into two symmetrical phases. In the first phase, each strategy was given to each study participant during four weeks and then evaluated. The order in which the strategies were given to each patient was randomized. In the second stage of the study, the strategies were given in reverse order with respect to the first phase. Again after 4 weeks each strategy was evaluated. Therefore, the total length of the study for each subject was 24 weeks. The study participants were selected because of their good hearing abilities in quiet and noisy environments and for their motivation to listen to

music with their own clinical program. The participants were not informed about the strategy they were using.

Frequency Discrimination. The aim of this task was to determine if current steering strategies could deliver better pitch perception than classical sequential stimulation strategies. Frequency discrimination was evaluated with a three alternative-forced-choice task (3AFC) using an adaptive method test [29]. Audio signals were delivered to the cochlear implant recipient via the direct audio input of the PSP. Stimuli were generated and controlled by the Psycho-Acoustic Test Suite (PACTS) software developed by Advanced Bionics. The stimuli consisted of 500 milliseconds pure tones sampled at 17.4 kHz and ramped on and off over 10 milliseconds with a raised cosine. The reference frequencies were 1280 Hz and 2904 Hz. Each subject was presented with three stimuli in each trial. Two stimuli consisted of a tone

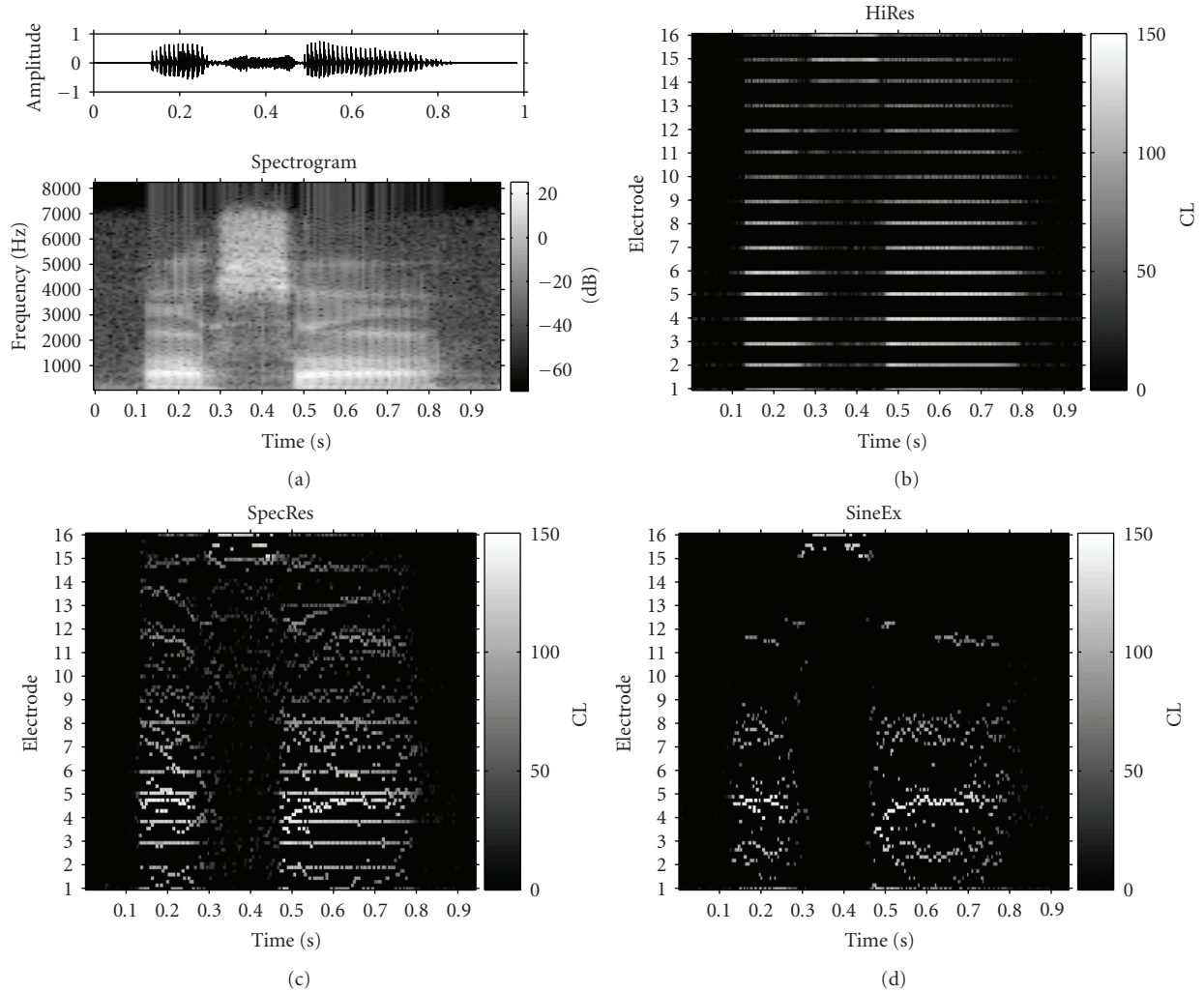


FIGURE 8: (a) Speech token “asa” uttered by a male voice and its spectrogram. (b) Stimulation pattern obtained with HiRes. (c) Stimulation pattern obtained with SpecRes. (d) Stimulation pattern obtained with SineEx. The horizontal axis represents time in seconds, and the vertical axis represents the electrode location. The level applied in current level (CL) is coded by the colors given in the color bars. The location of excitation is obtained by linearly interpolating the electrical amplitude applied to the pairs of simultaneous stimulated electrodes.

burst at the reference frequency. This frequency was fixed during the whole run. The third stimulus consisted of a tone burst at two times the reference frequency (probe frequency). The presentation order of the stimulus was randomized in the three intervals. The subject was asked to identify the interval containing the stimulus that was higher in pitch. After two consecutive correct answers, the frequency of the probe stimulus was decreased by a factor of $2^{1/12}$. After each incorrect answer, the frequency of the probe stimulus was increased by two times this factor, leading to an asymptotic average of 71% correct responses [29]. The procedure was continued until 8 reversals were obtained and the mean of the probe frequency of the last four reversals was taken as the result for that particular run. This result is termed the frequency difference limen (FDL). Intensity was roved by randomly varying the electrical output gain from 85% to 110% of the dynamic range, to minimize loudness cues. The experiment was performed twice for each subject and the mean value of both runs was calculated.

Speech Recognition Tests. Speech recognition was evaluated using the HSM sentence test [28]. The HSM test was administered in quiet, in noise, and with background speech interference (competing talker).

The aim of the speech-in-noise condition was to evaluate if current-steering strategies could improve speech intelligibility in noisy situations. For the noise condition, telephone noise was added to the HSM test according to the Committee Communication International Telephone and Telegram recommendation 227 [30]. The signal-to-noise-ratio was 10 dB. The aim of the speech-in-competing condition was to evaluate if current steering strategies could provide better speech intelligibility in the presence of multiple talkers. For the evaluation of speech recognition with background speech interference, a second German voice was added to the HSM sentence test. This was accomplished by mixing the HSM test with the Oldenburger sentence test (OLSA) [31]. Every word of the HSM sentence test was overlapped in time by at least one word of the OLSA test. The

signal-interference ratio was 5 dB. The patients were asked to repeat only those sentences corresponding to the HSM sentence test and the number of correct words was counted.

For each condition (quiet, noise, and competing talker) 2 lists of 20 sentences were presented in each stage of the study. The subjects had to repeat each sentence, and results were based on the number of correct words repeated correctly. All tests were conducted by connecting a CD player directly to the audio input of the speech processor.

Music and Speech Subjective Appreciation Tests. Subjective sound perception with each strategy was evaluated using questionnaires that assessed the overall benefits of the implant in daily life. The questionnaires supplemented the data available from conventional tests of speech perception [32].

The questionnaires asked subjects to rate music and speech quality [33]. At each stage of the study, the questionnaire was completed by the patient such that for each strategy the same questionnaire was filled out two times.

The music questionnaire asked subjects to rate the pleasantness, distinctness, naturalness, and overall perception of music on a scale from 0 (extremely unpleasant, extremely indistinct, extremely unnatural, extremely bad) to 10 (extremely pleasant, extremely distinct, extremely natural, extremely good).

For the speech questionnaire asked subjects to rate different characteristics of speech on a 10-point scale. Characteristics included background interference; naturalness of female voices, male voices, own voice; clarity of speech; pleasantness of speech; overall quality of speech. Listeners were provided with definitions of each dimension for each scale.

3.1. Subjects. All subjects had clinical experience with the HiRes strategy and were users of the Harmony (HiRes 90k or CII) implant. This strategy can only be configured in monopolar stimulation mode. Demographic information for all test subjects is presented in Table 3. P7 completed only the first phase of the study protocol. For all strategies, the stimulation rate was derived from the HiRes clinical program and was kept constant throughout the conditions. Threshold and most comfortable levels also were kept constant. Only global modifications of these levels were allowed to accommodate loudness requests of the subjects, meaning that the THL or MCL levels were changed by the same amount for all the electrodes.

4. Results

All subjects reported that speech experienced using SpecRes and SineEx was understandable immediately. However, the sound perceived with the new strategies was significantly different from HiRes for some users. For example, SineEx was immediately described as brighter than the other strategies. Despite the sound quality differences, all subjects were willing to take part in the chronic phase of the study even though they were not allowed to change the strategy

during the study period. All subjects immediately reported that speech experienced using SpecRes and SineEx was understandable. For some users, the sound perceived with the new strategies was, however, significantly different from HiRes. For example, the sound with SineEx was immediately described as brighter than with any other strategy.

Frequency Discrimination Results. Frequency discrimination results are presented for the two reference frequencies (1280 Hz and 2904 Hz) in Figures 9(a) and 9(b). In mean value, all current steering strategies obtained an improvement in frequency discrimination with respect to HiRes. SpecRes produced slightly better frequency discrimination than SineEx for the 1280 Hz reference frequency.

Results were subject to the paired *t*-test significance test. No significant difference was found between HiRes, SpecRes, and SineEx for the two reference frequencies due to the large inter- and intrasubject variability. Particularly, the results were dominated by the large variability observed in P2 for both reference frequencies.

4.1. Speech Intelligibility Results

Speech Intelligibility in Quiet. Figure 10(a) presents the averaged scores for each subject for the HSM sentence test in quiet. All subjects, except P2, scored 90% or higher for all three strategies, thus demonstrating that they were good performers with all strategies.

Speech Intelligibility in Noise. Figure 10(b) presents the averaged results for each subject for the HSM sentence test in noise (SNR = 10 dB). The mean results show that, in general, SpecRes produced the highest scores. Three patients out of 9 (P2, P7, and P8) attained better speech recognition scores with HiRes than with SpecRes.

Speech Intelligibility with Competing Talker. The mean scores for each patient using the HSM sentence test with competing talker are presented in Figure 10(c). SpecRes produced the highest word recognition scores in this condition. P7 was not able to understand speech in this condition with any strategy.

Although SineEx produced the best frequency discrimination scores, this strategy was not able to improve word recognition with competing talker in most of the patients. Only patient 4 obtained better speech understanding with SineEx.

All the results were subject to the paired samples *t*-test. No significant difference were found between HiRes, SpecRes, and SineEx.

Subjective Music Appreciation Questionnaire. Figure 11 presents the results for the items clarity, naturalness, pleasantness, and overall music perception. No significant differences were found between the three strategies (paired *t*-tests) and the overall perception of music was rated similarly for all three strategies. However, music was rated as more clear, natural and pleasant with SineEx compared to HiRes and SpecRes.

TABLE 3: Subject demographics for the current steering study.

Patient id	Age	Duration of deafness in years	Cause of deafness	Implant experience in years	Electrode type	Usual strategy
P1	54	0	Sudden hearing loss	3	HiRes90k	HiRes 1080 pps
P2	70	0	Unknown	8	HiRes90k	HiRes 1080 pps
P3	51	4	After heard operation	6	Clarion CII	HiRes 900 pps
P4	26	1	Unknown	7	Clarion CII	HiRes 500 pps
P5	43	0	Sudden hearing loss	7	Clarion CII	HiRes 900 pps
P6	62	0	Genetic	4	Clarion CII	HiRes 900 pps
P7	53	0	Unknown	9	HiRes90k	HiRes 1200 pps
P8	49	16	Unknown	5	Clarion CII	HiRes 900 pps
P9	60	0	Ototoxika	7	Clarion CII	HiRes 1200 pps

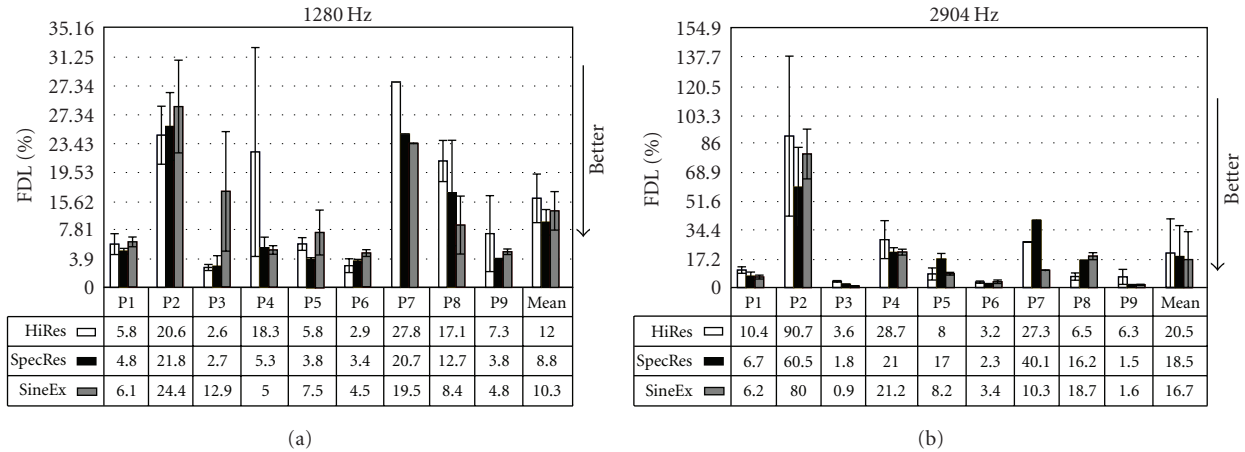


FIGURE 9: Frequency discrimination limen by subject (average and standard deviation) for the frequency discrimination test at different reference frequencies (a) 1280 Hz and (b) 2904 Hz with HiRes, SpecRes, and SineEx. The frequency discrimination limen is presented as percentage of the reference frequency.

Subjective Speech Appreciation Questionnaire. Figure 12 shows the results for the items speech quality with background interference, natural female voice, natural male voice, own voice, clarity, pleasantness, and overall speech quality.

There were no significant differences in the ratings of these speech characteristics. However, SpecRes produced better scores in perceived pleasantness, voice in background interference, and overall quality than HiRes and SineEx.

On the other hand, the naturalness of their own voice and male voice was rated higher with HiRes. Many patients reported that low frequency sounds were better perceived with HiRes, while the sound with the current steering strategies (especially with SineEx) was described as brighter. It is likely that the perception of female voices was rated

higher with SineEx, and male voices were rated higher with HiRes.

SineEx was rated higher for the perception of voice in background interference. These scores do not correlate with the speech intelligibility scores obtained by SineEx in the competing talker condition, which were lower than with SpecRes and equal to HiRes.

Clarity of voice was rated highly for all strategies. However SpecRes was rated slightly higher SineEx and HiRes.

5. Discussion

This study designed and evaluated new sound processing strategies that use current steering to improve the spectral and temporal information delivered to an Advanced Bionics

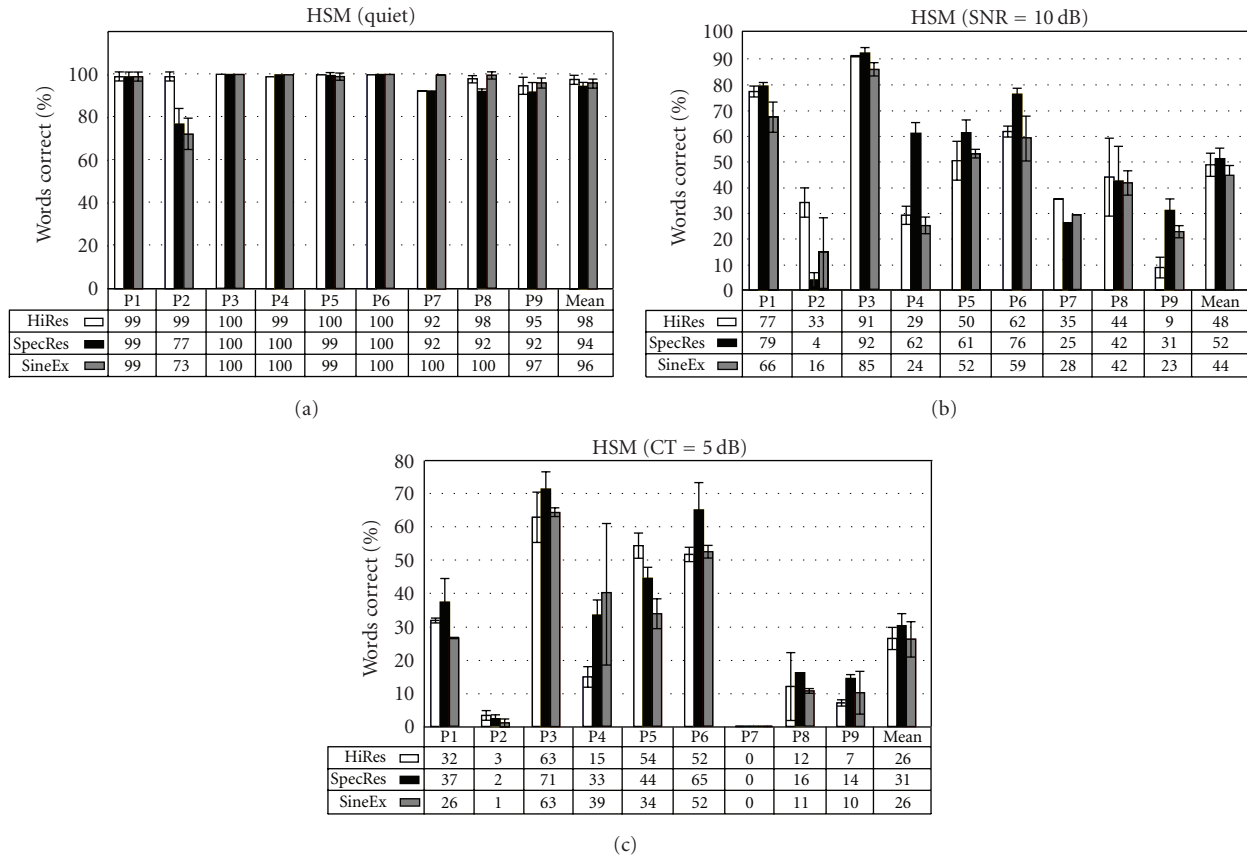


FIGURE 10: Percent of correct word by subject (average and standard deviation) for the HSM sentence test in (a) quiet, (b) with 10 dB SNR, and (c) 5 dB competing talker.

Harmony cochlear implant. Current steering stimulates pairs of electrodes simultaneously so that virtual channels are created intermediate to the physical electrodes. In SpecRes an FFT is used together with a spectral peak locator to extract the most dominant frequencies. In SineEx, the audio signal is modeled with sinusoids and the frequencies of those sinusoids are delivered to the implant using current steering and a psychoacoustic masking model.

SpecRes and SineEx are the first signal processing strategies implemented in a commercial device using the current steering technique. These strategies were evaluated in a chronic study comparing them to the standard HiRes strategy that does not use current steering. All patients were able to use the new strategies during a long-term test protocol. In the chronic study, frequency discrimination, speech intelligibility, and subjective ratings of music and speech were evaluated. Overall, the sound performance achieved with the three strategies evaluated in this study was similar and the results exhibited large inter- and intrasubject variability.

For frequency discrimination, there were nonsignificant improvements for the SineEx strategy over the HiRes strategy from 12% to 10.3% and from 20.5% to 16.7%, when the reference frequencies were 1280 Hz and 2904 Hz, respectively. These improvements are assumed to be the result

of using the current steering technique. A 1.8 percentage points improvement in frequency discrimination for SineEx compared to the SpecRes strategy was observed only for the 2904 Hz reference frequency. That improvement can be attributable to the robust method for modeling the audio signal with sinusoids and the use of a perceptual model to select the sinusoidal components.

There are several reasons why we could not demonstrate a significant improvement in frequency difference limen for current steering strategies with respect to sequential stimulation strategies. First, pure tones were used to estimate the FDL. Pure tones when presented to a cochlear implant may activate several bands (specially for HiRes and SpecRes). Therefore, the use of pure tones cannot provide an accurate assessment of specific locations along the electrode array as shown when stimulating the electrode array with just a pair of simultaneous stimulated pulses. Second, a large intra- and intersubject variability was observed. For example, study participant P2, who obtained worse speech intelligibility in quiet than the rest of patients, showed also very poor results in FDL with large variability. Third, it has to be remarked that all study participants used HiRes in daily life and therefore had more experience when hearing to sequential stimulation strategies than to current steering strategies.

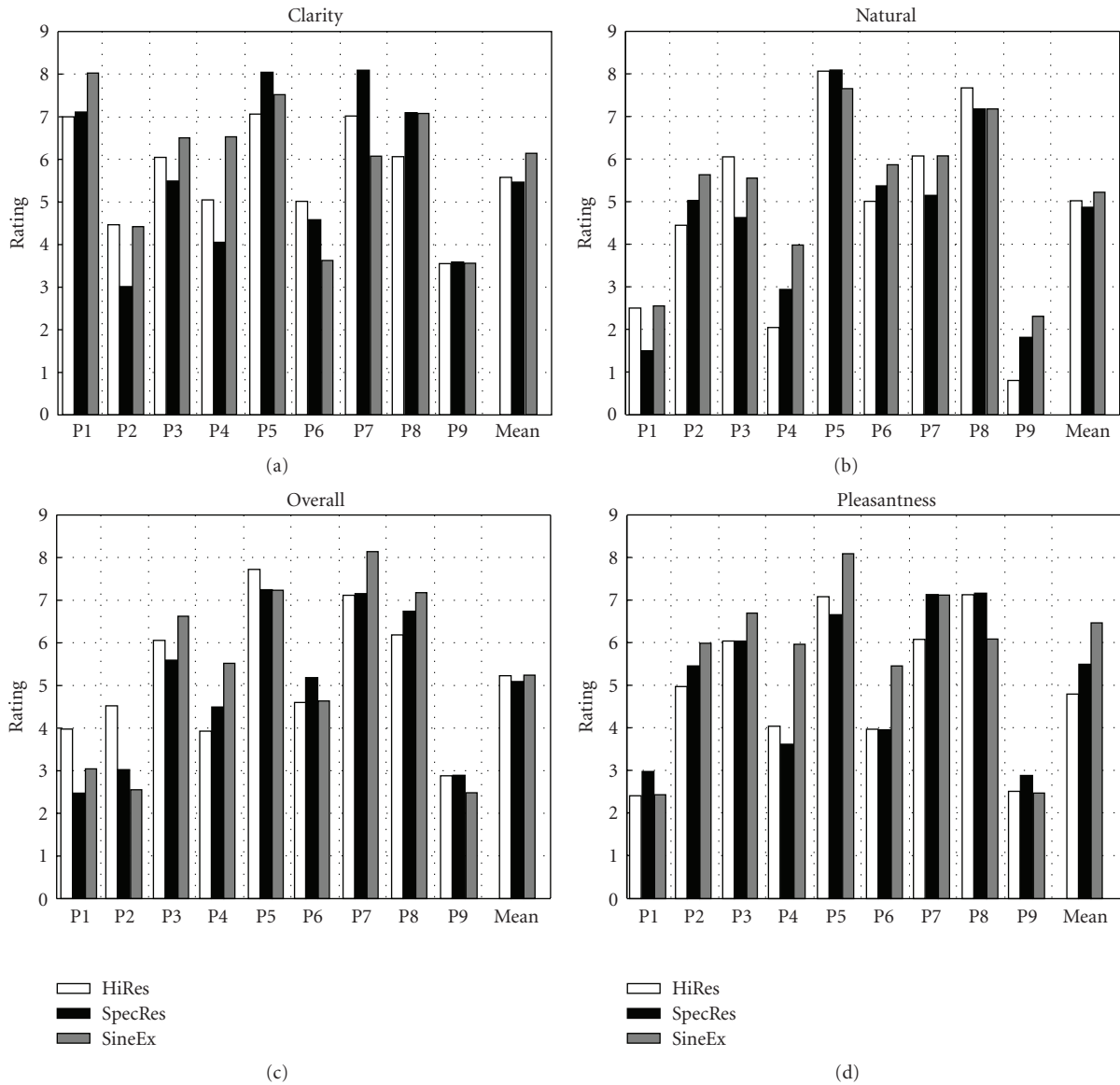


FIGURE 11: Results of the music quality questionnaire using HiRes, SpecRes, and SineEx. Different items were rated on a 10-point scale. The cochlear implant subjects were asked to rate different aspects of music perception. The items are (a) clarity of music, (b) natural music perception, (c) pleasantness hearing music, and (d) overall music perception.

The frequency difference limen results obtained in this study are in the same range as those observed in the literature [34] using a similar experiment. In our study we observed a trend towards mean sound frequency difference limens increasing with increasing frequency, this trend has also been reported in the literature for cochlear implants [34] as well as for normal hearing listeners [35].

Speech intelligibility was evaluated using the HSM sentence test. The results for HSM in noise showed a nonsignificant improvement of speech recognition for the SpecRes strategy against the HiRes and SineEx of 4% and 8%, respectively. For the HSM with competing talker (speech background interference) the SpecRes achieved a nonsignificant improvement of 5% with respect to both the HiRes and SineEx.

In SineEx, appropriate source modeling and the selection of the components that describe this model are obviously of great importance to sound and music quality. In addition, SineEx stimulates, on average, half as many electrodes as SpecRes. Thus SineEx has the potential to reduce power consumption in actual devices because fewer electrodes are stimulated per frame. Because a direct relationship exists between the number of channels selected and the amount of channel interaction, stimulating fewer channels can lead to a reduction of interaction between channels.

Current steering strategies stimulate pairs of electrodes simultaneously. Therefore, the interaction between channels and the spread of the electrical fields produced in the cochlea increase with respect to nonsimultaneous stimulation strategies. In order to limit channel interaction in

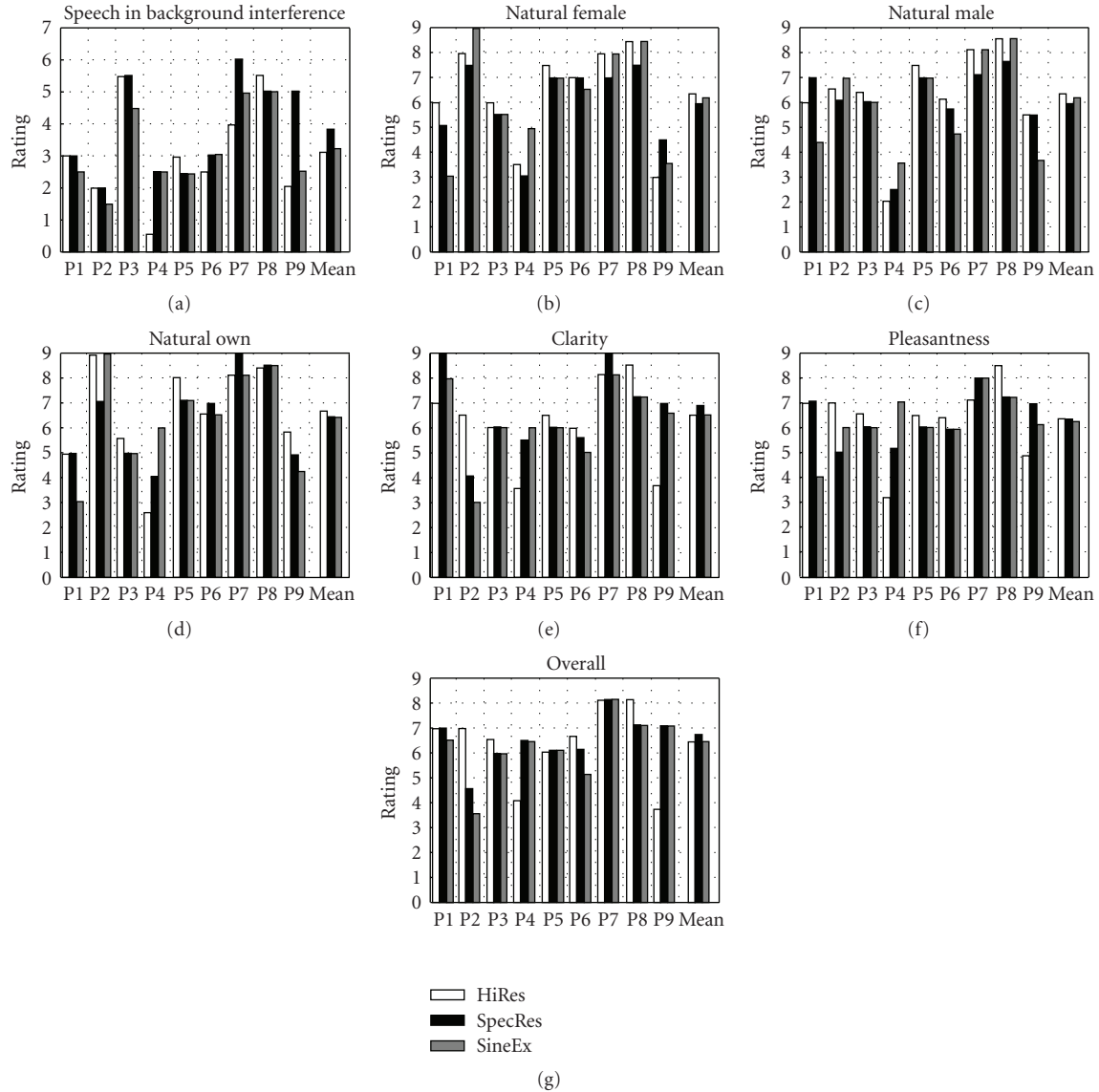


FIGURE 12: Results of the speech quality questionnaire using HiRes, SpecRes, and SineEx. Different items were rated on a 10-point scale. The cochlear implant subjects were asked to rate features of speech in different situations. The items are (a) speech in background interference, (b) natural female, (c) natural male, (d) natural own, (e) clarity, (f) pleasantness, and (g) overall speech quality.

SineEx, the frequency selection, which is related to the channels to be stimulated, also incorporates a simple version of a perceptual model based on masking effects.

In the version of SineEx evaluated in this study, sound perception is similar to the sound perceived with HiRes and SpecRes. However, this version of SineEx models the audio signal only with sinusoids, without considering noise components. Improvements to SineEx should include incorporating source models that process noise components as well as the sinusoidal components of the audio input signal because noise components comprise important cues for speech perception as well as for auditory scene analysis. An important aspect in designing new signal processing strategies for cochlear implants is the complexity of the algorithms. SineEx was implemented in the real-time Advanced Bionics bodyworn Platinum series processor, which uses a low-power

DSP. Therefore, implementation on a commercial behind-the-ear Harmony processor should not be a major problem. The Harmony processor can save up to three strategies at the same time. Given the large intersubject variability in speech and music perception with HiRes, SpecRes, and SineEx, the Harmony allows all three strategies to be placed on the same processor. This flexibility gives the user the opportunity to select between strategies depending upon the situation.

6. Conclusions

The SineEx strategy models the input audio signal with sinusoids and uses a psychophysical masking model and current steering to increase the spectral resolution in a cochlear implant while at the same time reducing channel

interaction. The SpecRes strategy uses an FFT and a spectral peak locator to extract the most dominant frequencies. SpecRes also uses current steering to improve the place-frequency accuracy of stimulation. Results show large variability among cochlear implant users in pitch perception, speech intelligibility, and music perception when comparing these two current steering strategies against HiRes, a sequential stimulation strategy. All patients tested until now performed equally well using the current steering technique as with conventional strategies. New modifications of the signal processing algorithms together with further investigation on simultaneous stimulation of the electrodes hold great promise for improving the hearing capabilities of cochlear implant users.

References

- [1] B. S. Wilson, C. C. Finley, D. T. Lawson, et al., "Better speech recognition with cochlear implants," *Nature*, vol. 352, no. 6332, pp. 236–238, 1991.
- [2] J. Laneau, *When the deaf listen to music Pitch perception with cochlear implants*, Ph.D. dissertation, Katholieke Universiteit Leuven, Faculteit Toegepaste Wetenschappen, Leuven, Belgium, 2005.
- [3] B. S. Wilson, R. Schatzer, E. A. Lopez-Poveda, X. Sun, D. T. Lawson, and R. D. Wolford, "Two new directions in speech processor design for cochlear implants," *Ear and Hearing*, vol. 26, no. 4, pp. 73s–81s, 2005.
- [4] L. Geurts and J. Wouters, "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 713–726, 2001.
- [5] Q.-J. Fu, F.-G. Zeng, R. V. Shannon, and S. D. Soli, "Importance of tonal envelope cues in chinese speech recognition," *Journal of the Acoustical Society of America*, vol. 104, no. 1, pp. 505–510, 1998.
- [6] K. Hopkins and B. C. J. Moore, "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," *Journal of Acoustical Society of America*, vol. 125, no. 1, pp. 442–446, 2009.
- [7] Boston Scientific, "HiRes with fidelity 120 sound processing," A Report from Advanced Bionics, The Auditory Business of Boston Scientific, 2006.
- [8] Z. M. Smith, B. Deleguette, and A. J. Oxenham, "Chimeric sounds reveal dichotomies in auditory perception," *Nature*, vol. 416, pp. 87–90, 2002.
- [9] A. Büchner, C. Frohne-Buechner, L. Gaertner, A. Lesinski-Schiedat, R.-D. Battmer, and T. Lenarz, "Evaluation of advanced bionics high resolution mode," *International Journal of Audiology*, vol. 45, no. 7, pp. 407–416, 2006.
- [10] G. S. Donaldson and H. A. Kreft, "Place-pitch discrimination of single- versus dualelectrode stimuli by cochlear implant users (1)(a)," *Journal of the Acoustical Society of America*, vol. 118, no. 2, p. 623, 2005.
- [11] B. S. Wilson, R. Schatzer, and E. A. Lopez-Poveda, "Possibilities for a closer mimicking of normal auditory functions with cochlear implants," in *Cochlear Implants*, pp. 48–56, Thieme Medical, New York, NY, USA, 2nd edition, 2006.
- [12] D. B. Koch, M. Downing, M. J. Osberger, and L. Litvak, "Using current steering to increase spectral resolution in CII and HiRes 90K users," *Ear and Hearing*, vol. 28, no. 2, supplement, pp. 38S–41S, 2007.
- [13] B. Townshend, N. Cotter, D. van Compernelle, and R. L. White, "Pitch perception by cochlear implant subjects," *Journal of the Acoustical Society of America*, vol. 82, no. 1, pp. 106–115, 1987.
- [14] H. J. McDermott and C. M. McKay, "Pitch Ranking with non-simultaneous dual-electrode electrical stimulation of the cochlea," *Journal of Acoustical Society of America*, vol. 96, pp. 155–162, 1994.
- [15] J. H. M. Frijns, R. K. Kalkman, F. J. Vanpoucke, et al., "Simultaneous and non-simultaneous dual electrode stimulation in cochlear implants: evidence for two neural response modalities," *Acta Oto-Laryngologica*, vol. 129, pp. 433–439, 2009.
- [16] M. A. Stone, B. C. J. Moore, J. I. Alcántara, and B. R. Glasberg, "Comparison of different forms of compression using wearable digital hearing aids," *Journal of the Acoustical Society of America*, vol. 106, no. 6, pp. 3603–3619, 1999.
- [17] A. J. C. Wilson, "The location of peaks," *British Journal of Applied Physics*, vol. 16, no. 5, pp. 665–674, 1965.
- [18] J. O. Smith III and X. Serra, "PARSHL: an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," in *Proceedings of the International Computer Music Conference (ICMC '87)*, pp. 290–297, Urbana-Champaign, Ill, USA, August 1987.
- [19] D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," *Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592–2605, 1990.
- [20] G. S. Stickney, P. C. Loizou, L. N. Mishra, P. F. Assmann, R. V. Shannon, and J. M. Opie, "Effects of electrode design and configuration on channel interactions," *Hearing Research*, vol. 211, no. 1-2, pp. 33–45, 2006.
- [21] H. Purnhagen, N. Meine, and B. Edler, "Sinusoidal coding using loudness-based selection," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '02)*, vol. 2, pp. 1817–1820, Orlando, Fla, USA, May 2002.
- [22] W. Nogueira, A. Büchner, Th. Lenarz, and B. Edler, "A psychoacoustic "NoFM"-type speech coding strategy for cochlear implants," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 18, pp. 3044–3059, 2005.
- [23] B. Edler, H. Purnhagen, and C. Ferekidis, "ASAC—analysis/synthesis codec for very low bit rates," in *Proceedings of the 100th Audio Engineering Society Convention (AES '96)*, Copenhagen, Denmark, May 1996, Preprint 4179.
- [24] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [25] H. Purnhagen and N. Meine, "HILN—the MPEG-4 parametric audio coding tools," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '00)*, vol. 3, pp. 23–31, Geneva, Switzerland, May 2000.
- [26] 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.
- [27] F. Baumgarte, C. Ferekidis, and H. Fuchs, "A nonlinear psychoacoustic model applied to the ISO MPEG layer 3 coder," in *Proceedings of the 99th Audio Engineering Society Convention (AES '95)*, New York, NY, USA, October 1995, Preprint 4087.
- [28] 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, no. 6, supplement, p. 83, 1997.

- [29] H. Levitt, "Transformed up-down methods in psychoacoustics," *Journal of the Acoustical Society of America*, vol. 49, no. 2, part 2, pp. 467–477, 1971.
- [30] International Telecommunication Union (ITU) G. 227, "International analogue carrier systems -General characteristics common to all analogue carrier-transmission systems—General characteristics common to all analogue carrier-transmission systems—Conventional telephone signal," ITU 1988, 1993.
- [31] K. Wagener, V. Kuehnel, and B. Kollmeier, "Entwicklung und evaluation eines satztests fuer die deutsche sprache i: design des oldenburgers satztests," *Zeitschrift für Audiologie*, vol. 38, pp. 4–15, 1999.
- [32] G. Clark, *Cochlear Implants: Fundamentals and Applications*, Modern Acoustics and Signal Processing, Springer, AIP Press, Berlin, Germany, 2003.
- [33] A. Gabrielsson and H. Sjogren, "Perceived sound quality of sound reproducing," *Journal of the Acoustical Society of America*, vol. 65, pp. 1019–1726, 1979.
- [34] N. B. Spetner and L. W. Olsho, "Auditory frequency resolution in human infancy," *Child Development*, vol. 61, no. 3, pp. 632–652, 1990.
- [35] E. J. Propst, K. A. Gordon, R. V. Harrison, S. M. Abel, and B. C. Papsin, "Sound frequency discrimination in normal-hearing listeners and cochlear implantees," *The University of Toronto Medical Journal*, vol. 79, no. 2, 2002.