Speech processing in vocoder-centric cochlear implants

Philipos C. Loizou
Department of Electrical Engineering
Jonsson School of Engineering and Computer Science
University of Texas at Dallas
P.O.Box 830688
Richardson, TX 75083-0688
loizou@utdallas.edu

ABSTRACT

The principles of most recent cochlear implant processor are similar to that of the channel vocoder, originally used for transmitting speech over telephone lines with much less bandwidth than that required for transmitting the unprocessed speech signal. An overview of the various vocoder-centric processing strategies proposed for cochlear implants since the late 1990s is provided including the strategies used in different commercially available implant processors. Special emphasis is placed on reviewing the strategies designed to enhance pitch information for potentially better music perception. The various noise suppression strategies proposed over the years based on multi-microphone and single-microphone inputs are also described.

1. Introduction

This Chapter presents an overview of the various vocoder-centric processing strategies proposed for cochlear implants since the late 1990s (a review of earlier strategies can be found in [1]). This Chapter also offers a review of the strategies used in different commercially available implant processors.

1.1. HISTORICAL BACKGROUND

In 1939, at the World's Fair in New York City, people watched with intense curiosity the first talking machine. The machine spoke with the help of a human operator seating in front of a console, similar to a piano keyboard, consisting of 10 keys, a pedal and a wrist bar. Inside the machine were analog circuits of bandpass filters, switches and amplifiers connected to a loudspeaker. The talking machine contained the first artificial speech-synthesis system implemented in hardware. This speech synthesis system, pioneered by Homer Dudley from Bell Laboratories, came to be known as the channel vocoder (voice coder) [2]. Dudley's vocoder idea had a profound impact not only on telephony and speech transmission applications [3,4], but also much later in the development of cochlear implant processors. The latest and most successful signal-processing strategies used in cochlear implants are based on vocoding analysis principles. All cochlear implant devices today are programmed (now digitally) with a modified version of the vocoder analysis algorithm.

2. The channel vocoder

The channel vocoder [2,4] consists of a speech analyzer and a speech synthesizer (see Figures 1 and 2). In speech transmission applications, the analyzer would be utilized at the transmitter and the synthesizer at the receiver end. The incoming signal is first filtered into a number of contiguous frequency channels using a bank of band-pass filters (10 filters were used in Dudley 's1939 demonstration). The envelope of the signal in each channel is estimated by full-wave rectification and low-pass filtering and is then downsampled and quantized for transmission. In addition to envelope estimation, the vocoder analyzer makes a voiced/unvoiced decision and estimates the vocal pitch (F0) of the signal. These two pieces of information are transmitted alongside the envelope information. The synthesizer modulates the received envelopes by the appropriate excitation as determined by the voiced/unvoiced (binary) signal. The excitation signal consists of random noise for unvoiced speech segments and a periodic pulse generator for voiced speech, with the period of the pulse generator being controlled by F0. The modulated signals are subsequently bandpass-filtered by the same filters and then added together to produce the synthesized speech waveform.

Current cochlear implant processors (sixty years later) utilize the same blocks of the channel vocoder analyzer shown in Figure 1. At present, only the vocoder analyzer is used for transmitting envelope information to the individual electrodes, but recently there has been a shift in research focus toward implementing blocks of the synthesizer as well [5,6]. Interestingly, early devices based on feature extraction strategies modulated the estimated formant amplitudes by F0 [1]. These strategies, however, were abandoned due to the inherent difficulties associated with F0 extraction in noisy environments. It is also interesting to note that the acoustic cochlear implant simulations often used to study performance of cochlear implant patients in the absence of confounding factors (e.g., duration of deafness, insertion depth) utilize the synthesizer (Fig. 2). By choosing random noise as the excitation signals for all segments of speech, we get the noise-band cochlear implant simulations [7]. Similarly, by choosing sine waves with frequencies set to the center frequencies of the bandpass filters as the excitation signals, we get the sine wave simulations [8].

3. Vocoder-centric strategies for cochlear implants

There are currently two variations of the channel vocoder (Fig. 1) that are used in all implant processors. The first implementation uses the analyzer of the channel vocoder in its original form (as per Fig. 1). The second implementation also uses the analyzer of the channel vocoder, but selects only a subset of the envelope outputs for stimulation. This section describes in detail these two variations.

3.1 Continuous Interleaved Sampling (CIS) strategy

The first cochlear implant device to adopt a channel-vocoder strategy was the Ineraid device manufactured by Symbion, Inc., Utah. The signal was first compressed using an automatic gain control, and then filtered into four contiguous frequency bands, with center frequencies at 0.5, 1, 2, and 3.4 kHz [9]. The filtered waveforms went through adjustable gain controls and then sent directly through a percutaneous connection to four intracochlear electrodes. The filtered waveforms were delivered simultaneously to four electrodes in analog form. A major concern associated with simultaneous stimulation is the interaction between channels caused by the summation of electrical fields from individual electrodes. Neural responses to stimuli from one electrode may be significantly distorted by stimuli from other electrodes. These interactions may distort speech spectral information and therefore degrade speech understanding.

A simple solution to the channel interaction problem was proposed by researchers at the Research Triangle Institute (RTI) via the use of non-simultaneous, interleaved pulses [10]. They proposed modulating the filtered waveforms by trains of biphasic pulses that were delivered to the electrodes in a non-overlapping (non-simultaneous) fashion, that is, in a way such that only one electrode was stimulated at a time (Figure 3). The amplitudes of the pulses were derived by extracting the envelopes of the band-passed waveforms. The resulting strategy was called the Continuous Interleaved Sampling (CIS) strategy.

The block diagram of the CIS strategy is shown in Figure 3. The signal is first pre-emphasized and then applied to a bank of bandpass filters. The envelopes of the outputs of these bandpass filters are then full-wave rectified and low-pass filtered (typically with 200 or 400 Hz cutoff frequency). The envelopes of the outputs of the bandpass filters are finally compressed and used to modulate biphasic pulses. A non-linear compression function (e.g., logarithmic) is used to ensure that the envelope outputs fit the patient's dynamic range of electrically evoked hearing. Trains of balanced biphasic pulses, with amplitudes proportional to the envelopes, are delivered to the electrodes at a constant rate in a non-overlapping fashion.

Figure 3 shows the basic configuration for the CIS strategy. Many variations of the CIS strategy have emerged and are currently used by the three implant manufacturers. Some devices for instance use the fast Fourier transform (FFT) for spectral analysis and some use the Hilbert transform to extract the envelope instead of full-wave rectification and low-pass filtering. Although the CIS strategy is employed by all three manufacturers, it is based on different implementations.

3.1.1 CIS design parameters

The CIS strategy can be configured in a number of ways by varying design parameters (e.g., filter spacing, envelope cut-off frequencies, etc) of the vocoder. These parameters include, among other things, the envelope detection method, stimulation rate (i.e., the number of pulses delivered to the electrodes per second), shape of compression function and filter spacing. A subset of these parameters may be varied to optimize speech recognition performance for each patient.

3.1.2. Stimulation rate

The pulse rate (the number of pulses per sec (pps) delivered to each electrode) may be as low as 250 pulses/sec or as high as 5000 pulses/sec in some devices. It would be reasonable to expect that better recognition performance should be obtained with high pulse-rates, since high pulse-rate stimulation can better represent fine temporal modulations. This is illustrated in Figure 4, which shows the pulsatile waveforms of the syllable /t/ obtained at different rates. As shown in Fig. 4, the unvoiced stop consonant /t/ is marked by a period of silence (closure) followed by a burst and aspiration. As the pulse rate increases, the burst becomes more distinctive, and perhaps more salient perceptually. There seems to be no evidence of the burst at low rates, 200 or 400 pulses/s. This example clearly demonstrates that lower rates do not provide a good, if any at all, temporal representation of the burst in stop consonants.

Despite the theoretical advantages of higher stimulation rates, the outcomes from several studies have not been consistent. While the majority of those studies [11-15] found a positive effect of high stimulation rates a few studies [16,17] found no significant effect. It is, however, consistent across these studies that some patients received large benefits with high stimulation rates while other patients received, little, or no benefit. Possible reasons for the discrepancies in the outcomes between the various studies include: (a) differences in implementation of the CIS strategy, (b) differences in speech materials used, and (c) differences in electrode design between devices. Each manufacturer has its own implementation of the CIS strategy. In the Nucleus device for instance, the FFT is used for spectral analysis in lieu of the bank of bandpass filters. Limited by the FFT analysis frame rate, extremely high stimulation rates can be obtained by repeating stimulus frames. Therefore, higher stimulation rates might not necessarily introduce new information, which explains the lack of improvement with high stimulation rates [16].

The influence of speech materials when examining the effect of parametric variations of the CIS strategy was demonstrated in the study by Loizou *et al.* [12] which assessed speech recognition as a function of stimulation rate in six Med-El/CIS-Link cochlear implant (CI) listeners. Results showed that higher stimulation rates >2100 pulses/sec produced a significantly higher performance on word and consonant recognition than lower stimulation rates (800 pulses/sec). The effect of stimulation rate on consonant recognition was highly dependent on the vowel context. The largest benefit was noted for consonants in the /iCi/ and /uCu/ contexts, while the smallest benefit was noted for consonants in the /aCa/ context. This finding suggests that the /aCa/ consonant test, which is widely used, is not sensitive enough to parametric variations of implant processors.

The advantages of high stimulation rates are unfortunately offset by the increased channel interaction associated with extremely high stimulation rates. Since each manufacturer uses different number of electrode contacts with different electrode spacing (see Table 1), it is reasonable to assume that a wider spacing between electrodes will yield smaller amounts of channel interaction. Consequently, the electrode spacing confounds the effect of high stimulation rates on speech recognition when comparing different devices. The Nucleus device has the smallest electrode spacing (0.7mm) while the Med-El device has the widest electrode spacing (2.4mm) (The Ineraid device has in fact the widest spacing (4mm), but is not commercially available). It is therefore not surprising that most of the benefits reported with high stimulation rates were with Med-El users and not with Nucleus users. Significant benefits were reported in [15,18] with Nucleus users, but with those users fitted with a spectral-maxima strategy running at high stimulation rates.

As mentioned above, some patients do receive significant benefit with the use of high stimulation rates. The "optimal" pulse rate, however, as far as speech recognition performance is concerned, varies from patient to patient. Wilson *et al.* [11], for instance, reported that some patients obtain a maximum performance on the 16-consonant recognition task with a pulse rate of 833 pulses/sec and pulse duration of 33 μsec/phase. Other patients obtain small but significant increases in performance as the pulse rate increases from 833 pps to 1365 pps, and from 1365 pps to 2525 pps, using 33 μsec/phase pulses. Unfortunately, there are no known methods for identifying the “optimal” pulse rate for each patient, other than trying out different values and examining their performance.

Current commercial implant processors are operating at stimulation rates ranging from 800 pulses/sec/channel to 2500 pulses/sec/channel, depending on the device. Use of very high-rates (>5000 pulses/sec) is being investigated by some as a means of restoring the stochastic independence of neural responses, which is lost with the overly synchronized electrical stimulation. In acoustic hearing, it is known that the nature of the neuron responses is stochastic in that the firing of a particular auditory-nerve fiber has no effect on the probability of a neighboring fiber firing. In electric stimulation, however, the response of single neurons is highly synchronized and also entrained with the stimulus, in that neurons fire on every stimulation cycle, up to rates of 800 Hz [19,20]. The stochastic nature (i.e., the independence) of the neural responses is lost with electrical stimulation since all the neurons in a local region fire at the same time (i.e., in synchrony). To restore the stochastic independence of neuron responses, Rubinstein *et al.* [21] proposed the use of high-frequency (5000 pulses/sec) desynchronizing pulse trains over the stimulus delivered by the processor. Litvak *et al.* [22] demonstrated that the use of desynchronizing pulse trains can improve the representation of both sinusoidal and complex stimuli (synthetic vowels) in the temporal discharge patterns of auditory nerve fibers for frequencies up to 1000 Hz. The addition of un-modulated high-rate pulse trains over the electrical stimulus can also result in significant increases in psychophysical dynamic range [23]. Another method proposed for restoring the stochastic independence of neural responses is the addition of appropriate amount of noise to the acoustic stimuli [24,25].

3.1.3. Compression function

The compression of envelope amplitudes is an essential component of the CIS processor because it transforms acoustical amplitudes into electrical amplitudes. This transformation is necessary because the range in acoustic amplitudes in conversational speech is considerably larger than the implant patient's dynamic range. Dynamic range is defined here as the range in electrical amplitudes between threshold (barely audible level) and loudness uncomfortable level (extremely loud). In conversational speech, the acoustic amplitudes may vary within a range of 30-50 dB [26,27]. Implant listeners, however, may have a dynamic range as small as 5 dB. For that reason, the CIS processor compresses, using a non-linear compression function, the acoustic amplitudes to fit the patient's electrical dynamic range. The logarithmic function is commonly used for compression because it matches the loudness between acoustic and electrical amplitudes [28,29]. It has been shown that the loudness of an electrical stimulus in microamps is analogous to the loudness of an acoustic stimulus in dB.

Logarithmic compression functions of the form $Y = A \log(1+C\square) + B$ are typically used, where $\square$ is the acoustic amplitude (output of envelope detector), A, B and C are constants, and Y is the (compressed) electrical amplitude. Other types of compression functions used are the power-law functions of the form:

$$y = Ax^p + B \tag{1.1}$$

where p<1. The advantage of using power-law functions is that the shape, and particularly the steepness of the compression function, can be easily controlled by simply varying the value of the exponent *p*. The constants A and B are chosen such that the input acoustic range is mapped to the electrical dynamic range [THR, MCL], where THR is the threshold level and MCL is the most comfortable level measured in μamps [1]. The input acoustic range, also known as input dynamic range (IDR), is adjustable in some devices and can range from 30-70 dB. The effect of IDR on speech recognition was examined in several studies (e.g., [27,30]).

The effect of the shape of the compression function on speech recognition has been investigated in a number of studies [12,31-34]. Loizou *et al.* [12] modified the shape of the amplitude mapping functions ranging from strongly compressive to weakly compressive by varying the power exponent in Eq. (1.1) from p=-0.1 (too compressive) to p=0.5 (nearly linear). Results indicated that the shape of the compression function had only a minor effect on performance, with the lowest performance obtained for nearly linear mapping functions.

3.1.4 Envelope detection

Two different methods can be used to extract the envelopes of filtered waveforms. The first method includes rectification (full-wave or half-wave) followed by low-pass filtering at 200-400 Hz. The second method, currently used by the Med-El device, uses the Hilbert transform. No clear advantage has been demonstrated for the use of one method over the other for envelope extraction.

The first method is simple to implement as it involves full-wave or half-wave rectification and low-pass filtering. The low-pass filter is a smoothing filter and also serves as an antialiasing filter, which is required prior to donwsampling (Figure 1) the filtered waveforms. The stimulation rate needs to be at least two times higher (Nyquist rate) than the cutoff-frequency of the low-pass filter. Psychophysics studies [35] suggest that it should be at least four times the envelope cutoff frequency. Pitch increased with sinusoidally amplitude-modulated pulse trains up to a modulation frequency of about 200-300 Hz, provided the carrier rate (stimulation rate) was at least four times the modulation frequency [35]. Similar findings were also reported in intracochlear evoked potential studies [36].

The cut-off frequency of the low-pass filter controls the modulation depth of the envelopes. The lower the cutoff frequency is, the smaller the modulation depth of the filtered waveform (see examples in Figure 7), i.e., the flatter the envelopes are. Simulation studies [7,37] demonstrated no significant effect of the envelope cutoff frequency on speech recognition by normal-hearing listeners. This was also confirmed with studies from our lab with cochlear implant patients (see Figure 5) tested on consonant and melody recognition tasks [38]. No significant effect of envelope cutoff frequency on consonant and melody recognition was found.

The second envelope detection method is based on the Hilbert transform [39], a mathematical tool which can represent a time waveform as a product of slowly-varying envelope and a “carrier” signal, containing fine structure information (see example in Fig. 6). More specifically, the filtered waveform, $x_i(t)$, in the *i*th band (channel) can be represented as

$$x_i(t) = a_i(t)\cos\phi_i(t) \tag{1.2}$$

where $a_i(t)$ represents the envelope of the *i*th band at time *t*, and $\cos\left(\phi_i(t)\right)$ represents the fine-structure waveform of the *i*th band. Note that $\phi_i(t)$ is called the instantaneous phase of the signal, and the derivative of $\phi_i(t)$ produces the instantaneous frequency (carrier frequency) of the signal, which varies over time. The fine-structure waveform is a frequency modulated (FM) signal (Fig. 6) since the carrier frequency is not fixed but varies with time. Figure 6 shows an example of the decomposition of the time-domain waveform of the vowel /a/ into its envelope and fine-structure. It is clear from Fig. 6 that the Hilbert envelope contains periodicity information and therefore is not the same as the envelope defined by Rosen [40]. The Hilbert transform renders Rosen’s [40] three-way partition of the temporal structure of speech into a two-way partition: the envelope, which also contains periodicity information, and the fine structure. This envelope/fine-structure decomposition of the signal (Fig. 6) can be done independently for each channel. Figure 7 shows examples of envelopes extracted using the above two methods: the Hilbert transform and rectification followed by low-pass filtering. Of the two methods, the Hilbert transform produces more accurate estimates of the envelope. Use of higher envelope cutoff frequencies, however, yields envelopes close to those extracted by the Hilbert transform (Fig. 7).

Current implant devices transmit envelope information ($a_i(t)$) and discard fine-structure information ($\cos\left(\phi_i(t)\right)$) as they implement only the analysis part of the vocoder and not the synthesis part (compare Figure 1 with Figure 3). Simulation studies [41-43] with normal-hearing listeners demonstrated the potential of including limited amounts of fine-structure information. It is not yet clear, however, how to incorporate fine-structure information in cochlear implants in a way that they can perceive it [44].

3.1.5. Filter spacing

For a given signal bandwidth (e.g., 0-8 kHz), there exist several ways of allocating the filters in the frequency domain. Some devices use a logarithmic spacing while others use a linear spacing in the low-frequencies (< 1300 Hz) and logarithmic spacing thereafter (> 1300 Hz). The effect of filter spacing on speech recognition, melody recognition and pitch perception has been investigated in a number of studies [45-48].

Fourakis *et al.* [47] advocated the placement of more filters in the F1/F2 region for better representation of the first two formants. They investigated the effect of filter spacing by modifying the electrode frequency boundary assignments of Nucleus 24 patients so as to include additional filters in the F1/F2 region. Small but significant improvements were noted on vowel recognition with an experimental MAP which included one additional electrode in the F2 region. No significant improvements were found on word recognition. The fixed number of frequency tables provided by the manufacturer, limited the investigators from assigning more electrodes in the F2/F3 region. The majority of the Nucleus-24 CI users tested preferred the experimental MAP over their everyday processor.

Similar findings were also found in our lab using newly implanted Clarion CII patients fitted with 16-channels of stimulation. The effect of three different filter spacings, which included log, mel [49] and critical-band [50] spacing, was investigated on recognition of 11 vowels in /hVd/ format. Results (see Figure 8) indicated that some subjects obtained a significant benefit with the critical-band spacing over the log spacing. Performance obtained with the mel frequency spacing was the lowest compared to the other two frequency spacing. This may be attributed to the number of frequency bands allotted in the F1 and F2 range. The mel-frequency spacing had the smallest number (4) of bands allocated in the 0-1 kHz range, which is the F1 range for most vowels. In contrast, both the critical-band and the log spacing had 6 bands in the F1 range. In addition, the critical-band spacing had 7 bands in the 1-3 kHz range (F2 range), while the log spacing had 6.

The effect of filter spacing on pitch perception has been investigated in [51,52], and will be discussed later (see Section 4.1). In brief, existing data support the idea that the number of filters allocated in the F1/F2 region can have a significant effect on performance, at least on vowel recognition tasks.

3.2. Spectral-maxima strategy

The spectral-maxima strategy implemented as the ACE (previously SPEAK) strategy on Cochlear Corporation devices [53]) and as the "n-of-m" strategy in other devices [54,55], has antecedents in the channel-picking vocoders of the 1950s [56] as well as Haskins Laboratories' Pattern Playback speech synthesizer [57]. The principle underlying the use of this strategy is that speech can be well understood when only the peaks in the short-term spectrum are transmitted. In the case of the Pattern Playback, only 4–6 of 50 harmonics needed to be transmitted to achieve highly intelligible speech—as long as the "picked" harmonics defined the first two or three formants in the speech signal.

The spectral-maxima strategy is similar to the CIS strategy with the main difference being that the number of electrodes stimulated is smaller than the total number of analysis channels. In this strategy, the signal is processed through m bandpass filters from which only a subset n ($n<m$) of the envelope amplitudes are selected for stimulation. More specifically, the n maximum

envelope amplitudes are selected for stimulation. The spectral-maxima strategy is sometimes called the "n-of-m" strategy or peak-picking strategy and is available in both Med-El and Nucleus-24 devices. In the Nucleus-24 device, out of a total of 20 envelope amplitudes, 10-12 maximum amplitudes are selected for stimulation in each cycle. The ACE (and SPEAK) strategy continuously estimates the outputs of the 20 filters and selects the ones with the largest amplitude. In the SPEAK strategy, the number of maxima selected varies from 5 to 10 depending on the spectral composition of the input signal, with an average number of six maxima. For broadband spectra, more maxima are selected and the stimulation rate is slowed down. For spectra with limited spectral content, fewer maxima are selected and the stimulation rate increases to provide more temporal information.

Several studies compared the performance of spectral-maxima and CIS strategies [15,18,58,59]. Cochlear implant simulation studies by Dorman *et al.* [59] indicated high performance with the spectral-maxima strategy even when a small number of maxima were selected in each cycle. A 3-of-20 processor (i.e., a processor that selected three maximum amplitudes out of 20 amplitudes in each cycle) achieved a 90% correct level of speech understanding for all stimulus material (sentences, vowels and consonants) presented in quiet. In contrast, it required 4, 6, and 8 channels of stimulation by CIS-type processors to achieve similar levels of performance for sentences, consonants, and vowels respectively. Hence, provided that there exist a large number of output analysis filters, only a small number of maxima need to be selected, an outcome consistent with the Pattern Playback studies. In noise (0 dB S/N), a minimum of 10 maxima needed to be selected for asymptotic performance on sentence recognition.

A study by Skinner *et al*. [15] compared the performance of Nucleus-24 implant patients fitted with the SPEAK, ACE and CIS strategies, after the patients used each strategy for a period of 4-6 weeks. Results indicated that the group mean score obtained with the ACE strategy on sentence recognition was significantly higher than the scores obtained with the SPEAK and CIS strategies. The SPEAK and ACE strategies are both spectral-maxima strategies selecting roughly the same number of envelope maxima (8-12) out of a total of 20 envelope outputs. The two strategies differ, however, in the stimulation rate. ACE's stimulation rate is significantly higher than SPEAK's and ranges from 900-1800 pps while SPEAK's rate is fixed at 250 pps. The higher scores obtained with ACE can therefore be attributed to its higher stimulation rate.

4. Speech coding strategies used in commercial devices

There are currently three cochlear implant processors in the United States approved by the Food and Drug Administration (FDA): the Nucleus 24, the Clarion and the Med-El processor. This section provides an overview of the signal processing strategies used in commercially available implant processors.

4.1 Advanced Bionics Corporation (Clarion CII/Auria device)

The Advanced Bionics Corporation's (ABC's) implant has undergone a number of changes in the past decade. ABC's first generation implant (Clarion S-Series) included an electrode array with 8 contacts and supported a number of stimulation strategies including a simultaneous (analog-type) stimulation strategy (see review in [60]). ABC's second generation device (termed Clarion CII) includes a 16-contact electrode array (HiFocus II) and supports simultaneous, partially simultaneous and non-simultaneous stimulation strategies. Temporal bone studies have shown that the placement of the implanted Clarion's HiFocus II electrode array is extremely close to the modiolar wall [61].

The Clarion CII device supports a high-rate CIS strategy, which can be delivered either non-simultaneously or partially simultaneously to 16 electrode contacts. Clarion's CIS strategy, called HiRes, differs from the traditional CIS strategy in the way it estimates the envelope. It uses half-wave rectification rather than full-wave rectification, and it does not use a low-pass filter. Instead, after the half-wave rectification operation, it averages the rectified amplitudes within each stimulation cycle. This averaging operation is in effect a low-pass filtering operation. The cutoff frequency of the low-pass filter depends on the number of samples to be averaged, i.e., it depends on the stimulation rate. The higher the stimulation rate is (i.e., the smaller the number of samples to average), the higher the cutoff frequency is.

In the HiRes strategy, the signal is first pre-emphasized and then bandpass filtered into 16 channels. The bandpass filters span the frequency range of 250 to 8000 Hz and are logarithmically spaced. The filtered waveforms are half-wave rectified, averaged and logarithmically compressed to the patients' electrical dynamic range. The compressed envelopes are transmitted via RF to the implant decoder, where are then modulated by trains of biphasic pulses for electrical stimulation. Comparisons between the conventional CIS strategy and the HiRes strategy were reported in [62,63].

The CII device utilizes a dual-action automatic gain control at the microphone input consisting of a slow-acting and fast acting stage. The slow-acting control has a compression threshold of 57 dB SPL with an attack time of 325 ms and a release time of 1000 ms. The second control is fast-acting and has a higher compression threshold of 65 dB SPL with an attack time of <0.6 ms and a release time of 8 ms.

The Clarion II device has 16 independent current sources that allow for simultaneous stimulation of two or more electrode contacts. When used in non-simultaneous mode of stimulation, HiRes operates at a stimulation rate of 2800 pps/sec using a pulse width of 11 μsces/phase. The stimulation rate can be further increased by the use of partially simultaneous stimulation whereby pairs of electrodes are stimulated simultaneously. To minimize potential channel interaction, non-adjacent pairs of electrodes are typically selected (e.g., 1-8, 2-7, etc.). For 16 electrodes configured with paired pulses and a narrow pulse width, the stimulation rate can exceed 5000 pps per channel. The combination of high rate stimulation and high cutoff frequency in the envelope detectors provides a fine temporal waveform representation of the signal at each channel. Some patients are able to utilize the fine temporal modulations present in the waveform at such high stimulation rates [12,62,63].

The presence of multiple current sources allows for the implementation of virtual channel processing strategies, currently under investigation by ABC. By properly manipulating (or steering) the current delivered simultaneously to adjacent electrodes, it is possible to elicit pitches intermediate to the pitches elicited by each of the electrodes alone. These intermediate pitches may introduce intermediate "virtual" channels of information. Different pitches can generally be elicited by controlling the proportion of current directed to each of the two electrodes [64]. Psychophysical studies have shown that simultaneous dual-electrode stimulation can produce as few as 2 and as many as 9 discriminable pitches between the pitches of single electrodes [65]. The motivation behind the virtual channel (also known as current-steering) idea is to produce intermediate pitches between electrodes in hopes of increasing the effective number of channels of information beyond the number of electrodes. The performance of the virtual channel strategy on music appreciation is currently being investigated by ABC. Anecdotal reports by some patients (e.g., [66]) fitted with the virtual channel strategy were very encouraging.

4.2 Cochlear Corporation (Nucleus-24 ESPrit 3G/ Freedom device)

The Nucleus-24 device (CI24M) is equipped with an array of 22 banded intra-cochlear electrodes and two extra-cochlear electrodes, one being a plate electrode located on the implant package and the other a ball electrode located on a lead positioned under the temporalis muscle [67]. The electrode contacts of the Nucleus 24 Contour array are oriented toward the modiolus minimizing possible current spread away from the target spiral ganglion cells. The electrodes can be stimulated in a bipolar, monopolar or common ground configuration. The extra-cochlear electrodes are activated during monopolar stimulation and can be used individually or together. Biphasic stimulus pulses are generated with electrode shorting during the inter-stimulus gap (about 8 μsecs) to remove any residual charge.

The CI24M processor can be programmed with the ACE and CIS strategies[16]. Both strategies estimate the input signal spectrum using a Fast Fourier Transform (FFT) rather than a bank of band-pass filters. The filter-bank is implemented using a 128 point Hanning Window and an FFT. Based on a sampling rate of 16 kHz this provides an FFT channel spacing of 125 Hz and a low-pass filter cut-off frequency of 180 Hz. The FFT bins, which are linearly spaced in frequency, are used to produce n (12-22) filter bands, which are typically linearly spaced from 188 to 1312 Hz and then logarithmically spaced up to 7938 Hz. A total of n (n=20) envelopes are estimated by summing the power of adjacent FFT bins within each of the n bands. In the ACE strategy, a subset of these m envelope amplitudes is then selected in each stimulation time frame. More specifically, 8-12 maximum amplitudes are selected for stimulation. In the CIS strategy, a fixed number of amplitudes are used for stimulation based on processing the signal through a smaller number of bands (10-12). The remaining electrodes are inactivated. Electrodes corresponding to the selected bands are then stimulated in a tonotopic basal to apical order.

The stimulation rate can be chosen from a range of 250 to 2400 pps per channel and is limited by a maximum rate of 14,400 pps across all channels. The stimulation rate can either be constant or jittered in time by a percentage of the average rate. When the jittered rate is programmed, the inter-stimulus gap (which is equal for all stimuli within one stimulation interval) is adjusted at every stimulation interval by a random amount. The resulting stimulation rate varies between consecutive stimulation intervals but has a fixed average rate.

For stimulation rates less than approximately 760 pps per channel, the filter-bank analysis rate is set to equal the stimulation rate. However, for higher stimulation rates, the analysis frequency is limited by the system to approximately 760 Hz and higher stimulation rates are obtained by repeating stimulus frames (stimuli in one stimulation interval) when necessary. For the 807 pps/channel rate, approximately one in every 17 or 18 stimulation frames is repeated. For the 1615 pps/channel rate, approximately every stimulus frame is repeated.

The majority of the Nucleus users are fitted with the ACE strategy [67]. Comparisons between the performance of the ACE, SPEAK and CIS strategies on multiple speech recognition tasks can be found in [15,67].

4.3 Med-El Corporation (Combi-40+/Tempo+/PULSARci[100] device)

The Med-El cochlear implant processor is manufactured by Med-El Corporation, Austria ([68]. The Med-El cochlear implant, also referred to as COMBI-40+ (C40+), uses a very soft electrode carrier specially designed to facilitate deep electrode insertion into the cochlea [69]. Because of the capability of deep electrode insertion (approximately 31 mm), the electrodes are spaced 2.4 mm apart spanning a considerably larger distance (26.4 mm) in the cochlea than any other commercial cochlear implant. The motivation for using wider spacing between electrode contacts is to increase the number of perceivable channels and to minimize potential interaction between electrodes.

The implant processor can be programmed with either a high-rate CIS strategy or a high-rate spectral-maxima strategy. The Med-El processor has the capability of generating 18,180 pulses/sec for a high-rate implementation of the CIS strategy in the 12 channel C40+ implant. The amplitudes of the pulses are derived as follows. The signal is first pre-emphasized, and then applied to a bank of 12 (logarithmically-spaced) bandpass filters. The envelopes of the bandpass filter outputs are extracted using the Hilbert transform [70]. Biphasic pulses, with amplitudes set to the mapped filter outputs, are delivered in an interleaved fashion to 12 monopolar electrodes at a default rate of of 1,515 pulses/sec per channel.

The latest Med-El device (PULSARci[100]) supports simultaneous stimulation of 12 electrodes. Higher (than the COMBI40+) stimulation rates are supported with aggregate rates up to 50,704 pulses/sec. For a 12-channel processor, rates as high as 4,225 pulses/sec/channel can be supported. Different stimulation techniques, including the use of triphasic pulses, are currently being explored by Med-El to reduce or minimize channel interaction associated with simultaneous stimulation.

5.0 Strategies designed to enhance F0 information

The above strategies were originally designed to convey speech information but fall short on many respects in conveying adequate vocal pitch (F0) information. Speakers of tonal languages, such as Cantonese and Mandarin, make use of vocal pitch variations to convey lexical meaning. Several researchers have demonstrated that CI users fitted with current strategies have difficulty discriminating between several tonal contrasts [71,72]. Also, CI users are not able to perceive several aspects of music including identification of familiar melodies and identification of musical instruments [73,74]. Hence, strategies designed to improve coding of F0 information are critically important for better tonal language recognition and better music perception.

Pitch information can be conveyed in cochlear implants via temporal and/or spectral (place) cues [52,75-79]. Temporal cues are present in the envelope modulations of the band-pass filtered waveforms (see example in Figure 7). Pitch can be elicited by varying the stimulation rate (periodicity) of a train of stimulus pulses presented on a single electrode, with high pitch percepts being elicited by high stimulation rates, and low pitch percepts being perceived by low stimulation rates. Once the stimulation rate increases beyond 300 Hz, however, CI users are no longer able to utilize such temporal cues to discriminate pitch [77]. Pitch may also be conveyed by electrode place of stimulation due to the tonotopic arrangement of the electrodes in the cochlea. Stimulation of apical electrodes elicits low pitch percepts while stimulation of basal electrodes elicits higher pitch percepts. Access to spectral cues is limited however, by the number of electrodes available (ranging from 12-22 in commercial devices), current spread causing channel interaction and possible pitch reversals due to suboptimal electrode placement.

A number of strategies have been proposed to enhance spectral (place) cues and/or temporal cues, and these strategies are described next.

5.1. Enhancing spectral (place) cues

Two different strategies have been explored to improve place coding of F0 information. The first approach is based on the use of virtual channels via the means of dual-electrode (simultaneous) stimulation. By properly manipulating (or steering) the current delivered simultaneously to adjacent electrodes, it is possible to elicit pitches intermediate to the pitches elicited by each of the electrodes alone. These intermediate pitches may introduce intermediate "virtual" channels of information. The virtual-channel approach is still in its infancy stages, and is currently being evaluated by several labs.

The second approach is based on modifying the shape of the filter response and/or the filter spacing. Such an approach was taken in [48,51,52]. A new filter bank was proposed by Geurts and Wouters [51] based on a simple loudness model used in acoustic hearing. The filter was designed such that the loudness of a pure tone sweeping through the filter increased linearly with frequency from the lower 3-dB cutoff frequency to the center frequency of each band, and decreased linearly from the center frequency to the upper boundary frequency. The resulting shape of the filters was triangular, with considerable overlap between adjacent filters. More filters were allocated in the low frequencies compared to a conventional filter-bank which was based on log spacing. The new filter-bank was tested on an F0 detection task in the absence of temporal cues by applying a 20-Hz low pass filter to the filter bank envelope signals. The new technique provided lower detection thresholds to F0 for synthetic vowel stimuli compared to a conventional filter bank approach. However, when temporal cues to F0 were reintroduced, differences in detection thresholds between filter banks were reduced indicating that the temporal cues also provided some information about F0.

Kasturi and Loizou [48] proposed the use of semitone-based filter spacing for better music perception. Results with cochlear implant simulations indicated that the semitone filter spacing consistently yielded better performance than the conventional filter spacing. Nearly perfect melody recognition was achieved with only four channels of stimulation based on the semitone filter spacing. Subsequent studies with Clarion CII users indicated that some subjects performed significantly better with 6 channels based on semitone spacing than with 16 channels spaced logarithmically as used in their daily strategy.

5.2. Enhancing temporal cues

The strategies designed to enhance temporal cues can be divided into two main categories: those that explicitly code F0 information in the envelope and those that aim to increase the modulation depth of the filtered waveforms in hopes of making F0 cues perceptually more salient.

The idea of modulating the extracted envelope by explicit F0 information is not new and dates back to the original channel vocoder synthesizer (Figure 2), which was based on a source-filter excitation approach. In channel-vocoded speech, voiced segments of speech are generated by exciting the vocal tract by a periodic (glottal) pulse train consisting of pulses spaced 1/F0 secs apart. Note that the F0-modulation idea was initially used in feature extraction strategies in the Nucleus device and later abandoned because of the inherent difficulty in extracting reliably F0 from the acoustic signal, particularly in noise. Jones *et al.* [80] investigated a strategy that provided F0 timing information on the most apical electrode. Results from several pitch perception tasks did not demonstrate any advantages with this approach.

Green and colleagues [5,81,82] adopted a similar approach to the enhancement of temporal pitch cues, based on the principle that *F*0 could be automatically extracted from voiced segments of the speech signal and used to appropriately modulate the envelopes. In the proposed strategy, amplitude envelopes were effectively split into two separate components. The first component contained slow rate information of 32 Hz conveying the dynamic changes in the spectral envelope that are important for speech. The second component presented *F*0 information in the form of a simplified synthesized waveform. More specifically, *F*0-related modulation was presented in the form of a saw-tooth waveform, on the assumption that "such a 'temporally sharpened' modulation envelope, with a rapid onset in each period, would lead to more consistent inter-pulse intervals in the neural firing pattern, and therefore to more salient temporal pitch cues" [5]. Implant users were required to label the direction of pitch movement of processed synthetic diphthong glides. Results indicated a significant advantage for the modified processing compared to standard CIS processing, demonstrating that the modified processing scheme was successful in enhancing the salience of temporal pitch cues. Subsequent studies by Green *et al.* [82], however, on tests of intonation perception and vowel perception indicated that the CI users performed worse with the F0-modified processing in vowel recognition compared to the conventional CIS strategy. The investigators concluded that while the modified processing enhanced pitch perception [5], it harmed the transmission of spectral information.

The above strategies assumed access to explicit F0 information. A number of strategies were proposed that did not rely on automatic extraction of F0 from the acoustic signal. These techniques focused on "sharpening" the envelopes so as to make the F0 information more apparent or perceptually more salient. This was accomplished by increasing the modulation depth of the envelopes. Geurts and Wouters [83] proposed a simple modification to the estimation of the envelope. Two fourth-order low-pass filters were first employed with cutoff frequencies of 400 Hz and 50 Hz. Note that the envelope output of the 400-Hz filter contained F0 modulations, but the envelope output of the 50-Hz filter did not. The modulation depth of the envelope was increased by subtracting an attenuated version of the 50-Hz (flat) log-compressed envelope from the 400-Hz log-compressed envelope. The resulting envelope was half-wave rectified (negative values set to zero), scaled and finally encoded for stimulation (note that the envelopes were already compressed to the patient's dynamic range prior to the subtraction operation). Despite the increase in modulation depth with the modified envelope processing, no significant differences in F0 discrimination of synthetic vowels were observed compared to the conventional CIS strategy [83]. Figure 9 shows examples of envelopes extracted with the above scheme and compared with envelopes extracted with conventional rectification and low-pass filtering (400 Hz).

The subtraction of the 400-Hz envelope amplitude from the 50-Hz envelope is equivalent to subtraction of the mean (dc component) of the rectified envelope, and constitutes a simple method for increasing envelope modulation depth. This idea was incorporated in one of the strategies proposed by Vandali *et al.* [6] to increase the modulation depth of the envelopes. The so called, Multi-channel Envelope Modulation (MEM) strategy utilized the envelope of the broadband signal (input acoustic signal prior to bandpass filtering), which inherently contains F0 periodicity information, to modulate the envelopes derived from the ACE filterbank. The envelope of the broadband signal was first estimated by full wave rectifying the broadband signal and then applying a 300-Hz, fourth-order low-pass filter (LPF). The modulation depth in the envelope signal was then expanded by applying an 80-Hz, second-order high-pass filter (HPF), which effectively increased the modulation depth of the envelope signal level by removing the mean (dc) component. Note that this step is equivalent to that of subtracting the mean of the rectified signal as done in [83]. The low-pass filtered signal, obtained prior to the HPF stage, was scaled and added to the output of the HPF stage. The expanded envelope signal was then half-wave rectified, to remove any negative values, and scaled. Finally, the narrow-band envelope signals estimated by the ACE filter bank were low-pass filtered, using a 50-Hz, second-order LPF, and then modulated by the normalized F0 envelope signal derived from the broadband signal. Figure 10 shows an example of the processed signal at different stages of the algorithm. As can be seen the derived envelope has large modulation depth and the F0 periodicity is evident in the envelopes. Note also that the envelopes are temporally synchronized (across all electrodes) with the input (broadband) waveform.

The second strategy (termed Modulation Depth Enhancement – MDE- strategy) evaluated by Vandali *et al*. [6] provided explicit modulation expansion by decreasing the amplitude of the temporal minima of the envelope. Modulation depths smaller than a specified level were expanded using a third-order power function, and modulation depths above this level, but below an upper limit of 20 dB, were linearly expanded. The modulation depth expansion was implemented by decreasing the amplitude of temporal minima in the signal while preserving the peak levels (a sliding time window, of duration 10 ms, was employed to track the peaks and minima). The modified envelope signals replaced those of the original envelope signals derived from the filter bank, and processing continued as per the normal ACE strategy. Comparison of the above strategies with the conventional ACE strategy indicated significantly higher scores with the MDE and MEM strategies on pitch ranking tasks. Comparison of the new strategies, however, on speech recognition tasks indicated no significant differences in scores with the conventional ACE strategy.

In brief, most of the above F0-enhancement strategies have been shown to improve pitch perception on tasks requiring discrimination of small pitch differences. Further work is needed, however, to investigate the efficacy of these F0-enhancement strategies on tonal language recognition and music perception tasks requiring perception of much finer pitch differences across a wide range of frequencies.

6. Noise reduction strategies

Perhaps one of the most common complaints made by CI listeners is that their performance decreases rapidly in noisy environments. This is not surprising given the limited amount of spectral information that they receive with their cochlear implant [84,85]. In noise, a larger number of channels are needed to understand speech [86,87]. Increasing the number of effective channels of spectral information, however, has been one of the biggest challenges in cochlear implants. For that reason, several researchers have focused on the development of noise reduction algorithms that either pre-process the noisy signal and feed the “enhanced” signal to the input of the processor or somehow suppress the noise present in the envelope amplitudes.

Several noise-reduction algorithms have been proposed for CI users. Some of those algorithms were based on the assumption that two or more microphones were available, while other algorithms assumed that the acoustic signal was picked up by a single microphone.

6.1. Multi-microphone methods

In some hearing aids and implant devices (e.g., Nucleus Freedom), a group of microphones with two or more entry ports are used with front and/or backward directivity. Some two-port microphones can easily reduce background noise simply by subtracting and delaying mechanically the input signals coming from each port of the diaphragm. Alternatively, the signals picked up by the two ports can be processed by an adaptive algorithm for better noise suppression.

The basis of the most sophisticated multi-microphone adaptive algorithms is the Griffiths-Jim beamforming algorithm [88] shown in Figure 11. When the target signal comes from the front, the subtracter output at the bottom input (mic 2) should contain primarily noise since the outputs from the two microphones will cancel each other. In contrast, the output of the adder in the top input (mic 1) should contain a mixture of the noise and the signal of interest. These two outputs containing the noisy signal and reference noise signals respectively are fed as input to an adaptive filter shown to the right in Fig. 11. The LMS algorithm [89] is used to adapt the filter coefficients in such a way as to minimize the power of the output error (Fig. 11). The error signal happens to be also the “enhanced” signal that is fed to the input of a hearing aid or cochlear implant device. The above beamforming algorithm (Fig. 11) has been found to work well in situations where there is only one noise source present and there is no reverberation.

Van Hoesel and Clark [90] tested an adaptive beamforming technique, similar to that shown in Figure 11, with four Nucleus-22 users. The adaptive beamforming (ABF) method used signals from two microphones – one behind each ear- to reduce noise coming from 90^{o} of the patients. The results of their study indicated that adaptive beamforming with two microphones can bring substantial benefits to CI users in conditions for which reverberation is moderate and only one source is predominantly interfering with speech. The ABF strategy yielded significantly higher intelligibility scores compared to a strategy in which the two microphone signals were simply added together. Hamacher *et al*. [91] evaluated the performance of two adaptive beamforming algorithms in different everyday-life noise conditions. The benefit of the two algorithms was evaluated in terms of the dB reduction in speech reception threshold (SRT). The mean benefit obtained using the beamforming algorithms for four CI users (wearing the Nucleus device) varied between 6.1 dB for meeting-room conditions to 1.1 dB for cafeteria noise conditions.

Margo *et al.* [92] evaluated a two-microphone beamforming algorithm with eight Nucleus users in a take-home trial for a period of 5-8 weeks. Subjective reports from the CI users indicated that the beamforming algorithm produced better sound quality and was preferred in noisy environments to their daily device. Wouters and van Berghe [93] evaluated the performance of an adaptive beamforming technique using a two-microphone array contained in a BTE hearing aid. The output of the noisy speech was pre-processed by the beamforming strategy and fed monaurally to the input of a LAURA cochlear implant processor. Speech was presented from the front and noise was presented from 90^{o} of the patients. Results indicated significant improvement in speech intelligibility corresponding to an SNR improvement of about 10 dB.

In brief, multi-microphone-based methods can bring substantial benefits to speech intelligibility in noise particularly in situations where there is a single interferer present and there is no reverberation.

6.2 Single-microphone methods

In the above studies, it was assumed that two (or more) microphones were available and in some cases, that each microphone was placed behind each ear. Adding, however, a second microphone contralateral to the implant is ergonomically difficult without requiring the CI users to wear headphones or a neck-loop, something that most patients would find cosmetically unappealing. Single-microphone noise reduction algorithms are therefore more desirable. These algorithms can be divided into two main categories: those that pre-process the noisy speech signal by a standard noise reduction algorithm and feed the "enhanced" output to the input of the CI processor, and those that are embedded or integrated within the subject's CI coding strategy.

A number of pre-processing noise-reduction strategies have been proposed for cochlear implants, some of which were implemented on old cochlear implant processors that were based on feature extraction strategies. Hochberg *et al*. [94] used the INTEL noise reduction algorithm to pre-process speech and presented the processed speech to 10 Nucleus implant users fitted with the F0/F1/F2 and MPEAK feature-extraction strategies [1]. Consonant-vowel-consonant (CVC) words embedded in speech-shaped noise at S/N ratios in the range -10 to 25 dB were presented to the CI users. Significant improvements in performance were obtained at S/N ratios as low as 0 dB. The improvement in performance was attributed to more accurate formant extraction, as the INTEL algorithm reduced the errors caused by the feature extraction algorithm. This was quantified later in a study by Weiss [95] who demonstrated that fewer formant extraction errors were made when the signal was first pre-processed with the INTEL algorithm.

A few pre-processing algorithms were also evaluated using the latest implant processors. Yang and Fu [96] evaluated the performance of a spectral-subtractive algorithm using subjects wearing the Nucleus-22, Med-El and Clarion devices. Significant benefits in sentence recognition were observed for all subjects with the spectral-subtractive algorithm, particularly for speech embedded in speech-shaped noise. Loizou *et al.* [97] evaluated a sub-space noise reduction algorithm [98] which was based on the idea that the noisy speech vector can be projected onto "signal" and “noise” subspaces. The clean signal was estimated by retaining only the components in the signal subspace and nulling the components in the noise subspace. The performance of the subspace reduction algorithm was evaluated using 14 subjects wearing the Clarion device. Results indicated that the subspace algorithm produced significant improvements in sentence recognition scores compared to the subjects' daily strategy, at least in continuous (stationary) noise.

All the above methods, including the multi-microphone methods, were based on pre-processing the noisy signal and presenting the “enhanced” signal to the CI users. The pre-processing approach has three main drawbacks, however: (1) pre-processing algorithms sometimes introduce unwanted distortion (e.g., musical noise [99]) in the signal despite the fact that these algorithms improve the SNR, (2) pre-processing algorithms can be highly complex (power hungry) and do not work synergistically with existing CI strategies, and (3) there is no simple approach for optimizing the algorithm to individual users, and consequently we often do not know why some users benefit while others do not. Ideally, noise reduction algorithms should be easy to implement and be integrated into the existing coding strategies. Only a few algorithms [100,101] were proposed along this direction.

Toledo *et al.* [100] proposed a simple envelope subtraction algorithm based on the principle that the clean (noise-free) envelope can be estimated by simply subtracting the noisy envelope from the noise envelope. This approach requires estimate of the noise envelope, which can be obtained using a noise estimation algorithm – an algorithm that continuously tracks the noise envelope even during speech activity. Results with four Clarion users indicated that some benefited with the envelope subtraction strategy. The lack of consistent improvement was attributed to inaccurate estimates of the noise envelope, which in turn might have produced speech distortion.

Loizou *et al*. [101] proposed the use of S-shaped compression functions in place of the conventional logarithmic compression functions for noise suppression. The motivation behind the use of S-shaped functions is to suppress the signal falling below the noise floor (and dominated by noise) while retaining the signal above the noise floor (and dominated by speech). This can be accomplished by the use of an expansive function for signal levels below the noise floor and a compressive function for signal levels above the noise floor (see Figure 12). In a way, the expansive segment of the function serves as a signal attenuator while the compressive segment serves as a signal amplifier. Key to the application of this S-shaped function is the choice of the knee point, which in [101] was set to the noise floor estimated using an algorithm. This knee point is not fixed, but adapted from cycle to cycle to the current estimate of the noise floor. Note that a similar input-output function is used in hearing aids [102], but with the knee-point fixed at a specific input level (e.g., 50 dB SPL). In [101], a noise estimation algorithm [103] was used to track continuously the noise floor and adapt the knee-point accordingly. The S-shaped function was evaluated with seven Clarion CII users using IEEE sentences [104] embedded in +5 dB multi-talker babble. Results (see Figure 13) showed significant improvements with the S-Shaped compression compared to the log compression used in the subject's daily strategy.

7. Summary and future challenges

This Chapter provided an overview of the various speech processing strategies developed for cochlear implants since the late 1990s. Many of those strategies, if not all, were variants of the Dudley's channel vocoder developed sixty years ago [2]. In fact, the latest attempts to design strategies to enhance F0 cues were similar to those used in the vocoder synthesizer (Fig. 2). These strategies have been shown to improve pitch perception but have not yet been shown to improve music or tonal language perception. This Chapter also presented an overview of the signal processing strategies available in commercial processors. It also described work already in progress in our lab and elsewhere in developing noise suppression strategies based on multi-microphone and single-microphone inputs.

The overview presented in this Chapter was by no means comprehensive. Several other strategies have been proposed but not described in this Chapter. These include the strategies designed to enhance onset cues [105-107], the strategies designed to enhance spectral contrast [101,108] and the strategies designed to provide a closer mimicking of the function of the normal cochlea [109]. Such strategies are currently under evaluation.

Despite the success of the current strategies in improving speech understanding, there still remains several challenges ahead including (but not limited to) the following:

- development of strategies for better music perception,
- development of noise suppression strategies for improved speech recognition in noisy environments,
- development of strategies tailored to individual patients (such strategies will bridge the gap between the poorly-performing users and the "star" users).

The development of such strategies will no doubt require a better understanding of:

(a) The mechanisms used for complex pitch perception in electrically evoked hearing, particularly, pertaining the interaction of temporal and place cues [51,110],

(b) The acoustic cues used by CI users for understanding speech in noise [111] and the factors influencing CI users' ability to receive release of masking when listening to speech embedded in fluctuating maskers [112,113],

(c) The factors influencing individual CI user's performance and the methods needed to assess the degree at which CI users are able to perceive temporal and/or spectral information. Such methods will help us design strategies that are tailored to individual CI user's perceptual capabilities.

Aside from the increased effort in the community to improve the design of speech coding strategies, there has also been effort to extend the capabilities of existing cochlear implant devices. Recent developments in cochlear implants include the use of bilateral implants and combined acoustic and electric stimulation (EAS) for subjects with residual hearing (see review in [109]). Results with bilateral implant patients [114] and EAS patients [115] have shown great promise in improving speech understanding in noise. Bilateral implants have also improved the ability of CI users to localize sounds [114]. Further research is needed in developing strategies that use coordinated stimulation of the two implant processors (currently, operating independently of one another) for perhaps better preservation of interaural time delay (ITD) cues.

In closing, it seems safe to expect further improvements in implant design and performance in the future, particularly regarding complex listening tasks such as listening to (and enjoying) Mozart's symphonies and conversing in crowded restaurants.

ACKNOWLEDGMENTS

The writing of this Chapter was partially supported by Grant No. R01 DC007527 from the National Institute of Deafness and other Communication Disorders, NIH.

References

[1] Loizou P: Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants. IEEE Signal Processing Magazine 1998; 15(5):101-130.

[2] Dudley H: Remaking speech. J Acoust Soc Am 1939; 11:1969-1977.

[3] Schroeder M: Vocoders: Analysis and synthesis of speech. Proc IEEE 1966; 54:720-734.

[4] Gold B, Rader C: The channel vocoder. IEEE Trans Audio Electroacoustics 1967; AU-15(4):148-161.

[5] Green T, Faulkner A, Rosen S: Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J Acoust Soc Am 2004; 116:2298-2310.

[6] Vandali A, Sucher C, Tsang D, McKay C, Chew J, McDermott H: Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies. J Acoust Soc Am 2005; 117(5):3126-3138.

[7] Shannon R, Zeng F-G, Kamath V, Wygonski J, Ekelid M: Speech recognition with primarily temporal cues. Science 1995; 270:303-304.

[8] Dorman M, Loizou P, Rainey R: Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J Acoust Soc Am 1997; 102(4):2403-2411.

[9] Eddington D: Speech discrimination in deaf subjects with cochlear implants. J Acoust Soc Am 1980; 68(3):885-891.

[10] Wilson CF, Lawson D, Wolford R, Eddington D, Rabinowitz W: Better speech recognition with cochlear implants. Nature 1991; 352:236-238.

[11] Wilson B, Lawson D, Zerbi M: Advances in coding strategies for cochlear implants. Advances in Otolaryngology - Head and Neck Surgery 1995; 9:105-129.

[12] Loizou P, Poroy O, Dorman M: The effect of parametric variations of cochlear implant processors on speech understanding. J Acoust Soc Am 2000; 108(2):790-802.

[13] Brill S, Gstottner W, Helms J, Ilberg Cv, Baumgartner W, Muller J et al.: Optimization of channel number and stimulation rate for the fast CIS strategy in the COMBI 40+. Am J Otol 1997; 18:S104-S106.

[14] Frijns J, Klop W, Bonnet R, Briare J: Optimizing the number of electrodes with high-rate stimulation of the Clarion CII cochlear implant. Acta Otolaryngolog 2003; 123(2):138-142.

[15] Skinner M, Holden L, Whitford L, Plant K, Psarros C, Holden T: Speech recognition with the Nucleus 24 SPEAK, ACE and CIS speech coding strategies in newly implanted adults. Ear Hear 2002; 23(3):207-223.

[16] Vandali A, Whitford L, Plant K, Clark G: Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system. Ear Hear 2000; 21(6):608-624.

[17] Friesen L, Shannon R, Cruz R: Effects of stimulation rate on speech recognition with cochlear implants. Audiol Neurootol 2005; 10(3):169-184.

[18] Kim H, Shim Y, Chung M, Lee Y: Benefit of ACE compared to CIS and SPEAK coding strategies. Adv Otorhinolaryngol 2000; 57:408-411.

[19] Javel E. Acoustic and electrical encoding of temporal information. In: Miller J, Spelman F, editors: Cochlear Implants: Models of the Electrically Stimulated Ear. NY: Springer-Verlag, 1990: 247-295.

[20] Van den Honert C, Stypulkowski P: Physiological properties of the electrically stimulated auditory nerve II. Single fiber recordings. Hear Res 1984; 14:225-243.

[21] Rubinstein J, Wilson B, Finley C, Abbas P: Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation. Hear Res 1999; 127:108-118.

[22] Litvak L, Delgutte B, Eddington D: Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains. J Acoust Soc Am 2003; 114(4):2099-2111.

[23] Hong R, Rubinstein J: High-rate conditioning pulse trains in cochlear implants: Dynamic range measures with sinusoidal stimuli. J Acoust Soc Am 2003; 114(6):3327-3342.

[24] Morse R, Evans E: Enhancement of vowel coding for cochlear implants by addition of noise. Nature 1996; 2:928-932.

[25] Zeng F-G, Fu Q, Morse R: Human hearing enhanced by noise. Brain Research 2001; 869:251-255.

[26] Boothroyd A, Erickson FN, Medwetsky L: The hearing aid input: A phonemic approach to assessing the spectral distribution of speech. Ear Hear 1994; 6:432-442.

[27] Zeng F-G, Grant G, Niparko J, Galvin J, Shannon R, Opie J et al.: Speech dynamic range and its effect on cochlear implant performance. J Acoust Soc Am 2002; 111(1):377-386.

[28] Eddington D, Dobelle W, Brachman D, Mladevosky M, Parkin J: Auditory prosthesis research using multiple intracochlear stimulation in man. Annals of Otology,Rhinology and Laryngology 1978; 87(Suppl. 53):1-39.

[29] Zeng F-G, Shannon R: Loudness balance between acoustic and electric stimulation. Hearing Research 1992; 60:231-235.

[30] Fu Q-J, Shannon R: Effect of acoustic dynamic range on phoneme recognition in quiet and noise by cochlear implant users. J Acoust Soc Am 1999; 106(6):L65-L70.

[31] Fu Q-J, Shannon R: Effect of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. J Acoust Soc Am 1998; 104(5):2570-2577.

[32] Fu Q-J, Shannon R: Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping. J Acoust Soc Am 1999; 106(2):L18-L23.

[33] Wilson B, Lawson D, Zerbi M, Wolford R: Speech Processors for Auditory Prostheses. NIH Project N01-DC-8-2105, Third Quarterly Progress Report 1999.

[34] Zeng F-G, Galvin J: Amplitude mapping and phoneme recognition in cochlear implant listeners. Ear Hear 1999; 20:60-74.

[35] McKay CM, McDermott HJ, Clark GM: Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. J Acoust Soc Am 1994; 96:2664-2673.

[36] Wilson B, Zerbi M, Finley C, Lawson D, van Honert C: Speech Processors for Auditory Prostheses. NIH Project N01-DC-5-2103, Eighth Quarterly Progress Report 1997.

[37] Drullman R: Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 1995; 97(1):585-592.

[38] Lobo A, Toledo F, Loizou P: The effect of envelope cutoff frequency on consonant and melody recognition by CI listeners. J Acoust Soc Am 2002; 112(5):2245.

[39] Hartmann W. Signals, sounds and sensation. New York: Springer-Verlag, 1998.

[40] Rosen S: Temporal information in speech: Acoustic, auditory and linguistic aspects. Phil Trans R Soc Lond 1992; 336:367-373.

[41] Zeng F-G, Nie K, Stickney G, Kong Y, Vongphoe M, Weit C et al.: Speech recognition with amplitude and frequency modulations. Proc Nat Acad Sciences 2005; 102(7):2293-2298.

[42] Smith ZM, Delgutte B, Oxenham AJ: Chimaeric sounds reveal dichotomies in auditory perception. Nature 2002; 416:87-90.

[43] Xu L, Pfingst B: Relative importance of temporal envelope and fine structure in lexical-tone perception. J Acoust Soc Am 2003; 114(6):3204-3207.

[44] Wilson B, Sun X, Schatzer R, Wolford R: Representation of fine structure or fine frequency information with cochlear implants. International Congress Series 2004; 1273:3-6.

[45] Skinner M, Holden L, Holden T: Effect of frequency boundary assignment on speech recognition with the SPEAK speech-coding strategy. Annals of Otology, Rhinology, and Laryngology 1995; 104(Suppl. 166):307-311.

[46] Fu Q-J, Shannon R: Frequency mapping in cochlear implants. Ear Hear 2002; 23(4):339-348.

[47] Fourakis M, Hawks J, Holden L, Skinner M, Holden T: Effect of frequency boundary assignment on vowel recognition with the Nucleus 24 ACE speech coding strategy. J Am Acad Audiol 2004; 15:281-289.

[48] Kasturi K, Loizou P: Effect of filter spacing and correct tonotopic representation on melody recognition: Implications for cochlear implants. Proceedings of ARO, New Orleans, LA 2005.

[49] Fant G. Speech Sounds and Features. Boston: MIT Press, 1973.

[50] Zwicker E, Fastl H. Psychoacoustics, Facts and Models. Berlin: Springer Verlag, 1990.

[51] Geurts L, Wouters J: Better place-coding of the fundamental frequency in cochlear implants. J Acoust Soc Am 2004; 115(2):844-852.

[52] Laneau J, Moonen M, Wouters J: Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. J Acoust Soc Am 2004; 106(6):3606-3619.

[53] McDermott H, McKay C, Vandali A: A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. J Acoust Soc Am 1992; 91:3367-3371.

[54] Wilson B, Finley C, Farmer J, Lawson D, Weber B, Wolford R et al.: Comparative studies of speech processing strategies for cochlear implants. Laryngoscope 1998;1069-1077.

[55] Wilson B, Finley C, Farmer J, Lawson D, Weber B, Wolford R et al.: Comparative studies of speech processing strategies for cochlear implants. Laryngoscope 1988;1069-1077.

[56] Peterson G, Cooper F: Peakpicker: A bandwidth compression device. J Acoust Soc Am 1957; 29:777(A).

[57] Cooper F, Liberman A, Borst J: Preliminary studies of speech produced by a pattern playback. J Acoust Soc Am 1950; 22:678.

[58] Kiefer J, Muller J: Speech understanding in quiet and in noise with the CIS speech coding strategy (Med-El Combi-40) compared to the multi-peak and SPEAK strategies. ORL 1996; 58:127-135.

[59] Dorman M, Loizou P, Spahr T, Maloff E: A Comparison of the Speech Understanding Provided by Acoustic Models of Fixed-Channel and Channel-Picking Signal Processors for Cochlear Implants. J Speech, Lang Hear Res 2002; 45(4):783-788.

[60] Loizou P, Stickney G, Mishra L, Assmann P: Comparison of speech processing strategies used in the Clarion implant processor. Ear Hear 2003; 24(1):12-19.

[61] Balkany T, Esharaghi A, Yang N: Modioloar proximity of three perimodiolar cochlear implant electrodes. Acta Otolaryngol 2002; 122(4):363-369.

[62] Filipo M, Mancini P, Ballantyne D, Bosco E, D'Elia C: Short-term study of the effect of speech coding strategy on the auditory performance of pre- and post-lingually deafened adults implanted with the Clarion CII. Acta Otolaryngol 2004; 124(4):368-370.

[63] Ostroff J, David E, Shipp D, Chen J, Nedzelski J: Evaluation of the high-resolution coding strategy for the Clarion CII cochlear implant system. J Otolaryngol 2003; 32(2):81-86.

[64] Wilson B, Zerbi M, Lawson D: Speech processors for auditory prostheses. NIH Contract N01-DC-2-2401, 3rd Quarterly Progress Report, February 1- April 30 1993.

[65] Donaldson G, Kreft H, Litvak L: Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users. J Acoust Soc Am 2005; 118(2):623-626.

[66] Chorost M: My bionic quest for Bolero. Wired Magazine 2005; 11:144-159.

[67] Parkinson A, Arcaroli J, Staller S, Arndt P, Cosgriff A, Ebinger K: The Nucleus 24 Contour cochlear implant system: Adult clinical trial results,". Ear Hear 2002; 23(1S):41S-48S.

[68] Bassim M, Buss E, Clark M, Kolln K, Pillsbury C, Pillsbury H et al.: MED-EL Combi-40+ cochlear implantation in adults. Laryngoscope 2005; 115:1568-1573.

[69] Kos M, Boex C, sigrist A, Guyot J, Pelizzone M: Measurements of electrode position inside the cochlea for different cochlear implant systems. Acta Otolaryngol 2005; 125(5):474-480.

[70] Zierhofer C. Cochlear implant system. United States Patent 5,983,139. 1999 1999.

[71] Barry J, Blamey P, Martin L, Lee K, Tang T, Ming Y et al.: Tone discrimination in Cantonese-speaking children using a cochlear implant. Clin Linguist Phon 2002; 16(79):99.

[72] Fu Q-J, Zeng F-G, Shannon R, Soli S: Importance of tonal envelope cues in Chinese speech recognition. J Acoust Soc Am 1998; 104(1):505-510.

[73] Gfeller K, Lansing CR: Melodic, rhythmic, and timbral perception of adult cochlear implant users. J Speech Hear Res 1991; 34:916-920.

[74] Gfeller K, Woodworth G, Robin DA, Witt S, Knutson JF: Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. Ear Hear 1997; 18:252-260.

[75] Tong Y, Clark G: Absolute identification of electric pulse rates and electrode positions by cochlear implant patients. J Acoust Soc Am 1985; 77:1881-1888.

[76] Townshend B, Cotter N, van Compernolle D, White RL: Pitch perception by cochlear implantees. J Acoust Soc Am 1987; 82:106-115.

[77] Zeng F-G: Temporal pitch in electric hearing. Hear Res 2002; 174:101-106.

[78] Carlyon R, Wieringen A, Long C, Deeks J, Wouters J: Temporal pitch mechanisms in acoustic and electric hearing. J Acoust Soc Am 2002; 112(2):621-623.

[79] Laneau J, Wouters J: Multi-channel place pitch sensitivity in cochlear implant recipients. JARO 2004; 5(3):285-294.

[80] Jones P, McDermott H, Seligman, Millar J: Coding of voice source information in the Nucleus cochlear implant system. Ann Otol Rhinol Laryngol 1995; 104:363-365.

[81] Green T, Faulkner A, Rosen S: Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous interleaved sampling cochlear implants. J Acoust Soc Am 2002; 112:2155-2164.

[82] Green T, Faulkner A, Rosen S, Macherey O: Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification. J Acoust Soc Am 2005; 118(1):375-385.

[83] Geurts L, Wouters J: Coding of the fundamental frequency in CIS processors for cochlear implants. J Acoust Soc Am 2001; 109(2):713-726.

[84] Fishman K, Shannon R, Slattery W: Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant processor. J Speech Hear Res 1997; 40:1201-1215.

[85] Friesen L, Shannon R, Baskent D, Wang X: Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J Acoust Soc Am 2001; 110(2):1150-1163.

[86] Fu Q-J, Shannon R, Wang X: Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. J Acoust Soc Am 1998; 104(6):3586-3596.

[87] Dorman M, Loizou P, Fitzke J, Tu Z: The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels. J Acoust Soc Am 1998; 104(6):3583-3585.

[88] Griffiths L, Jim C: An alternative approach to linearly constrated adaptive beamforming. IEEE Trans Antennas Propagation 1982; AP-30:27-34.

[89] Widrow B, Stearns S. Adaptive Signal Processing. Englewood Cliffs, NJ: Prentice Hall, 1985.

[90] Van Hoesel R, Clark G: Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients. J Acoust Soc Am 1995; 97(4):2498-2503.

[91] Hamacher V, Doering W, Mauer G, Fleischmann H, Hennecke J: Evaluation of noise reduction systems for cochlear implant users in different acoustic environments. Am J Otol 1997; 18:S46-S49.

[92] Margo V, Terry M, Schweitzer C, Shallop J: Results of take-home trial for a non-linear beamformer as a noise reduction strategy for cochlear implants. J Acoust Soc Am 1995; 98(5):2984.

[93] Wouters J, Berghe J: Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system. Ear Hear 2001; 22(5):420-430.

[94] Hochberg I, Boorthroyd A, Weiss M, Hellman S: Effects of noise and noise suppression on speech perception by cochlear implant users. Ear Hear 1992; 13(4):263-271.

[95] Weiss M: Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor. J Rehab Res Dev 1993; 30(1):117-128.

[96] Yang L, Fu Q: Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. J Acoust Soc Am 2005; 117(3):1001-1004.

[97] Loizou P, Lobo A, Hu Y: Subspace algorithms for noise reduction in cochlear implants. J Acoust Soc Am 2005; 118(5):2791-2793.

[98] Hu Y, Loizou P: A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans on Speech and Audio Processing 2003; 11:334-341.

[99] Berouti M, Schwartz R, Makhoul J: Enhancement of speech corrupted by acoustic noise. Proc IEEE Int Conf Acoust , Speech, Signal Process 1979;208-211.

[100] Toledo F, Loizou P, Lobo A: Subspace and envelope subtraction algorithms for noise reduction in cochlear implants. Conf of IEEE Engineering in Medicine and Biology Society 2003; 3:17-21.

[101] Loizou P, Kasturi K, Turicchia L, Sarpeshkar R, Dorman M, Spahr T: Evaluation of the companding and other strategies for noise reduction in cochlear implants. Abstr of 2005 Conference on Implantable Auditory Prostheses 2005.

[102] Dillon H. Hearing Aids. NY: Thieme, 2001.

[103] Rangachari S, Loizou P, .: A noise estimation algorithm for highly non-stationary environments. Speech Communication 2006; 28:220-231.

[104] IEEE Subcommittee: IEEE Recommended Practice for Speech Quality Measurements. IEEE Trans Audio and Electroacoustics 1969; AU-17(3):225-246.

[105] Geurts L, Wouters J: Enhancing the speech envelope of CIS processors for cochlear implants. J Acoust Soc Am 1999; 105:2476-2484.

[106] Vandali A: Emphasis of short-duration acoustic speech cues for cochlear implant users. J Acoust Soc Am 2001; 109:2049-2061.

[107] Holden L, Vandali A, Skinner M, Fourakis M, Holden T: Speech recognition with the ACE and transient emphasis spectral maxima strategies in Nucleus 24 recipients. J Speech, Lang , Hear Research 2005; 48:681-701.

[108] Turicchia L, Sarpeshkar R: A bio-inspired companding strategy for spectral enhancement. IEEE Transactions on Speech and Audio Processing 2005; 13(2):243-253.

[109] Wilson B, Lawson D, Muller J, Tyler R, Kiefer J: Cochlear implants: Some likely next steps. Annu Rev Biomed Eng 2003; 5:207-249.

[110] McDermott H, McKay CM: Musical pitch perception with electrical stimulation of the cochlea. J Acoust Soc Am 1997; 101(3):1622-1631.

[111] Munson B, Nelson P: Phonetic identification in quiet and in noise by listeners with cochlear implants. J Acoust Soc Am 2005; 118(4):2607-2617.

[112] Qin M, Oxenham A: Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J Acoust Soc Am 2003; 114(1):446-454.

[113] Stickney G, Zeng F-G, Litovsky R, Assmann P: Cochlear implant speech recognition with speech maskers. J Acoust Soc Am 2004; 116(2):1081-1091.

[114] Van Hoesel R, Tyler R: Speech perception, localization, and lateralization with bilateral cochlear implants. J Acoust Soc Am 2003; 113(3):1617-1630.

[115] Kiefer J, Tillein J, von Ilberg C, Pfennigdorff T, Strurzebecher E: Fundamental aspects and first results of the clinical application of combined electric and acoustic stimulation of the auditory system. Cochlear Implants - An update 2002;569-576.

[116] Hartmann WM, Johnson D: Stream segregation and peripheral channeling. Music Percept 1991; 9:155-184.

[117] Hillenbrand J, Getty L, Clark M, Wheeler K: Acoustic characteristics of American English vowels. J Acoust Soc Am 1995; 97:3099-3111.

Device	Processor name	Electrodes		Stimulation
		Number	Spacing	
Nucleus	ESPrit/Freedom	22	0.7 mm	Sequential
Clarion II	Auria	16	1.1 mm	Sequential/simultaneous
Med-El	Combi40+/ Tempo+/ PULSARci[100]	12	2.4 mm	Sequential/simultaneous[1]

Table 1: Characteristics of commercially available cochlear implant devices.

[1] Supported only in the PULSARci[100] processor.

FIGURE CAPTIONS

Figure 1. The channel vocoder analyzer. The signal processing blocks enclosed in the dashed rectangle are used in most cochlear implant devices.

Figure 2. The channel vocoder synthesizer.

Figure 3. Block diagram of the signal processing involved in the CIS strategy. (BPF=bandpass filter, LPF=low-pass filter)

Figure 4. The pulsatile waveforms of the syllable /t i/ obtained at five different stimulation rates [12]. These waveforms were obtained by bandpass filtering the syllable /t i/ into 6 channels, performing envelope detection, and sampling the rectified envelopes at the rates indicated. Only the waveforms for channel 5 (with a center frequency of 3316 Hz) are shown. The bottom panel shows the original speech envelopes of channel 5. This figure shows the effect of stimulation rate in detecting short-duration segments (e.g., burst) of speech. As the pulse rate increases, the burst becomes more distinctive, and perhaps more salient perceptually (Reprinted with permission from [12]. Copyright © 2005, American Institute of Physics).

Figure 5. Consonant and familiar melody identification as a function of envelope cutoff frequency (Hz). Plots show mean identification scores (% correct) for five Clarion-S users fitted with the SAS strategy. Error bars indicate standard errors of the mean. The melodies were taken from a set of 34 simple melodies with all rhythmic information removed [116] and consisted of 16 equal-duration notes synthesized using samples of a grand piano. Prior to the melody recognition test, the subjects selected ten melodies (e.g., “Twinkle Twinkle”, “Old McDonald”) that they were familiar with. The consonant test included 16 consonants in /aCa/ format produced by a male speaker.

Figure 6. Decomposition of a signal (taken from the vowel /a/) into its envelope and fine-structure using the Hilbert transform.

Figure 7. Examples of envelope extraction based on full-wave rectification and low-pass filtering (top three panels) and the Hilbert transform (bottom panel). Envelopes are shown for three different envelope cutoff frequencies.

Figure 8. Vowel recognition as a function of filter spacing (logarithmic, critical-band and mel) for six newly implanted Clarion CII users. The vowel test included vowels in /hVd/ format produced by 7 male speakers, 6 female speakers and 9 children. The stimuli were drawn from a set developed by Hillenbrand *et al.* [117]. Asterisks indicate significant

difference between the scores obtained with critical-band and log frequency spacing. * $p<0.05$, ** $p<0.005$.

Figure 9. Envelope output (bottom panel) produced by the algorithm proposed by Geurts and Wouters [83] for enhancement of F0 cues. Top panel shows the filtered waveform of channel 6 (centered at 1 kHz) taken from the syllable /pa/ produced by a male speaker. Middle panel shows the corresponding envelope extracted using full-wave rectification and low-pass filtering (400 Hz), and subsequently log compressed to fit within a narrow electrical dynamic range. Bottom panel shows the envelope obtained by subtracting an attenuated version of the 50-Hz (flat) log-compressed envelope from the 400-Hz log-compressed envelope.

Figure 10. Envelope outputs obtained at different stages of the MEM algorithm proposed by Vandali *et al.* [6] for enhancement of F0 cues. Top panel shows the input (broadband) signal taken from the vowel /i/ produced by a female speaker (F0=188 Hz). Middle panel shows the full-wave rectified signal of the filtered waveform of channel 3 (centered at 486 Hz). The 50-Hz envelope extracted using full-wave rectification and low-pass filtering is superimposed. Bottom panel shows the envelope output produced by the MEM algorithm. All waveforms are shown prior to compression.

Figure 11. Block diagram of the processing involved in the beamforming strategy based on two microphone inputs (Mic 1 and Mic 2).

Figure 12. The S-shaped input-output function.

Figure 13. Recognition (in terms of percent of words identified correctly) of sentences embedded in +5 dB multi-talker babble by seven Clarion CII patients for two different input-output functions, logarithmic (as used in their daily strategy) and S-shaped (Fig. 12)**.** The S-shaped input-output function (Fig. 12) was implemented using $y = A_1 x^{1.8} + A_2$ as the expansion function for signal levels below the knee point and $y = A_3 x^{-0.0001} + A_4$ as the compression function for signal levels above the knee point, where A_i are constants chosen to limit the acoustic dynamic range to the patient's electrical dynamic range. The differences in the mean scores were found to be statistically significant (***** $p<0.05$).

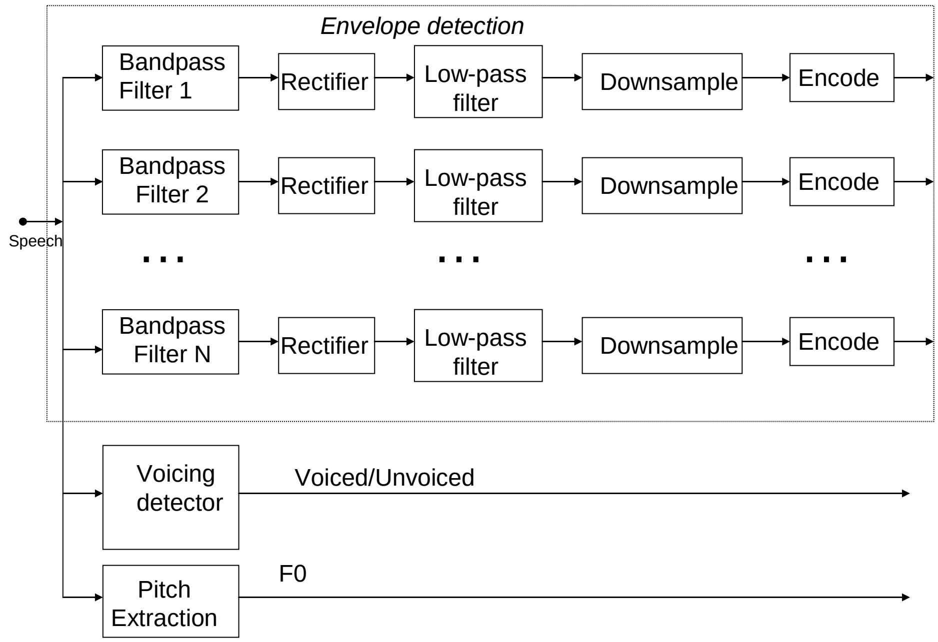

Envelope detection
Speech
Bandpass Filter 1
Rectifier
Low-pass filter
Downsample
Encode
Bandpass Filter 2
Rectifier
Low-pass filter
Downsample
Encode
Bandpass Filter N
Rectifier
Low-pass filter
Downsample
Encode
Voicing detector
Voiced/Unvoiced
Pitch Extraction
F0

Fig. 1

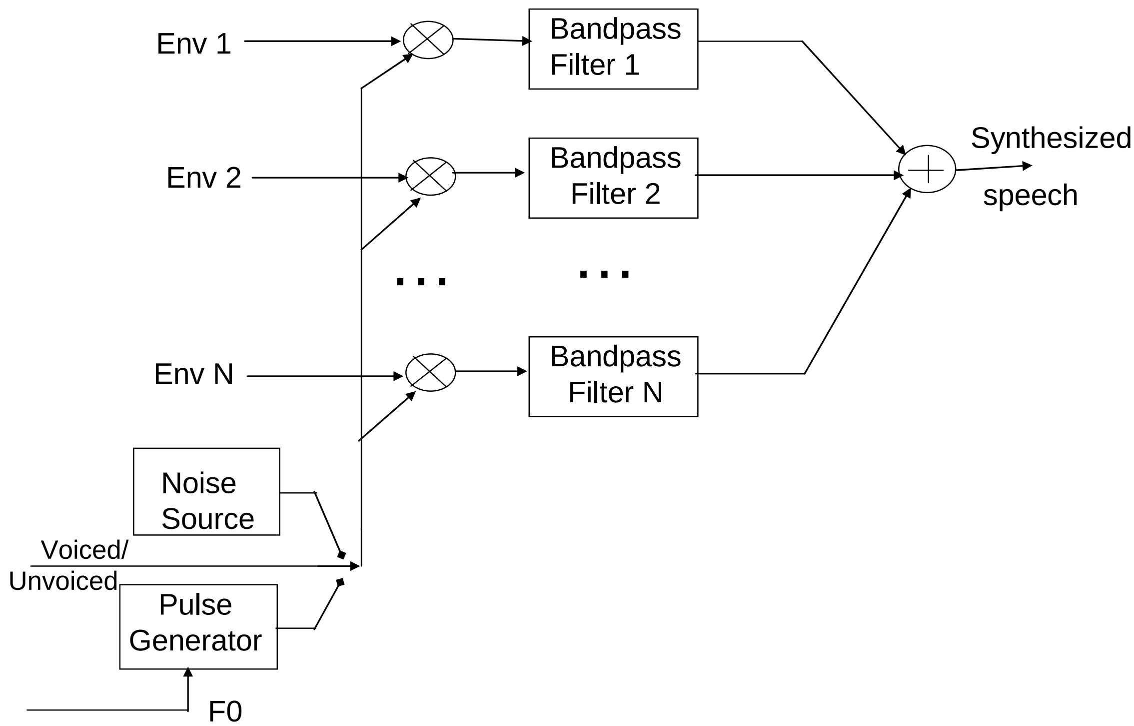

Env 1
Env 2
Env N
Bandpass Filter 1
Bandpass Filter 2
Bandpass Filter N
Synthesized speech
Noise Source
Voiced/ Unvoiced
Pulse Generator
F0

Figure 2.

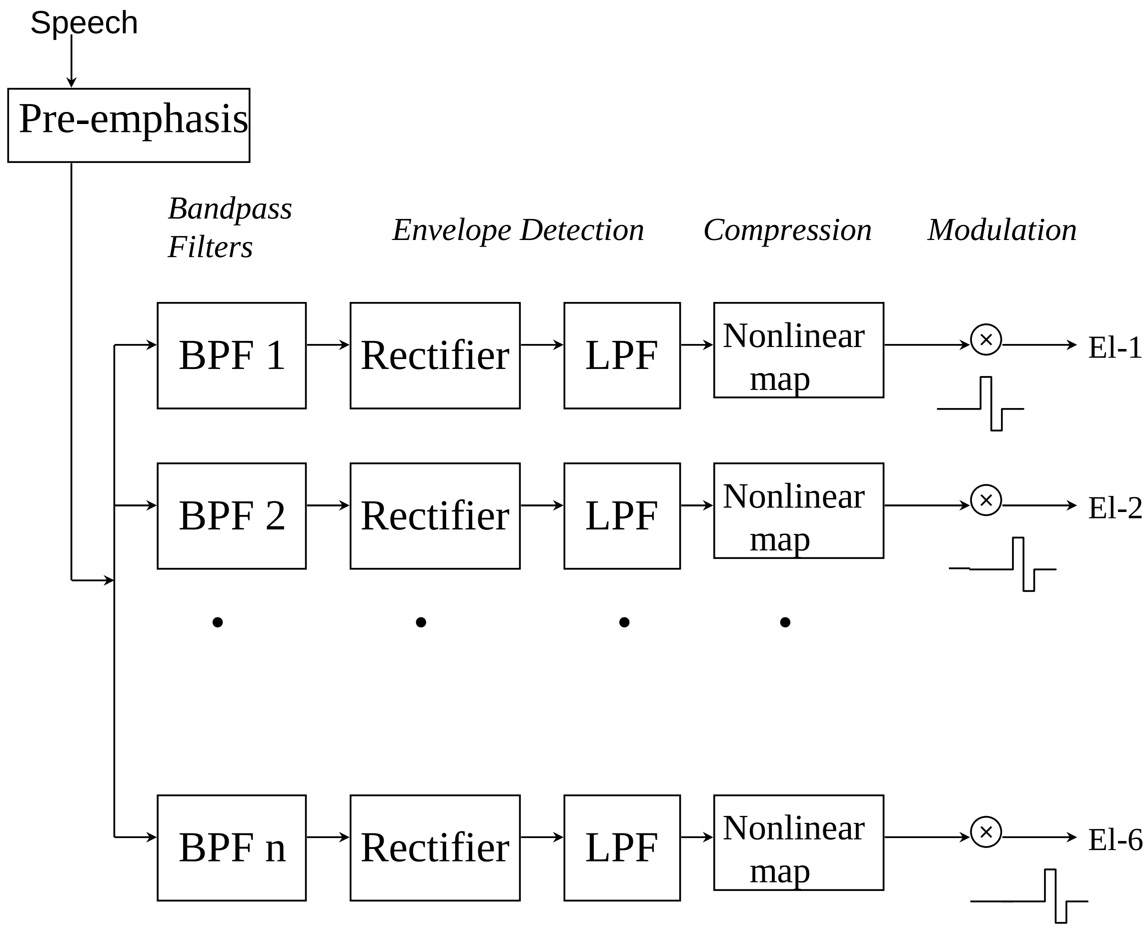

Speech
Pre-emphasis
Bandpass Filters
Envelope Detection
Compression
Modulation
BPF 1
Rectifier
LPF
Nonlinear map
El-1
BPF 2
Rectifier
LPF
Nonlinear map
El-2
BPF n
Rectifier
LPF
Nonlinear map
El-6

Figure 3.

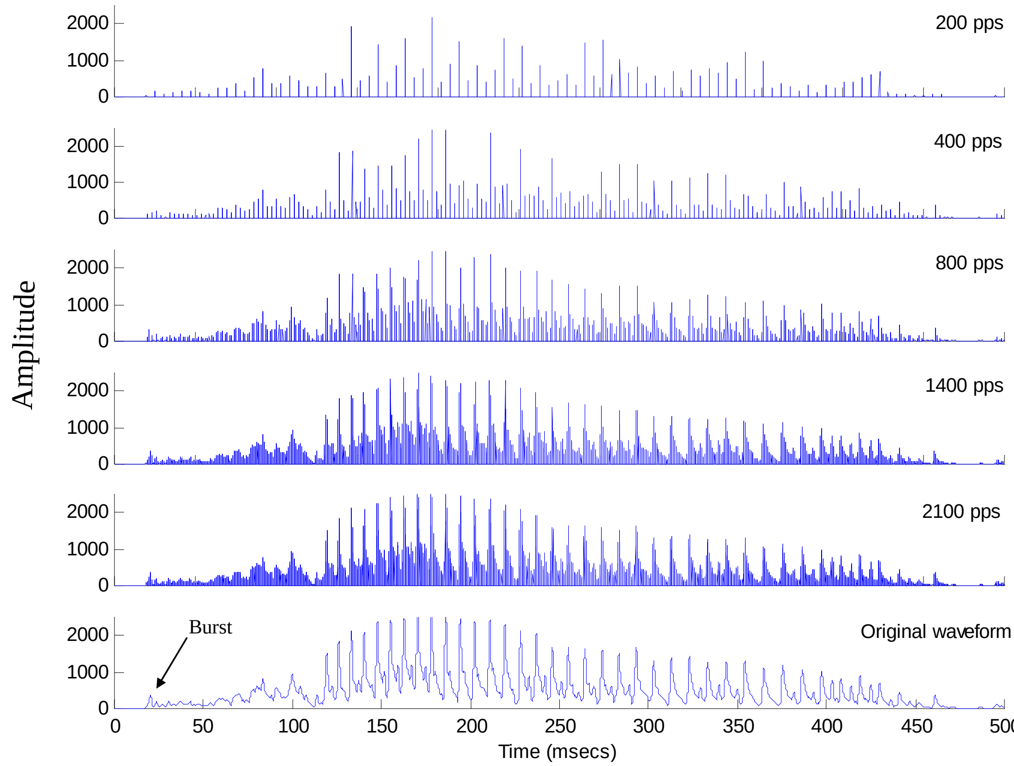
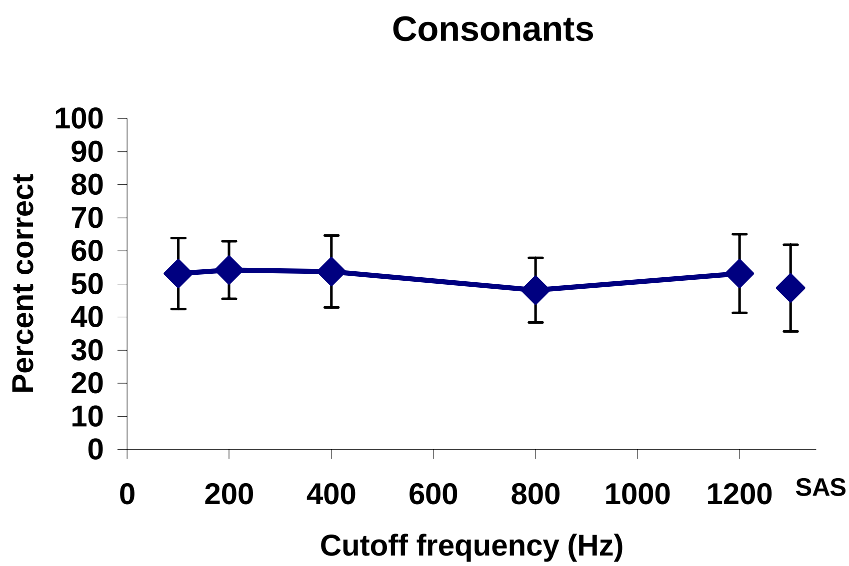

200 pps
400 pps
800 pps
1400 pps
2100 pps
Original waveform
Burst
Amplitude
2000
1000
0
0
50
100
150
200
250
300
350
400
450
500
Time (msecs)

Figure 4.


Consonants
Percent correct
100
90
80
70
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200
SAS
Cutoff frequency (Hz)

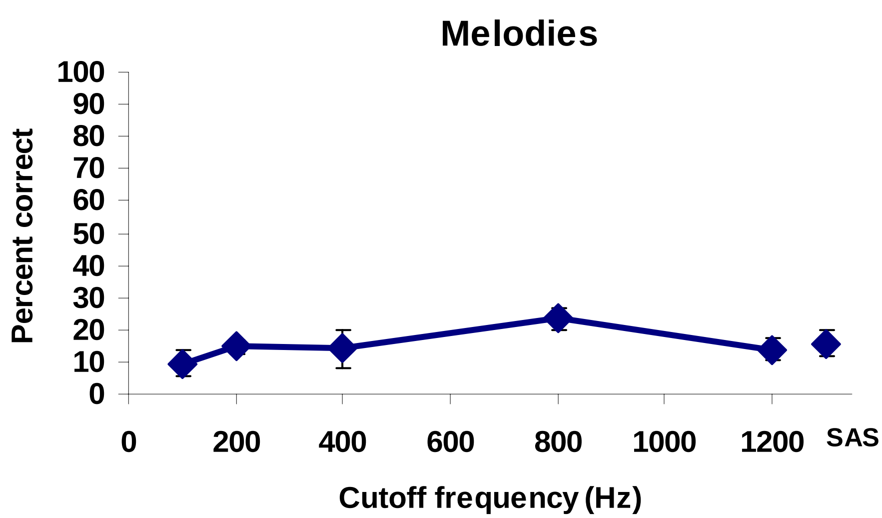

Melodies
Percent correct
100
90
80
70
60
50
40
30
20
10
0
0
200
400
600
800
1000
1200
SAS
Cutoff frequency (Hz)

Figure 5.

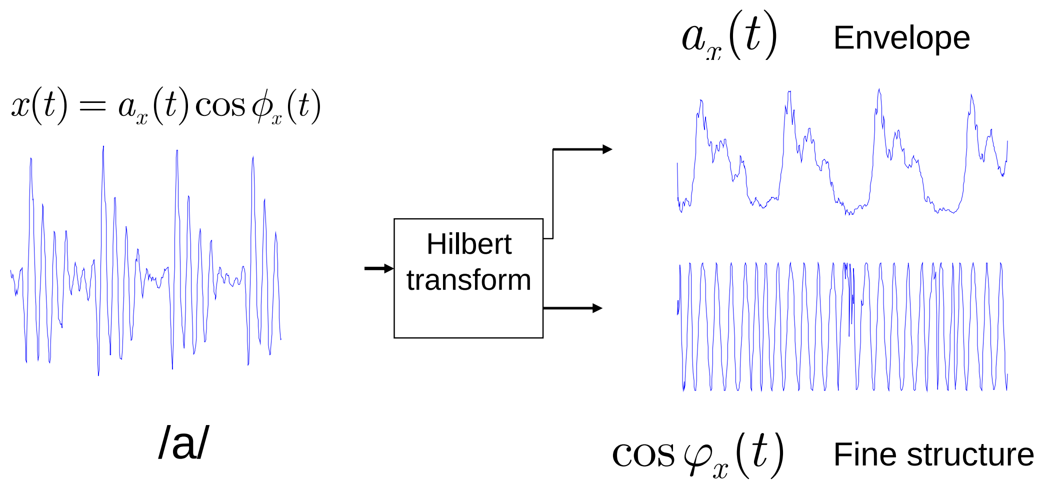

$x(t) = a_x(t)\cos\phi_x(t)$
/a/
Hilbert transform
$a_x(t)$ Envelope
$\cos\varphi_x(t)$ Fine structure

Figure 6.

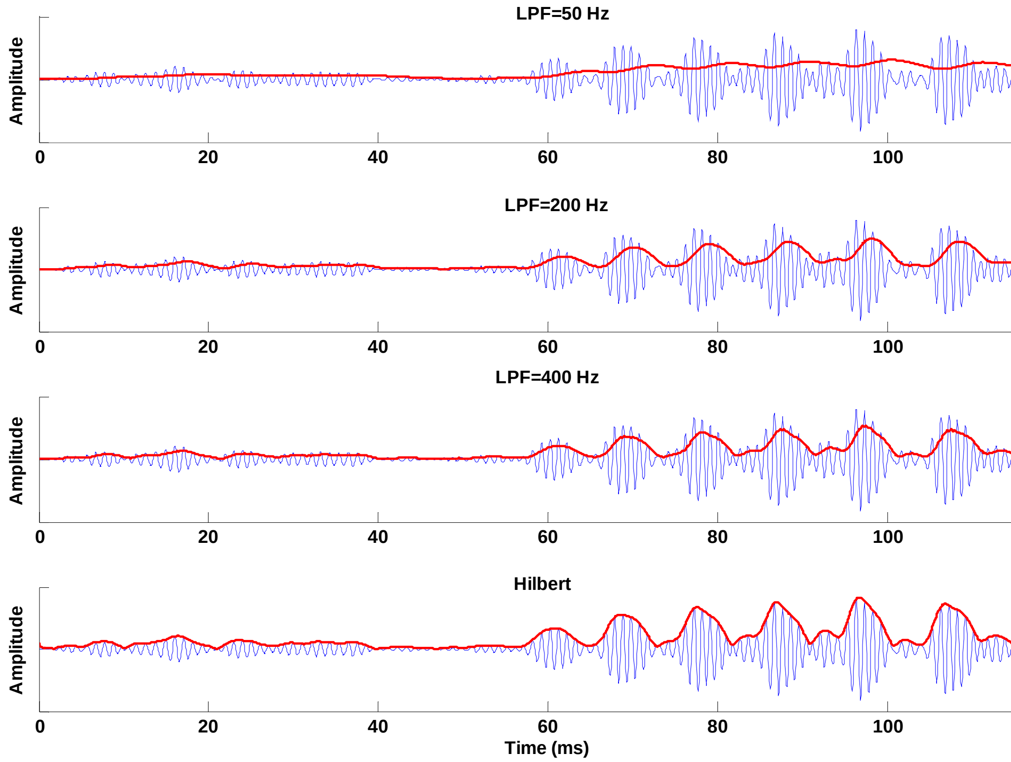

LPF=50 Hz
Amplitude
0
20
40
60
80
100
LPF=200 Hz
Amplitude
0
20
40
60
80
100
LPF=400 Hz
Amplitude
0
20
40
60
80
100
Hilbert
Amplitude
0
20
40
60
80
100
Time (ms)

Figure 7.

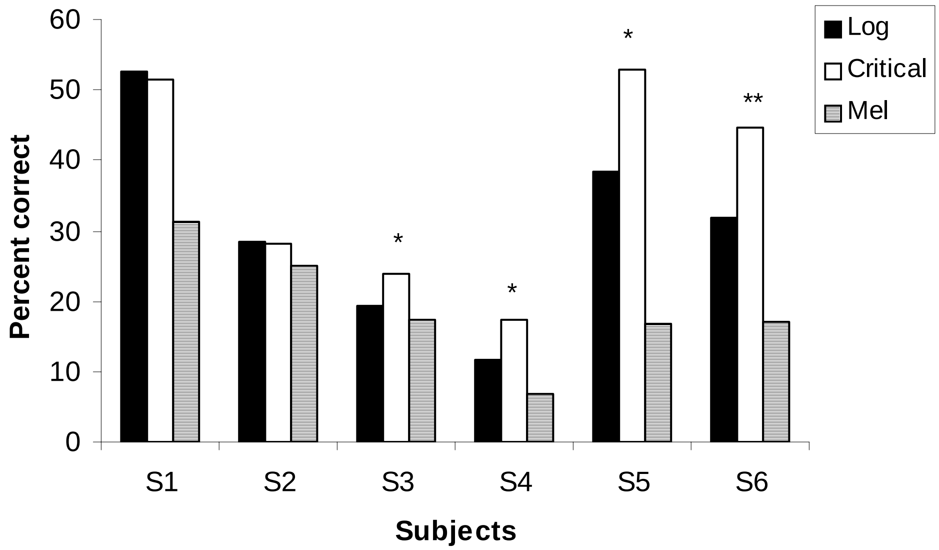

60
50
40
30
20
10
0
Percent correct
Log
Critical
Mel
*
**
S1
S2
S3
S4
S5
S6
Subjects

Figure 8.

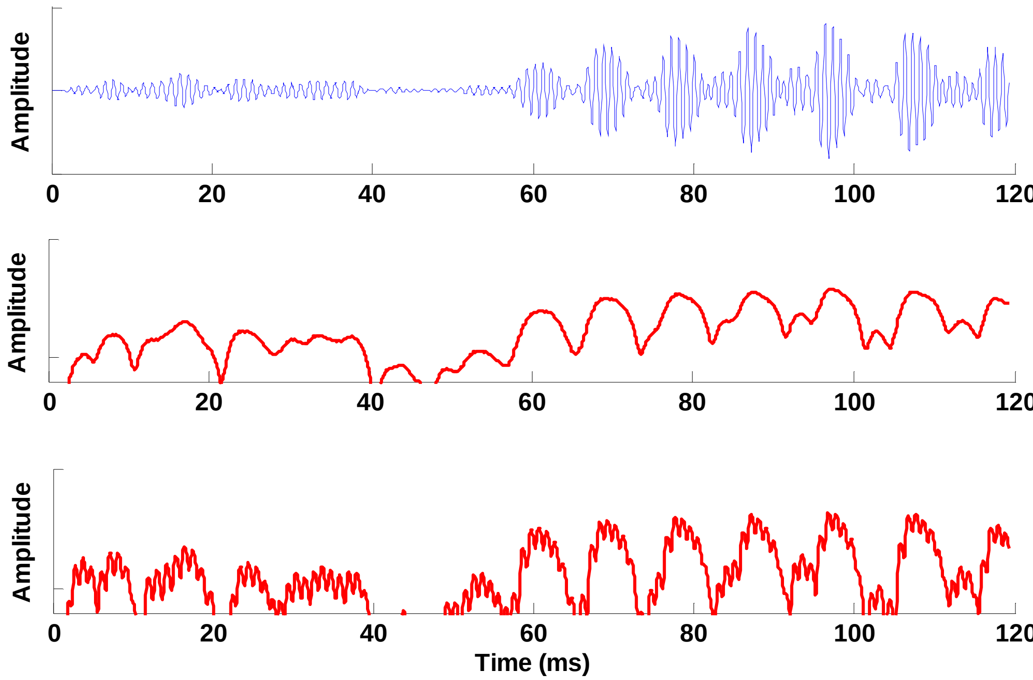

Amplitude
0
20
40
60
80
100
120
Amplitude
0
20
40
60
80
100
120
Amplitude
0
20
40
60
80
100
120
Time (ms)

Figure 9.

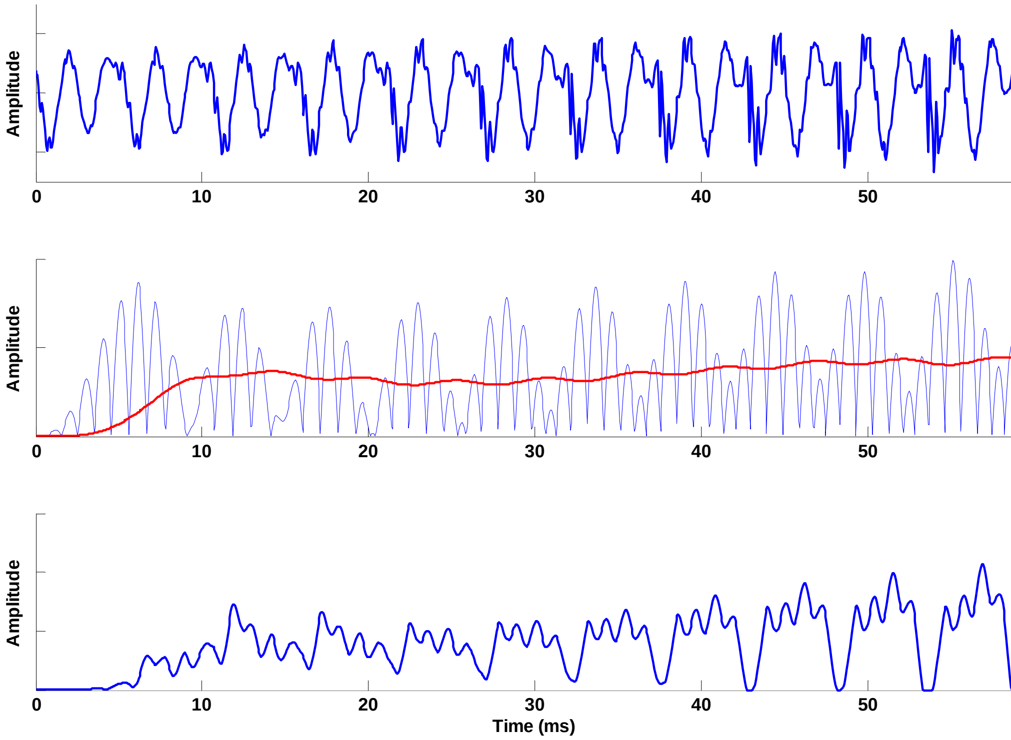

Amplitude
0
10
20
30
40
50
Amplitude
0
10
20
30
40
50
Amplitude
0
10
20
30
40
50
Time (ms)

Figure 10.

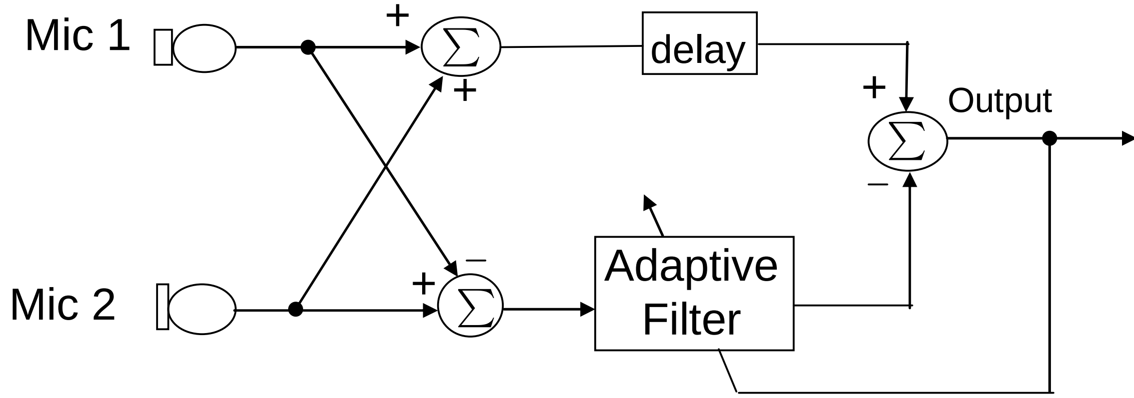
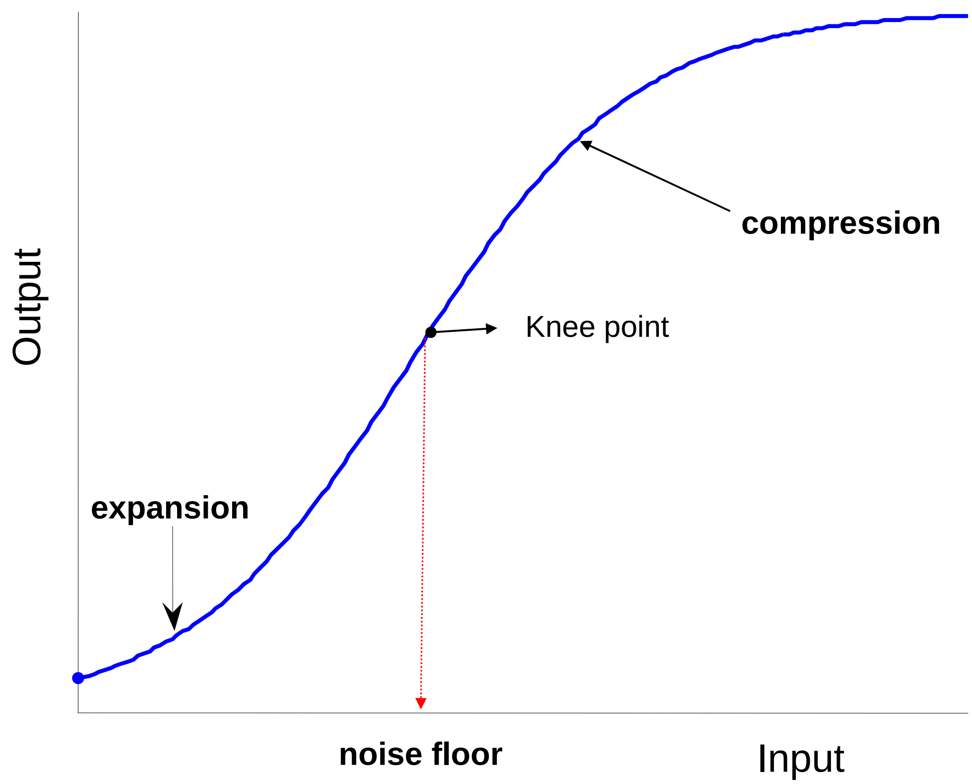


Figure 11.



Output
compression
Knee point
expansion
noise floor
Input

Figure 12.

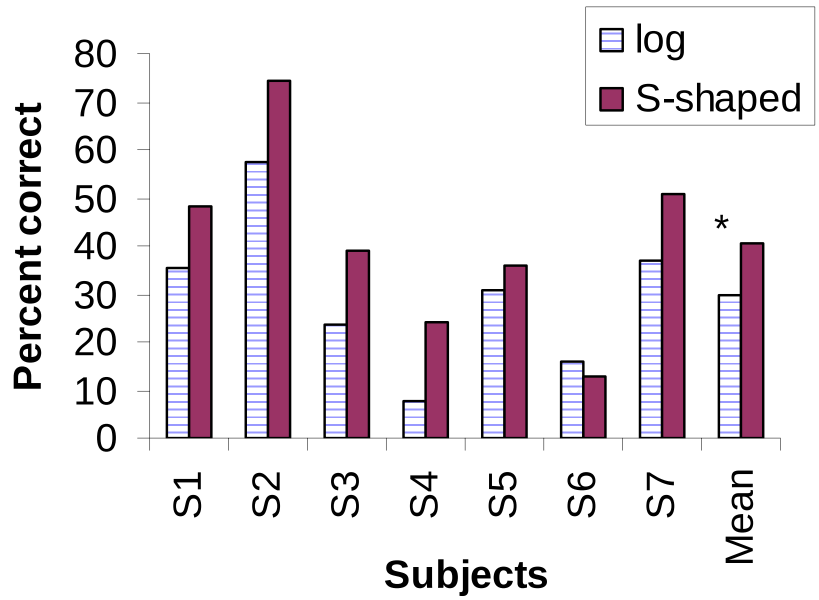

log
S-shaped
Percent correct
80
70
60
50
40
30
20
10
0
*
S1
S2
S3
S4
S5
S6
S7
Mean
Subjects

Figure 13.