Sound Coding in Cochlear Implants: From electric pulses to hearing — **Source link** ⧉

Jan Wouters, Hugh J. McDermott, Tom Francart

Institutions: Katholieke Universiteit Leuven, Bionics Institute

Published on: 10 Feb 2015 - IEEE Signal Processing Magazine (IEEE)

Topics: Cochlear implant, Speech processing, Auditory perception and Audio signal processing

Related papers:

- Better speech recognition with cochlear implants.

- Cochlear Implants: System Design, Integration, and Evaluation

- Electrically evoked auditory steady state responses in cochlear implant users

- Speech recognition with primarily temporal cues.

- Cochlear implants: a remarkable past and a brilliant future

Sound coding in cochlear implants: from electric pulses to hearing

Jan Wouters, Hugh J McDermott, Tom Francart

Abstract—Cochlear implantation is a life-changing intervention for people with a severe hearing impairment. For most cochlear implant (CI) users, speech intelligibility is satisfactory in quiet environments. Although modern CIs provide up to 22 stimulation channels, information transfer is still limited for the perception of fine spectro-temporal details in many types of sound. These details contribute to the perception of music and speech in common listening situations, such as where background noise is present. Over the past several decades, many different sound-processing strategies have been developed to provide more details about acoustic signals to CI users. In this article, progress in sound coding for cochlear implants is reviewed. Starting from a basic strategy, the current commercially most-used signal processing schemes are discussed, as well as recent developments in coding strategies that aim to improve auditory perception. This article focuses particularly on the stimulation strategies, which convert sound signals into patterns of nerve stimulation. The neurophysiological rationale behind some of these strategies is discussed, and aspects of CI performance that require further improvement are identified.

I. INTRODUCTION

The cochlear implant (CI) is the most successful man-made interface to the human neural system; i.e., a machine-brain interface. The auditory nerve is stimulated electrically which leads to a partial restoration of auditory perception for persons with severe hearing

Jan Wouters and Tom Francart are at KU Leuven, Dept. Neurosciences, ExpORL, Leuven, Belgium

Hugh McDermott is at the Bionics Institute of Australia and the University of Melbourne, VIC, Australia.

impairment. Speech understanding of CI recipients in quiet environments can be very good, but is considerably worse than that of normal-hearing listeners in realistic listening situations. Typically the presence of background noise greatly reduces the performance of CI systems. For example, the signal-to-noise ratio required for many CI users to attain 50% speech understanding is about 15 dB higher than for normal-hearing listeners.

At present more than 300,000 people worldwide with severe hearing impairment, of whom 80,000 are young children, have received CIs. CIs are a life-changing intervention [1] and the proportion of implanted children (aged below 2 years) is increasing due to the increasing deployment of neonatal hearing-screening programs in many countries. Early implantation can give profoundly deaf children access to important information to process auditory signals and master spoken language skills at a young age. In many countries a single CI is reimbursed by health insurance organizations, and in some countries also a second CI, primarily for children. About 80% of normally developing, severely hearing-impaired children with a CI eventually participate in the mainstream educational system.

Apart from the technological and surgical progress that has made cochlear implantation the success it is today, the preformed cochlear duct and the ease of surgical access via the middle ear have played a role in its proliferation and progress. How CIs work has been described before in several articles; e.g., in [2]–[4]. This article focuses on a review of stimulation strategies. These are the techniques which convert sound signals picked up by a microphone into patterns of electric stimuli that activate the auditory nerve. The remainder of this section provides a short overview on how we hear and how a CI works.

In the normal auditory system, sound is captured and transmitted by the outer ear, predominantly the pinna (external ear) and ear canal, and then transformed in the middle ear (via the ossicles – small bones which have a mechanical impedance-matching function) to movement of the fluids and membranes in the cochlea, or inner ear. The cochlea has a spiral structure typically about

10mm wide and 5mm high. Within the cochlea, there are numerous transducer structures – the inner and outer hair cells – which have stereocilia that are deflected in response to incoming sound waves. In a healthy ear, movement of the stereocilia of inner hair cells leads to streams of action potentials in the auditory nerve fibers. This electrical activity has patterns with temporal and tonotopic characteristics that ultimately enable identification and interpretation of sounds, including music, speech, and language, at higher neural levels [5]. Temporal information about sound signals is carried through the precise timing of action potentials both within and between nerve fibers, whereas spectral information is represented mainly in the spatial distribution of activity across the neural population; the latter is referred to as the tonotopic organization of auditory nerve.

The most common cause of deafness is damage to or loss of the stereocilia and hair cells, resulting from infections, trauma, exposure to high levels of noise, side-effects of certain drugs, and a range of physiological disorders. Hearing impairment may be acquired by adults who previously had normal hearing, or it may be present at birth. In many cases, the degree of hearing loss becomes progressively worse over time. When the hair-cells are absent or extensively damaged, the transduction of the acoustically induced motion in the cochlea to neural action potentials is disrupted. If the resulting hearing loss is severe, the amplification that can be provided by acoustic hearing aids may be insufficient to restore satisfactory perception of sounds.

A CI bypasses the deficient transducer structures and produces action potentials at the auditory nerve sites (or the residual neurons, depending on the degree and type of pathology) using direct electrical stimulation. Most of today's CI systems have an external and an internal part. The external part consists of a behind-the-ear (BTE) device connected to an external transmission coil which provides a radio-frequency (RF) link to a matching coil in the internal part, the implant. The implant consists of a miniature enclosure containing electronics connected to a number of electrodes. There are one or more reference electrodes on the enclosure or on a separate lead, and there is an array of multiple intracochlear electrodes, between 12 and 22 depending on the manufacturer and implant type. The stimulation currents flow between selected electrodes to activate the neural structures near the electrode-neuron interface. The electrode array is surgically inserted into the cochlea. Implantation of the complete internal system takes approximately 3 hours.

As illustrated in figure 1, sound is captured in the external BTE device by a microphone system (one or more microphones). Pre-processing is applied, for example, to optimize the input dynamic range relative to input signal levels and to adjust the spectrum shape using a pre-emphasis filter. In some systems there is also fixed or adaptive beamforming or other types of noise-reduction processing that typically exploit the differences between signals obtained from several microphones to enhance desired sounds while suppressing competing noise. The stimulation 'strategy' refers to the transformation of the input sound signal into a pattern of electrical pulses. Digital specifications of the required stimulation patterns produced by the stimulation strategy are coded in the transcutaneous RF transmission. The RF signal also provides power to the internal part. The specifications of the stimulation are decoded from the RF signal. The electronics of the implant include one or more current source(s) to deliver the electrical stimulation pattern to the electrode channels. A channel is defined as a set of two or more electrodes with currents flowing between them. The term "monopolar" stimulation is used to describe current passing between an intracochlear electrode and a remote reference electrode, whereas "bipolar" refers to stimulation current passing between two intracochlear electrodes. The implant also has measurement amplifiers on-chip for the recording of evoked neural activity from non-stimulating electrodes via outward telemetry.

A few weeks after implantation and at regular intervals thereafter, stimulation levels are adjusted ("fitted") to the individual patient. In each fitting session a patient-specific 'map' is set up containing all stimulation parameters. For each channel, minimal levels of stimulation (min) and levels of maximal comfortable loudness (max) are determined. In some cases also the shape of the growth function between min and max that converts the input acoustic levels to electric stimulation levels is determined. During a fitting session impedances of the stimulation channels can be measured (which may lead to deactivation of some electrodes if faults are detected) and parameters of the pre-processing stage can be adjusted [6].
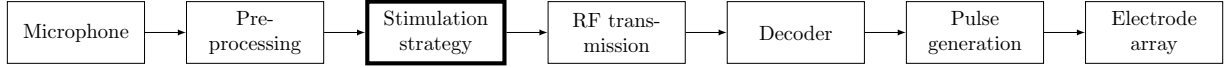
Fig. 1. Block diagram of complete CI system

Today's CIs have a high power consumption compared to hearing aids, which means that the batteries largely determine the size of the BTE sound processor, making it cumbersome and unsightly for users. This also means that users need to replace batteries often, typically every day with rechargeable cells and every two days for primary cells, which may be expensive and inconvenient. Therefore currently a lot of research and development is going into reducing power consumption. Another major comfort improvement would be a totally implantable CI. The major challenge of a totally implantable system is the capture of airborne target sound with microphones and accelerometers, while suppressing the high levels of unwanted noise emanating from inside the human body.

A major technical and basic scientific challenge, and the subject of this article, is the translation of the captured sounds, particularly speech or music, to electrical stimulation patterns across the intracochlear channels to optimize auditory perception and interpretation. Historically, the objective of CIs has mainly been to improve speech intelligibility. Speech intelligibility is determined by spectral and temporal characteristics of the acoustic signal. The spectral information is coarsely coded through multi-channel representation following the auditory system's natural tonotopic organization; i.e., acoustic spectral information is normally represented from low to high frequency in a corresponding spatial progression within the cochlea. Temporal speech information is commonly classified into three categories: (1) the speech envelope, defined as the fluctuations in overall amplitude at rates between 2 and 20 Hz, (2) the periodicity from around 50 to 500 Hz, usually due to the fundamental frequency (F0), and (3) temporal fine structure (TFS). TFS can be defined as the variations in wave shape within single periods of periodic sounds, or over short time intervals of aperiodic ones. It has dominant fluctuation rates from around 500 Hz to 10 kHz. Alternatively, from a perceptual point of view, TFS can be defined as the fast fluctuations in a signal that can be used by normal-hearing listeners to perceive pitch, to localize sounds, and to binaurally segregate different sound sources. The fine structure is modulated in amplitude by the temporal envelope and periodicity. For speech sounds, F0 is the frequency at which the vocal cords vibrate. Recently the transmission of F0 information, related to pitch perception, has attracted a

lot of interest because of the need to improve perception of music and tonal languages with CIs [7].

It is not easy to define pitch. It is defined by the American National Standards Institute (1994) as "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from high to low". From a musical point of view, it can be defined as "that attribute of sensation whose variation is associated with musical melodies". For periodic sounds, pitch is the perceptual counterpart of the fundamental frequency (F0), leading to the alternative definition that "a sound has a certain pitch if it can be reliably matched by adjusting the frequency of a pure tone of arbitrary amplitude" [8]. While F0 is a purely physical signal attribute, i.e., the frequency of the first harmonic of a complex tone, pitch is a perceptual attribute which arises after processing in the brain, and can not always be easily linked to physical signal attributes. Typical relevant signals that elicit a pitch percept are spoken vowels and sustained sounds produced by musical instruments. Aperiodic sounds can also elicit a pitch percept, but it is less well-defined.

In the normal auditory system, pitch is determined by three different physical cues: (1) place of stimulation in the cochlea, (2) TFS, and (3) periodicity. The cochlea is tonotopically organized, so sounds with different spectral content will activate distinct neural populations, leading to different percepts. In the case of a simple sinusoid there is a one-to-one relation between frequency and place of stimulation. For harmonic sounds, the situation is more complicated: the place of stimulation of the lowest harmonic still has a one-to-one relationship with F0, but the higher harmonics do not by themselves directly code F0. The spectral pitch mechanism is not very sensitive to small changes in F0, and the change in percept associated with a pure change in spectral pitch has been reported to correspond more to a change in timbre than a change in pitch [8]. Timbre, also called "tone color", "tone quality", or "brightness" is the quality of a sound that distinguishes different types of sound production, such as voices or musical instruments. The American Standards Association (1960) defines timbre by exclusion as "that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar".

The second pitch-related cue, TFS, can yield a strong and tonal pitch percept when individual harmonics are

3

coded by discrete neural populations and their frequency is lower than the maximal frequency to which neurons can phase-lock (around 1500 Hz), i.e., the neural action potentials tend to occur during a particular phase of the oscillation. When multiple harmonics excite the same hair cells and therefore neurons, information is carried mainly by the aggregate stimulation pattern. This is likely to happen at higher frequencies because harmonics of a given F0 are spaced linearly in frequency whereas the auditory periphery is organized logarithmically. This leads to unavailability of the TFS of individual harmonics. However, the auditory system can still make use of a third physical cue: the periodicity of the combined harmonics, which corresponds to the F0. Perception of periodicity is limited to around 300-500 Hz [9]. Periodicity pitch is weak compared to TFS pitch. For good pitch perception across a wide variety of types of sound, all three cues are needed.

Pitch perception with CIs is extremely poor. This is due both to limitations at the interface with electrical stimulation (spread of excitation) and to imprecise coding of temporal cues. The large spread of excitation in the cochlea and the small number of channels to code the low frequencies with electrical stimulation reduces the spectral resolution and therefore the precision of spectral pitch. Another limitation with electrical stimulation is the inability of CI users to perceive TFS. Therefore the only remaining mechanism is periodicity pitch perception, which is much weaker than TFS pitch and limited by the maximum frequency at which pitch changes are perceived, around 300 Hz [9]. Furthermore, temporal envelope fluctuations are not always accurately coded by current sound-processing strategies.

Currently, an increasing number of people are being implanted bilaterally, especially children. Also, due to relaxed implantation criteria, an increasing number of people can make use of bimodal stimulation. These CI recipients have residual hearing in the non-implanted ear which can be aided with an acoustic hearing instrument. Listeners with bilateral CIs or using bimodal stimulation can potentially perceive ITDs. Therefore another topic of intensive research is binaural hearing and the preservation of binaural cues in applications with bilateral and bimodal devices. Interaural time differences (ITDs), the difference in arrival time between the ears, are important binaural cues for NH listeners to localize sound sources and to separate multiple sound sources such as speech and noise. The latter is called binaural unmasking. ITDs range from 0 μs for sounds in front to around 700 μs for sounds from the side of the head. NH listeners can use ongoing temporal cues that are present in both the fine

structure and the envelope of sound signals [10]–[12], and temporal cues in the onset of signals [11].

In the following, an overview is given of basic stimulation strategies (section II) and sound-processing strategies implemented in commercial sound processors (section III), focusing on the 4 processing strategies that are used by more than 90% of CI recipients worldwide. Next the functional concepts underlying 4 examples of promising experimental processing strategies are outlined (section IV). In the general discussion (section V), important challenges are addressed, and conclusions are made. Extended use is made of illustrations to facilitate the comprehension of the physical differences between strategies. Sound coding approaches and applications across the majority of different cochlear implant types are covered.

II. BASIC STIMULATION STRATEGIES

Historically, the first main types of stimulation strategies can be classified as feature extraction strategies. In such strategies, estimates of F0 and formants F1 and F2 of speech signals are calculated in real-time. Formants are peaks in the spectral envelope, corresponding to resonances of the vocal tract. Formants are used by the auditory system to identify sounds such as vowels. The formant information is used predominantly to stimulate channels corresponding to F1 and F2. The F0 is used to control the pulse rate. The outcomes in speech understanding of these schemes are, on average, lower than those of more recent schemes, and therefore they are not normally used any more in commercial processors [2], [4].

A simple strategy, widely used in CI signal processing, is continuous interleaved sampling (CIS), see figure 2 and [13]. CIS is based on a running spectral analysis of the pre-processed digital input sound signal performed by a bank of band-pass filters or a fast Fourier transform (FFT). The filter bank has an overall bandwidth from approximately 100 to 8000 Hz, and the number of filters usually equals the number of stimulation channels at the electrode array-neuron interface. The filters have partially overlapping frequency responses and bandwidths that generally become broader with increasing frequency. Each filter is assigned to (at least) one intracochlear electrode following the frequency-place tonotopic organization of the cochlea. Although the correspondence of signal frequencies and filter bank outputs to depth of electrode insertion follows the tonotopy, the signal is not necessarily delivered to the normal anatomical or neurophysiological place because generally electrode arrays do not allow insertion beyond the anatomical

position corresponding to acoustic frequencies lower than 500-1000 Hz. However, studies have shown that with time of use of the CI, cortical plasticity can partly compensate for this mismatch [14]. Also, manufacturers have recently introduced CI systems with electrode arrays that allow deeper insertion depths to facilitate more apical stimulation. The rationale for more apical stimulation as well as a review of results is given in [15].

After the filter bank, the magnitude of the envelope in each channel is determined (block 4 in figure 2), for instance with an envelope detector using rectification or using a Hilbert transformation followed by low-pass filtering. The filter cut-off frequency should at least comprise the modulation frequencies below 20 Hz to preserve the speech envelope information. Typical cut-off frequencies are between 125 and 300 Hz. When spectral estimates are obtained via a FFT, magnitudes corresponding to each of the electrodes are obtained from the allocated FFT bins, summing the powers across adjacent FFT bins depending on the filter bandwidths.

The stimulation levels are related to the magnitudes of the band-limited input signals by user-specific functions. The output of the envelope detector is transformed to a value between the min and max levels according to a non-linear compression function because the electrical stimulation dynamic range (≈ 10 dB) is much smaller than the input dynamic range of the pre-processor (block 5 in figure 2). This mapping is patient-specific because min and max can vary widely across patients, stimulation channels, and electrode configurations (due to the status of the neural structures at the electrode-neuron interface and higher-level neural structures). Next, these transformed magnitudes modulate carrier waves of electrical pulses. Commonly, symmetric biphasic pulse trains are used in commercial CIs, and magnitude is coded by varying the pulse amplitude and/or the pulse width.

For practical reasons (many CIs have only 1 current source) but also for limiting across-channel interactions, pulsatile stimuli are used in an interleaved stimulation scheme (i.e., only one pulse is delivered at any time). Furthermore, all channels are activated in a temporally non-overlapping sequence, and a fixed stimulation carrier rate is used (typically 500 to 2000 pulses per second (pps)), with the total pulse rate equal to the number of active channels times the channel rate. The latter has no relationship with auditory neurophysiology, as neural fibers do not fire at fixed rates and stimulation rates are generally far higher than neural spike rates. However, it is simple from a signal-processing point of view and provides most CI recipients with adequate perception of sounds.

This strategy can faithfully represent the temporal speech envelope in the electrical stimulation patterns, leading to effective transmission of envelope information, which is a necessary condition for speech perception. CIS has been described by Wilson et al in 1991 [13]. Essentially the same sound processing scheme, albeit with relatively low stimulation rate (around 300pps), was previously used in an earlier French CI system [16].

In general, evaluation (and comparison) of strategies is mainly based on behavioral performance measures on identification and discrimination tasks related to speech understanding, music and tone perception, directional hearing, sound quality and preference measures. At present no validated model of these measures, nor objective neurophysiological markers, exists for electrical stimulation. So behavioral tests are the reference evaluation approach.

In the following, a range of stimulation strategies for CI sound coding is described. Along with a description of the technical features of each strategy, we highlight the rationale behind the strategy, where one can be identified. We also review selected published outcomes for speech understanding and, if relevant and available, also for music or tone perception. Some of these schemes are widely used in commercial processors, while others are experimental and still in development.

III. SOUND PROCESSING STRATEGIES IMPLEMENTED IN COMMERCIAL SOUND PROCESSORS

Since the introduction of the first stimulation strategies in commercial multi-channel CIs over 30 years ago, a number of diverse sound-processing strategies have been devised and evaluated. These strategies focus on better spectral representation, better distribution of stimulation across channels, and better temporal representation of the input signal. The 4 most commonly used strategies will be described. These are ACE (Advanced Combination Encoder) with channel selection based on spectral features, MP3000 (named after the MP3 digital audio format) with channel selection and stimulation based on spectral masking, FSP (Fine Structure Processing) based on enhancement of temporal features, and HiRes120 (High Resolution) with temporal feature enhancement and current steering to improve the spatial precision of stimulus delivery.

An overall outline of the sound-processing steps for the different stimulation strategies, with common and differentiating parts, is shown in the block diagram of figure 2. The outputs of the strategies are shown as electrodograms. An electrodogram is similar to a spectrogram, but the vertical axis indicates channel number

rather than frequency, and biphasic current pulses are represented as vertical lines with amplitudes between 0 (min level of map) and 1 (max level of map). In figures 3, 4, 5, and 6 respectively, electrodograms are shown of the synthesized vowel ah, a naturally spoken sentence in quiet taken from the HINT corpus [17], a selected word from the same sentence, and the same sentence in steady noise with a speech-weighted spectrum at a signal-to-noise ratio of 10 dB. The CIS and ACE based strategies and FSP were implemented in MATLAB. HiRes120 was implemented in C. The base stimulation rate per channel for ACE/CIS was 900 pps, for FSP 1500 pps, and for HiRes120 1856 pps.

Four manufacturers of CI systems are on the international market (with implementations of strategies described in this review): Cochlear (ACE, MP3000), Advanced Bionics (HiRes120), Med-El (FSP), and Oticon Medical / Neurelec.

A. ACE

ACE is the sound-processing scheme currently used by most recipients of CI systems manufactured by Cochlear. It is functionally very similar to the Spectral Maxima Sound Processor (SMSP) [18] and the Speak scheme [19] used with previous models of Cochlear CIs. The original development of the SMSP arose from the observation that sound-processing schemes based on presentation of selected acoustic features of speech signals were technically and perceptually limited. As mentioned above, most of those schemes provided CI users with partial information primarily about the two lowest speech formants (F1, F2) and the fundamental frequency (F0) [20]. While those schemes enabled many recipients to understand speech adequately in favorable listening conditions, performance was degraded by even moderate levels of background noise. This was mainly because of the technical difficulty of estimating parameters corresponding to the selected speech features in real time when the signal-to-noise ratio is poor. The SMSP and its successor schemes Speak and ACE (as well as closely related schemes provided by other CI companies) attempt to provide CI users with information about salient aspects of the acoustic spectral shape without explicitly estimating speech features. Indeed, there is no inherent assumption that the sound signals processed for CI recipients contain any speech.

ACE has many signal processing modules in common with CIS and almost all other current CI processing schemes (blocks (1-6) in Fig. 2). However, the major distinction with CIS, is that on each stimulation cycle, only a subset of the available electrodes is selected.

This is indicated by the "channel selection" block (7) in figure 2. The subset comprises the n electrodes that have the highest short-term signal levels; thus, this type of processing scheme is sometimes referred to as n-of-m. In Cochlear CI systems, typically 8 electrodes from the available set of 22 are selected for stimulation at a rate of 900 pps per electrode, although stimulation parameter values can be varied to optimize performance for individual recipients..

Figures 4-6 show that ACE represents some speech formant peaks and formant trajectories (i.e., changes in formant frequency over time) more distinctly than CIS, particularly when background noise is present. Because frequency bands containing relatively low signal levels are not represented in the stimulation pattern, ACE can enhance certain spectral features when perceived by CI users. This may be one reason that several studies of speech understanding have demonstrated slightly higher scores for ACE than CIS [21]. For example, Skinner et al [22] reported that CI listeners in two separate comparison studies scored about 6-9 percentage points higher, on average, in sentence tests when using ACE rather than CIS.

B. FSP

Although most CI users obtain good performance with sound-processing schemes such as ACE and CIS, unfortunately intelligibility of speech in competing noise is often unsatisfactory, and essential components of musical sounds – particularly pitch – are poorly perceived. Part of the reason may be the lack of TFS in the stimulation patterns. In general, TFS is characterized by the rapid amplitude variations within each of the band-pass filters that implement the initial spectral analysis of sound signals. In contrast, only the slowly varying envelope of the band-limited signals is used to modulate stimulation levels in schemes such as ACE and CIS.

In the quest for improved CI sound processing, numerous attempts have been made to introduce TFS cues explicitly. One such scheme, currently the default in systems manufactured by Med-El, is known as FineHearing Technology. The aim of FineHearing Technology is to represent TFS information present in the lowest frequencies of the input sound signals by delivering bursts of stimulus pulses on one or several of the corresponding CI electrodes. These bursts can consist of one or more stimulation pulses and are derived indirectly from the band-limited acoustic signals [23]. Each burst is triggered by a positive zero-crossing in the bandpass-filtered waveform, while stimulus pulses within the burst are delivered at a constant, high rate that
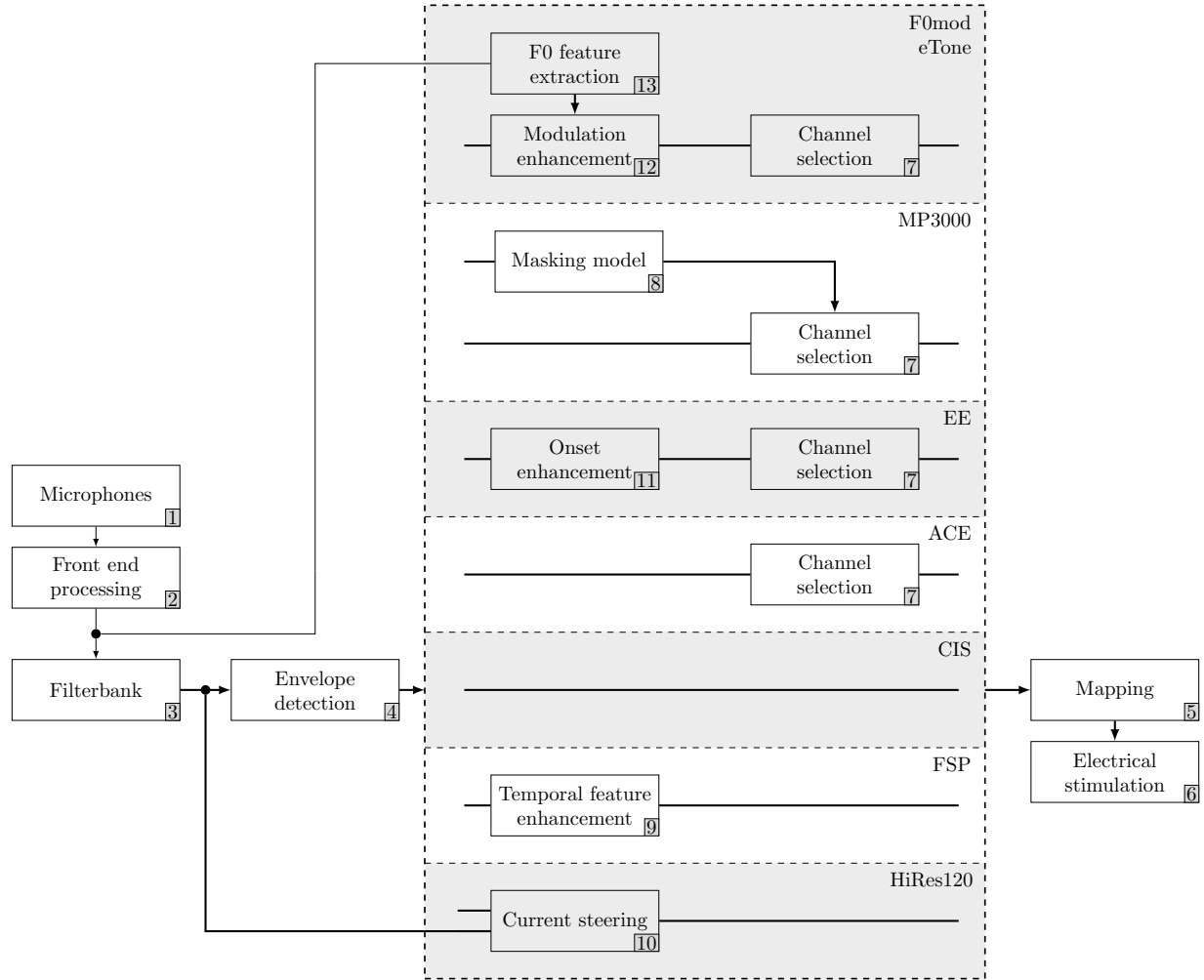
Fig. 2. Block diagram of all monaural strategies discussed. Common elements are shown on the left and right hand side, while strategy-specific elements are shown in the dashed box in the middle.

depends on user-specific settings (typically 5-10 kpps). The duration and amplitude-envelope modulation of each burst are predetermined to approximate the filtered acoustic waveforms after half-wave rectification. These bursts contain information about the TFS in the lower frequency bands that is not available in the envelope of those signals, potentially leading to improved perception for CI users. In essence, FineHearing Technology uses variable-rate coding to provide additional information about the TFS of the signal. Med-El has released the FSP (Fine Structure Processing), FS4, and FS4-p coding strategies. These strategies differ mainly in the frequency range across which TFS is presented. While in FSP, TFS is represented for frequencies up to 350-500 Hz, in FS4 and FS4-p, TFS is presented for frequencies up

to 750-950 Hz. In order to faithfully represent F0, these strategies cover an input frequency range from 100-8500 Hz by default, which differs from the CIS strategies from Med-El (250-8500 Hz). The FSP coding strategy is illustrated in Figures 3-6, where TFS pulse patterns are delivered by the two most-apical electrodes while the remaining electrodes convey CIS-like pulse trains.

In a number of studies several of the coding strategies available in the Med-El system have been compared. Most published studies evaluating the perception of CI recipients when using FSP relative to other sound-processing schemes (e.g. CIS) are difficult to interpret. In some cases, the sound-processor hardware and settings such as the input frequency range were altered at the same time as the processing algorithm was changed. In
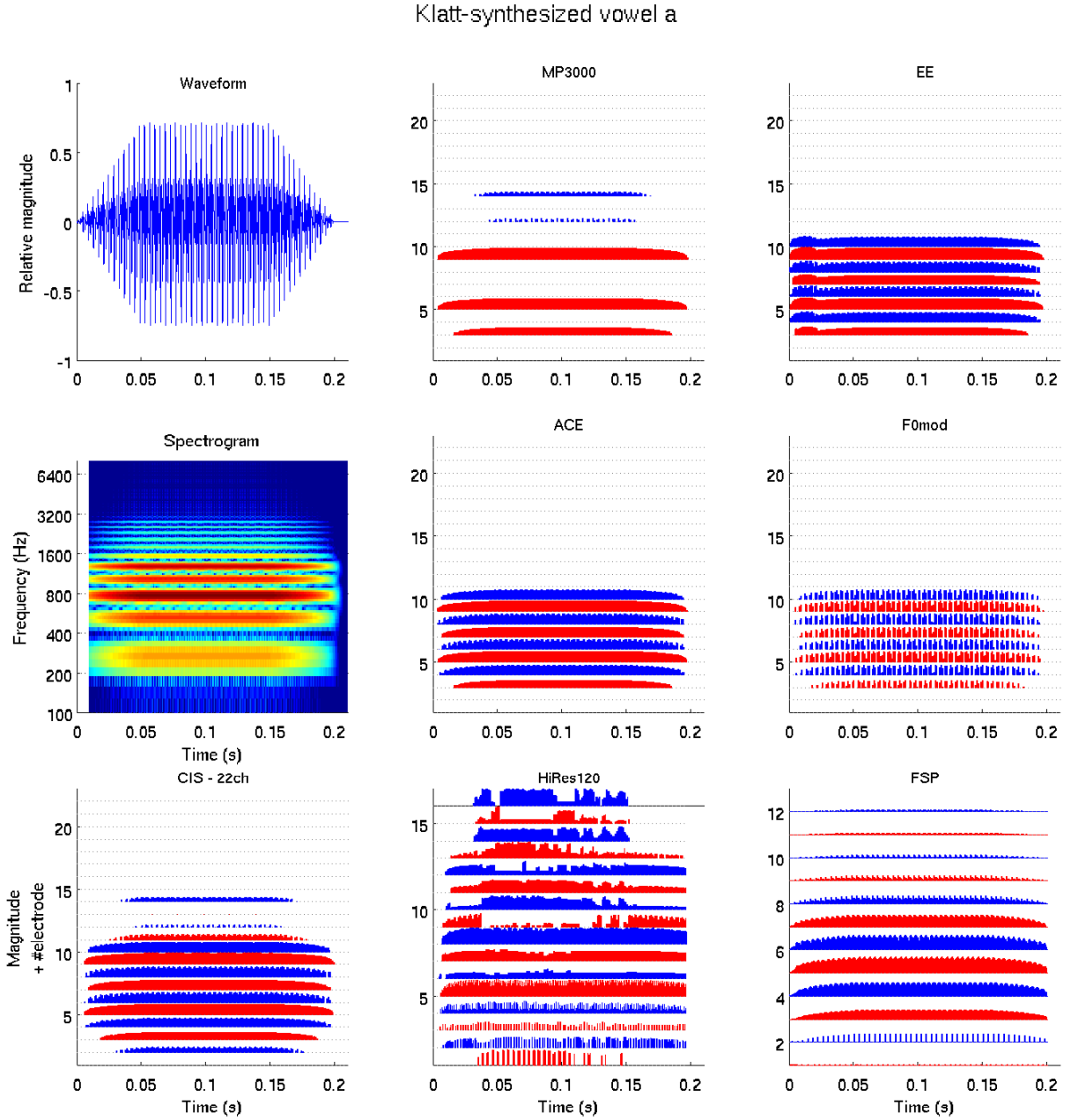
7

Fig. 3. Waveform, spectrogram and electrodograms for a Klatt-synthesized vowel with F0=100 Hz, and formant frequencies 700, 1220, and 2600Hz. The signal was presented at an average RMS level of 60 dB SPL. For the electrodograms, the vertical axis indicates the channel, and the height of each vertical line represents the magnitude of the pulse. The magnitude is expressed in different units for different strategies. The red and blue colors serve to visually distinguish adjacent channels and have no additional meaning. For the CIS, ACE, MP3000, EE, and F0mod strategies the channel magnitudes are shown between 0 and 1 before compression. For HiRes120, current was normalized by dividing by the maximum current, and normalized values below 0.1 were set to zero. HiRes120 uses simultaneous stimulation of adjacent electrodes to generate virtual channels, which is hard to distinguish on the current plot. For FSP the channel magnitudes between 0 and 1 are shown, which are linearly mapped to current, and multi-pulse sequences have been replaced by single pulse sequences for clarity.

one study of 46 experienced CI users where such differences were explicitly taken into account, no significant differences were found between FSP and a variant of CIS in speech perception tests, although the participants'
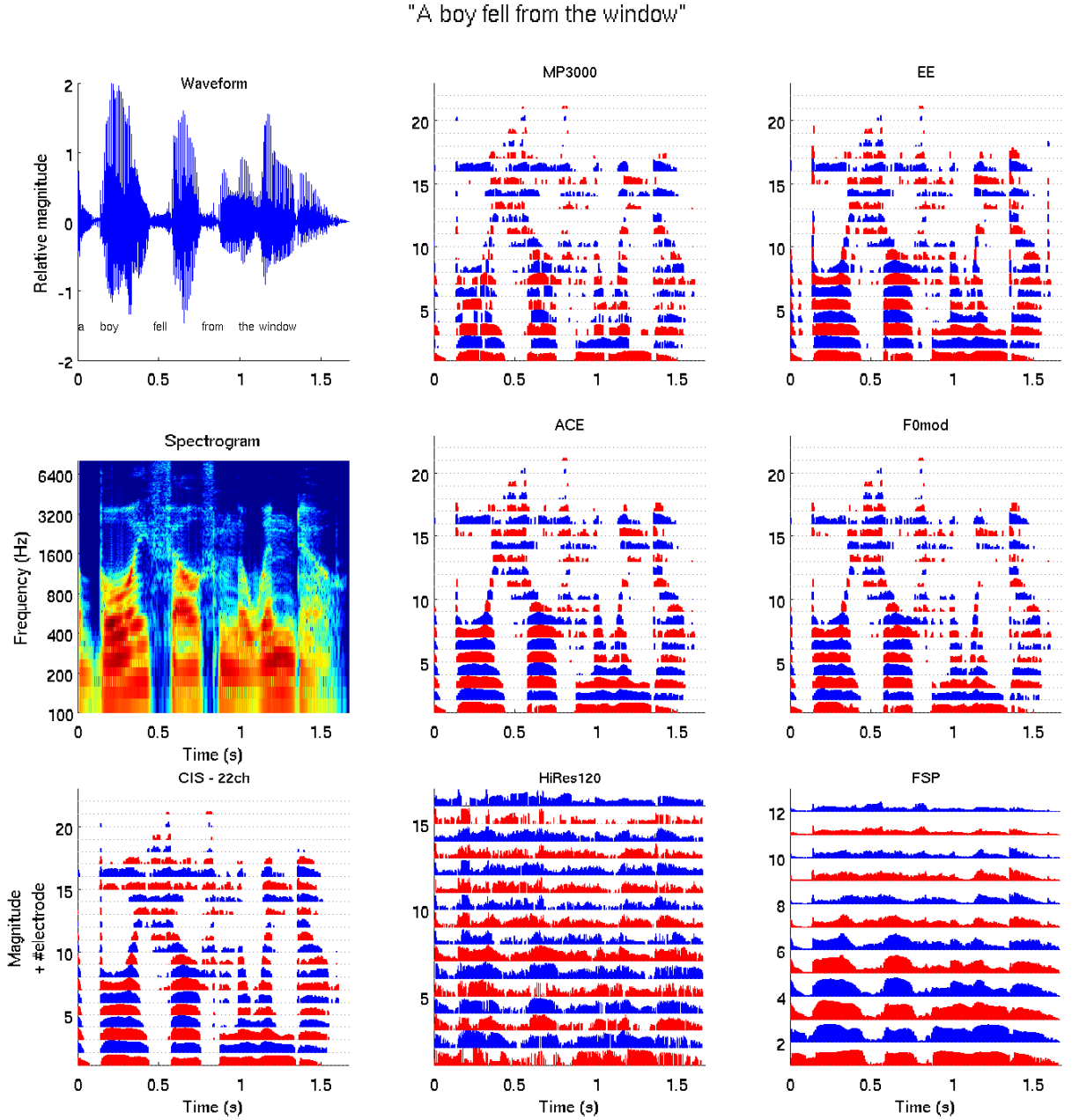
Fig. 4. Waveform, spectrogram and electrodograms of the sentence "A boy fell from the window" from the HINT corpus, uttered by a male speaker. All parameters are identical to those of figure 3.

subjective preferences generally favored FSP [24]. A similar overall result was reported from a different study with 20 CI users [25]. Moreover, it should be noted that in some experiments the fitting of the CI system to recipients was not altered when changing from CIS to FSP [25]. The study by Riss et al. [26] seems to indicate that at least some of the short-term improvements that have been seen with FSP can be attributed to the extended frequency range. In some studies also FSP was preferred for music. Studies with the newer FS4 and FS4-p strategies are ongoing. As studies with the newer FS4 and FS4-p strategies are ongoing, further research is needed to quantify perceptual outcomes more thoroughly.
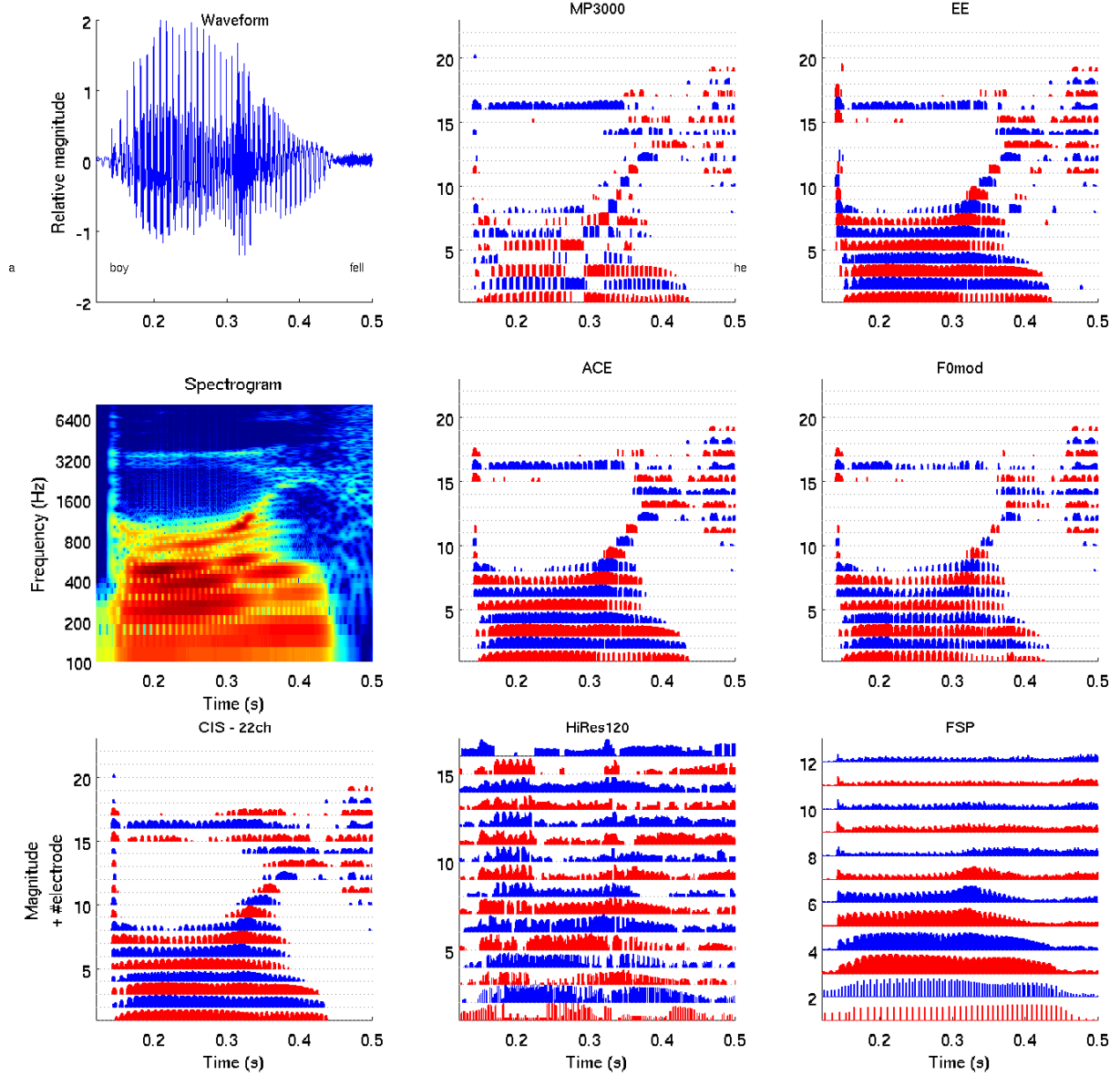
Fig. 5. Same as figure 4, but zoomed into the word "boy".

C. HiRes120

Another sound-processing scheme designed to enhance delivery of TFS information to CI recipients is used in systems manufactured by Advanced Bionics. Known as HiRes120, this scheme applies a technique to identify the dominant spectral peak within each of the band-pass filters that perform the spectral analysis of incoming sounds. The frequency of each spectral peak is used to control a synthetic modulator such that the modulations contain temporal information derived from each frequency band that is not present in the amplitude envelope of the band-limited signals [27]. These modulations are combined with the corresponding envelope levels and then sampled in synchrony with the pulses delivered to the electrodes. The typical pulse rate on each electrode is about 2 kpps. At the same time, the estimated

10
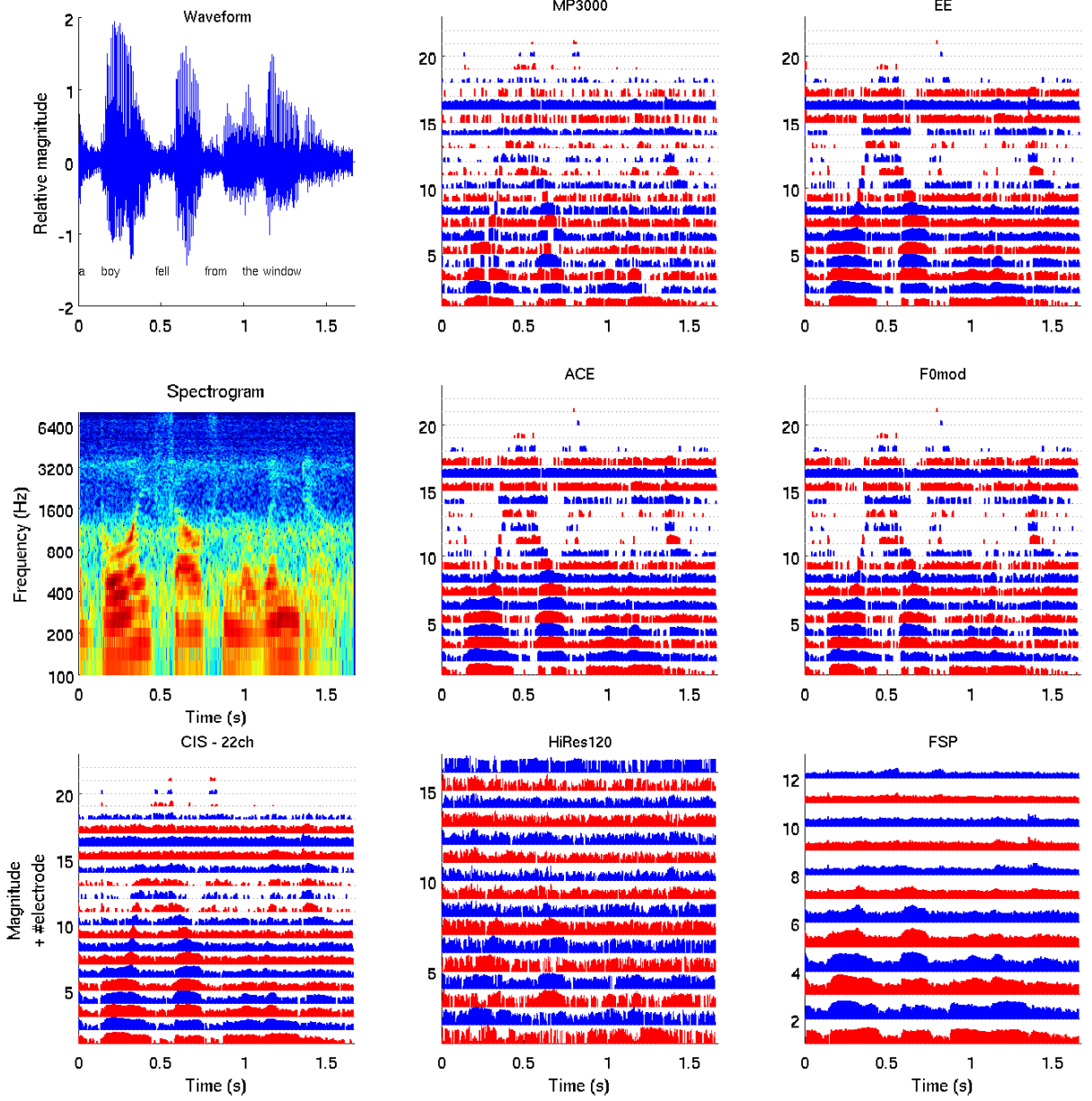
Fig. 6. Same as figure 4, but with noise added at an SNR of 10 dB.

peak frequency within each of the analysis filters is used to control the relative currents of pulses delivered simultaneously on two adjacent electrodes that are allocated to the filter. There are 16 intracochlear electrodes in the Advanced Bionics implant, and therefore 15 paired electrodes can be allocated to the filters. By varying the relative currents on the electrode pairs, so-called virtual channels are created, and it is assumed that the site of maximal neural activity can be steered with finer spatial resolution than is possible when the electrodes are activated one at a time [28]. With HiRes120, 8 different ratios of current are implemented, leading to 8

Fig. 7. Virtual channel plot for the HINT sentence "A boy fell from the window", processed by HiRes120. Color intensity indicates current. Integer numbers indicate "real" channels and the lines in between the 8 virtual channels, created by stimulating with two electrodes simultaneously

virtual channels per adjacent pair of physical electrodes. HiRes120 is claimed to provide improvements over sound-processing schemes such as CIS in both temporal and spatial resolution of the stimulation patterns. The main differences between 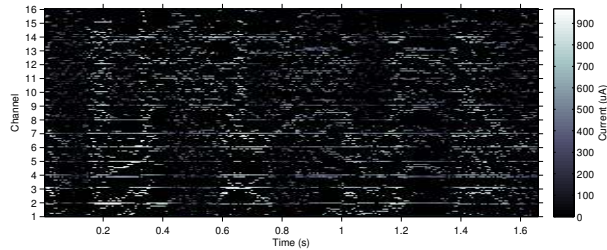these stimulation schemes are most clearly visible in the electrodograms of Figure 3-5. Additionally, a graphical representation of the virtual channels is shown in Fig. 7.

As with many publications in this field, studies reporting the performance of HiRes120 often have confounding factors that make it difficult to determine the specific effects of each technical change to the sound processing. In a study with 8 CI users, Firszt et al [29] compared perceptual performance between HiRes120 and HiRes, which is a CIS-like strategy without current-steering. Although significant improvements in perception were reported from some listening tests, it was unclear whether they could be attributed specifically to the addition of the current-steering feature. In [30] and [31] the current-steering stimulation strategy was compared with HiRes, both in 10 adult CI-recipients, on speech perception in quiet and in noise, music perception measures as well as other psychophysical measures: place-pitch sensitivity, spectral-ripple discrimination, Schroeder-phase discrimination and temporal modulation detection. There were no clear significant effects of the processing strategy on any of the speech and music perception abilities nor on temporal modulation detection. Furthermore, experience with the strategies did not seem to play a significant role. For some of the psychophysical measures differences were observed, but with varying results for HiRes120. Further research is needed to investigate the impact on more ecologically relevant outcome measures.

For all CI sound-processing strategies, the information throughput at the electrode-neural interface may be a fundamental limitation restricting improvements in perceptual performance. The limited perceptual effects of

introducing explicit information about the fine structure of acoustic signals in some CI sound-processing schemes such as HiRes120 and FSP may be a consequence of this "bottleneck" at the electrode-neural interface. In particular, if the spatial extent of the neural population activated by each electrode is broad and the populations associated with each electrode partially overlap, then temporal information from closely spaced electrodes will generally be combined at the neural level. Psychophysical studies have reported evidence that temporal patterns from nearby electrodes cannot be completely resolved by most CI recipients. This suggests that sound-processing schemes like HiRes120 and FSP which use very different approaches but rely on providing independent channels of information across adjacent electrodes may result in only limited benefits [32]. More carefully controlled studies of CI recipients' listening experiences using schemes such as HiRes120 and FSP over an extended time are needed to determine specifically whether provision of fine-structure information by these schemes is perceptually beneficial.

D. MP3000

The MP3000 strategy is based on the ACE scheme but uses a psychoacoustic masking model with the aim of improving sound perception for CI users based on more perceptually relevant channel selection. The masking model attempts to select the perceptually most important spectral components in the coding of any given input audio signal. The rationale for this development was that it should not be necessary to code sounds in parts of the spectrum that are masked. This approach reduces spread of excitation, and can lead to a more precise representation of the spectrum, which in turn could lead to improved speech intelligibility. Processing techniques based on auditory masking are widely used in common audio and music data-compression algorithms. These techniques also compress the audio signals by selecting only a subset of the frequency bands at a time. A well-known example is the MP3 compression algorithm. In principle, the n-of-m speech coding strategies such as ACE are similar to these data reduction or compression algorithms.

In MP3000 an additional processing stage is introduced between the envelope estimation and the channel selection modules (see figure 2, block 8). The psychoacoustic masking model used is derived from a body of data from psychoacoustic measurements in human auditory perception, such as studies on absolute thresholds of hearing and simultaneous masking [5], [33]. For each sound the envelopes of each channel of the filter bank are

inputs to the psychoacoustic model, and masking spread functions with 3 parameters (peak amplitude or attenuation, high- and low-frequency slope) are calculated. The masked threshold is calculated for each channel selected. The overall masked threshold from all channels is approximated by a non-linear superposition of the separate masked thresholds [34]. Subsequently, the n channels with highest levels relative to an estimate of the spread of masking are selected in each stimulation cycle. This selection of stimulation channels can be significantly different from the ACE standard scheme where only the n channels (typically n=8) with the highest envelope magnitudes are selected. This is clearly visible in figure 3, where in channel 14 a formant is coded with MP3000 that is not coded by ACE.

MP3000 has been implemented and evaluated in a within-subject repeated measures design with 221 subjects using an ABABA-design with A for ACE and B for MP3000. With a fixed pulse rate per channel, no significant difference was found for speech intelligibility and strategy preference between MP3000 (4 to 6 spectral maxima selected) and ACE (8 to 10 spectral maxima selected). The best results were found for MP3000 with 6 spectral maxima, leading to an increase in battery life of about 24% relative to ACE [35]. Thus when a lower number of stimulation channels is selected in each cycle, resulting in a lower overall stimulation rate, MP3000 has advantages. However, overall subject preferences were equally distributed between the two strategies, and additional parameters have to be fitted in the MP3000 mapping sessions.

IV. Experimental processing strategies

In this section some experimental stimulation strategies are briefly discussed to demonstrate the current limitations and opportunities with CI stimulation. Most of these strategies have been or are being considered for implementation in commercial speech processors for CIs. The following sections concern loudness-based strategies (SpeL for Specific Loudness, and SCORE for Stimulus Control to Optimise Recipient Experience), envelope enhancement based on a neural model (EE for Envelope Enhancement), enhancement of periodicity modulation (eTONE, F0mod), and bilateral stimulation strategies (PDT for Peak Derived Timing, MEnS for Modulation Enhancement Strategy). The loudness-based strategies are not shown in Fig. 2. They can be added onto any strategy by adding an extra block before the Mapping block (5). The bilateral strategies are not shown for reasons of clarity.

A. Loudness-based strategies (SpeL and SCORE)

A distinctive approach to sound processing for CIs has been explored in a range of experimental schemes, with the broad aim to improve the experience of loudness by CI recipients when listening to sounds with widely varying acoustic characteristics. Psychophysical studies have shown that CI users generally do not experience the loudness of sounds in the same way as listeners with NH, particularly when the spectral content and level of sound signals change over time [36].

In one such scheme, known as SpeL, the initial stages of sound processing are based on a running spectral analysis and the distribution of current levels across electrodes is determined such that the loudness experienced by the CI user is similar to that experienced by an average listener with natural hearing. More precisely, the levels are calculated using an estimate of the specific loudness for normal hearing related to the incoming sound [37]. The specific loudness function describes the contributions to the loudness of sounds from components in multiple frequency bands, and therefore depends on both the overall level and the detailed spectral shape of those sounds. Preliminary perceptual studies with CI recipients using SpeL confirmed that the relation between loudness and the level and bandwidth of sounds was closer to normal [38]. Furthermore, speech perception was similar on average to that obtained using ACE [39].

More recently, a simplified version of SpeL has been developed that uses the estimated specific loudness function to calculate the total loudness of sound signals in real time. This processing scheme, known as SCORE, uses the same methods as ACE to determine an initial set of stimulation parameters (i.e., stimulus levels across electrodes for constant-rate stimulation). The overall level of the set of stimuli is then adjusted so that the total loudness experienced by CI users is close to normal. Stimuli based on the adjusted parameters are delivered by the electrodes as for ACE. Tests of speech recognition with SCORE showed small but statistically significant improvements over ACE [40]. Further development enabled an extended version of SCORE to be used by CI recipients who benefit from simultaneous use of an acoustic hearing aid in the non-implanted ear. Experimental studies with this scheme (SCORE bimodal) have suggested that it may improve the ability of users to localize sounds, presumably because the loudness differences between ears that carry information about the direction of a sound source are conveyed more consistently [41].

B. Envelope enhancement (EE)

In a CI the electrical stimulation generates action potentials in auditory neurons directly, predominantly bypassing any remaining hair cells and synapse function. The synapse normally demonstrates neural short-term adaptation [42], i.e., an increased firing rate at the onsets of sounds. This short-term adaptation acts as an across-channel phonological timing cue [42] and, with conventional schemes such as CIS, is not present in the electrically stimulated auditory nerve as in the normal auditory system. Furthermore, recent studies have demonstrated that the transient parts of the speech envelope carry information that is important for speech intelligibility in NH listeners [43].

Based on this rationale and former investigations [44]–[46] the enhanced envelope strategy (EE) was developed and its feasibility studied for applications in auditory prostheses [47]. In this approach an additional processing stage is introduced after the envelope detection stage (see figure 2, block 11) wherein peaks, as a model for the short-term adaptation and dependent on the onset rise time, are added at the onsets in the envelope. This scheme is complementary to the main structure of ACE or CIS.

The EE algorithm was evaluated with CI users, using sentence materials in stationary speech-shaped noise and with an interfering talker [48]. All listeners demonstrated an immediate benefit with EE relative to ACE. With the onsets detected from the clean speech signal, speech intelligibility improvements were obtained resulting in a 2.1 dB improvement in speech reception threshold (SRT, i.e., the SNR at which 50% speech is intelligible) and also in stop consonant recognition. For a 2-speaker scenario comprising a talker and an interferer, the SRT improvement was 2.1 dB; when the onsets were enhanced for the target speaker alone. When processed for the noisy mixture of target and interfering speaker together, it was 1.0 dB. The latter example illustrates that benefits can be obtained without a priori knowledge of the clean speech signal [48].

The advantage of this enhanced envelope coding is due to emphasis on across-channel temporal coherence in the coded speech signal. This temporal marker is an important attribute for speech understanding in adverse listening situations, and for sound source segregation [49]; see also the electrodograms in figures 4, 5 and 6. The onset enhancement is particularly noticeable for the b-sound of the word "boy" in figure 5.

C. Periodicity modulation enhancement (eTone and F0mod)

From psychophysical studies it is known that periodicity cues are better perceived when modulation depth is high [50], [51] and modulations are synchronized across channels to some extent [32], [50]. This is probably due to spread of excitation: electrodes close together stimulate overlapping populations of neurons, which therefore receive the aggregate stimulation pattern of multiple electrodes. So if modulations are not synchronized across electrodes, the modulation depth at the neural level may be severely reduced. There is a trade-off in synchronizing modulations though: temporal modulations serve as grouping cues for the auditory system, to fuse parts of the spectrum into a single sound image, corresponding to a single sound source, and modulating too large a number of channels synchronously would remove this grouping cue, yielding potentially worse sound source segregation, which could severely affect speech intelligibility in noise.

From the electrodogram figures it is clear that with most commercial strategies temporal modulations are not well coded. In some channels modulation depth is quite shallow and the desynchronization across channels combined with spread of excitation leads to reduced modulation depth or even spurious modulations in the aggregate pattern that will be received by the auditory nerve fibers.

To improve this, based on the principle that synchronous and deep modulations can improve periodicity pitch, several strategies have been developed, e.g., the sawsharp strategy [52], [53], Peak Derived Timing (PDT), Modulation Depth Enhancement (MDE), F0 Synchronized ACE (F0Sync), Multi-channel Envelope Modulation (MEM) [54], F0 modulation (F0mod) [50], [55], and envelope enhanced tone (eTone) [56], [57]. While the signal processing to achieve it may differ, these strategies either expand modulation depth or remove existing modulations and explicitly modulate the envelope at the rate of F0. We will focus here on the F0mod and eTone strategies.

The F0 modulation (F0mod) strategy is a simple example of a periodicity enhancement strategy based on the ACE strategy. For each frame of samples it estimates F0 and voicing probability using an autocorrelation approach. If a frame is unvoiced, ACE processing is applied. If a frame is voiced, all channels are modulated synchronously using a sinusoidal modulator constructed based on the F0 estimate [50], [55]. The output of F0mod is shown in the electrodogram figures and its block diagram in figure 2. F0mod was evaluated in several studies.

Compared to ACE, F0mod was found to improve F0 discrimination of musical notes for different instruments, melody recognition of familiar Flemish songs (with all rhythm cues removed), estimates of musical pitch intervals, pitch ranking, and melodic contour identification [55], [58]. In a follow-up study, ACE was compared to F0mod for speech recognition of Mandarin Chinese, which is a tonal language, in which pitch determines the lexical meaning of certain phonemes [7]. An off-line implementation of F0mod was used, which made use of the clean speech signal to estimate F0. Significantly better lexical tone perception was found with the F0mod strategy than with ACE for the male voice. No significant difference in recognition of Mandarin Chinese sentences in noise was found between F0mod and ACE. In a next study, F0mod was implemented on a real-time system without access to the clean signal and the effect of F0mod on speech intelligibility in quiet and noise was investigated for Dutch words and sentences. No significant difference was found between F0mod and ACE [59].

The eTone strategy [56], [57] is based on the same principles as F0mod and other pitch strategies, but includes some new concepts. It includes an F0 estimator based on harmonic sieves, which is very precise and robust to noise, and the modulated envelope is mixed with the original envelope with a ratio depending on an estimate of harmonicity of that particular channel. Modulations are synchronous across channels and an exponential decay modulation shape is used.

Compared to ACE, it was found to improve pitch ranking for sung-vowel stimuli three semitones apart. No effect was found on recognition of English sentences in multi-talker babble, with a subject-dependent fixed SNR, ranging from 4 to 12 dB. While these results cannot be directly compared with those obtained with F0mod, due to the use of different evaluation procedures, they are qualitatively similar. eTone's F0 estimator is clearly superior to F0mod's, but it is not clear whether this has any perceptual consequences.

While periodicity enhancement strategies can clearly improve periodicity pitch perception, performance is still well below that of NH listeners. For good pitch perception, listeners need access to all three physical cues (Cf. section I), and spectral (place) and temporal cues need to be consistent. There are no current CI strategies that make this possible, and we hypothesize that with the current electrode design and stimulation paradigm it is not possible to provide sufficiently place-specific stimulation to achieve performance similar to NH. Note that for a good representation of temporal information, good place specificity is required as well: when a population of neurons is stimulated by information from several channels due to spread of excitation, the aggregate pattern will be coded.

A downside of the discussed strategies is their dependence on an F0 estimator, which needs to be fine-tuned and might fail for some signals, especially if multiple sound sources are present. However, while the advantage obtained with these strategies is modest, they do not negatively affect speech intelligibility, so if computational complexity allows it would seem worthwhile to include them in commercial processors, especially for users of tonal languages.

D. Bilateral strategies (PDT, MEnS)

In various studies with controlled stimulation in laboratory conditions, it has been found that bilateral CI users can be sensitive to ITDs. Mostly single-electrode stimuli have been used [60]–[62]. ITD thresholds vary widely across subjects, with the best thresholds around 50-100 μs and in the worst case no ITD sensitivity at all.

Best performance can be achieved with deeply modulated pulse trains, a modulation frequency $\approx$ 100 Hz, and sufficient dead time (off-time) between bursts of stimulation [61]–[64].

Sound-processing strategies use more complicated stimulation patterns than the simple single-electrode stimuli used in most ITD discrimination experiments. Ecological sounds, in particular speech, are broadband and require stimulation of more than one electrode. When multiple electrodes are stimulated, there appears to be no beneficial effect on ITD discrimination [64], as is present for normal-hearing listeners, and performance is best when stimulation patterns are synchronized across channels [65].

For bimodal stimulation (CI combined with contralateral hearing aid) similarly ITDs can be perceived with optimal stimuli and if there is sufficient residual hearing, both with single and multi-electrode stimuli [66]–[68].

Several studies have investigated ITD sensitivity with commercial processors, stimulating multiple electrodes [69]–[71]. Performance is generally lower with speech processors than with direct single-electrode stimulation. The influence of the speech processor is evident from the fact that performance is very stimulus-dependent. Unfortunately these studies do not allow identification of which aspects of the stimulation patterns cause the performance decrease compared to direct single-electrode stimulation.

Current commercial strategies do not precisely code ITDs. The delay and spectral characteristics of the

processing paths can be very different for the left and right auditory prosthesis, certainly for bimodal systems. This may lead to non-synchronous and non-coordinated (across channels) left and right auditory stimulation. Temporal cues in the envelope are in some cases preserved, depending on the interaction between the spectral shape of the sound and the magnitude and phase response of the CI filter bank, and the level of the signal in each channel. Even if the strategy does in principle code temporal fine structure, the clocks of the two processors would need to be synchronized for optimal precision. While onset cues are preserved, the time of the first or maximal pulse associated with an onset does not necessarily correspond to the first or peak acoustic stimulation. This is due to quantization effects and other non-linear processing such as maxima selection.

One of the first strategies developed to improve ITD coding with bilateral CIs is the Peak Derived Timing (PDT) strategy [61]. It operates by synchronizing stimulation pulses from the CI with amplitude peaks in the fine structure of the signals in the different channels of the filter bank. In this manner, fine pulse timing cues are transmitted, in contrast with CIS-type sound-processing techniques that provide only envelope information with fixed stimulation rates. PDT was implemented on a wearable processor and evaluated after 2-3 weeks take home experience. Compared to ACE no clear advantage was found for localization of sounds presented from eight loudspeakers, but there were some individual differences. There was some evidence of binaural unmasking of speech in noise, but it was small and performance was not compared with ACE. Pitch ranking with PDT was tested, but no significant difference was found with ACE [54].

Bilateral CI strategies like PDT can introduce temporal patterns that are not synchronized with the acoustic signal. This is appropriate for application with bilateral CIs because the same processing is done for both CIs, but it is a problem for bimodal stimulation where the binaural system compares the neural excitation of the electric and acoustic signals. Therefore [68] proposed the Modulation Enhancement Strategy (MEnS), which imposes a deeply modulated envelope on all frequency channels simultaneously, explicitly synchronized with peaks in the acoustic signal. This is similar to pitch enhancement strategies such as F0mod (see section IV-C). Improved ITD thresholds and improved lateralization were found with MEnS compared to ACE.

While some improvements in ITD perception have been obtained in laboratory tests with experimental strategies, the same caveats hold as with the pitch strategies described in section IV-C: performance is much poorer than with normal hearing. It should be pointed out though that thus far only acute experiments have been performed, while listeners potentially need long-term exposure to the novel stimulation paradigm to learn to use the binaural timing cues provided. Therefore if such strategies do not decrease speech perception and are not too computational expensive, it seems a good idea to conduct a long-term study.

V. General discussion

In this article a tutorial of CI stimulation strategies was presented, together with a review of concepts and rationales of different standard and experimental processing schemes. Some of the newer schemes have demonstrated significant improvements in understanding of speech and perception of other types of sound. Although each such strategy may lead to only a small benefit, it is plausible that appreciably larger benefits may be obtained when they are combined. Furthermore, some signal-processing approaches introduce speech enhancements in noisy conditions at the cost of significant signal distortions. These distortions may be detrimental for sound quality when appraised by listeners with normal or impaired acoustic hearing, but are hardly noticeable by most CI recipients [72]. This is an opportunity for further improvements in auditory perception for CI users.

However, the broad neural excitation profiles inherent to present-day electrode array technology and electrical stimulation parameters most probably limit the potential for improvement. The number of independent information transmission channels is still very small because of both technical and perceptual/neural sensitivity limitations. Not all CI users can discriminate all channels, but even if all actual and virtual stimulation channels and electrodes may be perceptually discriminated from each other, this does not imply that channels can be resolved, nor that different channels can effectively convey independent information.

It has become clear that some temporal aspects of the input sound, such as the speech envelope and partly periodicity can be transmitted faithfully by CIs. However, the TFS and F0 are not adequately represented in present-day CI processors and are therefore presented to the auditory neural system only imprecisely.

Auditory perception results can be spectacular for many CI recipients in quiet environments, particularly in early-implanted deaf children when neural plasticity can fully play its role and in adults with a largely intact neural periphery. However, hearing in realistic adverse listening situations, as well as music perception and

16

sound source localization are still major challenges for sound coding and electrical stimulation in CIs. Also, a wide variation in outcomes is observed across CI users. A significant proportion of recipients get limited benefit from their CI, at least in terms of speech understanding. Some investigations indicate that a better individual fitting of the stimulation parameters (the map) may result in substantial improvement, be it by better selection of active channels [73] or by development of closed-loop automatic fitting paradigms [74].

Another important factor is the neural survival at the electrode-neuron interface in the auditory periphery, which may be improved by application of drugs such as neurotrophins. Future research will include a greater focus on the combination of non-standard pulse waveforms [75], new stimulation modes to reduce across-channel interactions, and improved electrode designs. These approaches may result in the provision of more independent information channels in future CI systems.

Tom Francart, born 1981 received the M.S. and Ph.D. degrees in engineering from the University of Leuven, KU Leuven, Leuven, Belgium, in 2004 and 2008, respectively. Since 2013 he is a research professor at KU Leuven. His research interests include sound processing for auditory prostheses, binaural hearing and objective measures of hearing.

Hugh McDermott is the Deputy Director of the Bionics Institute of Australia and a Professorial Fellow of the University of Melbourne. He is a biomedical engineer and Fellow of the IEEE. He was elected Fellow of the Acoustical Society of America in 2002 for signal processing that improves speech recognition with CIs. For over 30 years he has contributed directly to the design, development, and evaluation of CIs, hearing aids, and neurostimulation devices. This research has frequently been translated to products manufactured by the worlds foremost companies in this field, particularly Cochlear Ltd, Australia, and Sonova AG, Switzerland. He is named as an inventor on 19 patent families, and has a further 10 patent applications currently being processed. He has authored more than 105 journal articles, 7 book chapters, and over 100 additional publications. In 2009 he was awarded the first Callier Prize in Communication Disorders, a biennial award from the University of Texas, Dallas, for leadership that has fostered scientific advances and significant developments in the diagnosis and treatment of communication disorders.

Jan Wouters, born 1960, obtained a Master and PhD in physics from the University of Leuven, KU Leuven, Belgium, in 1982 and 1989, respectively, with intermission for officer military service. From 1989 till 1992 he was a Postdoc Research Fellow with the National Fund for Scientific Research (FWO) at the Institute of Nuclear Physics (UCL Louvain-la-Neuve) and at NASA Goddard Space Flight Center (USA). Since 1993 he is a Professor at the Dept. of Neurosciences of the KU Leuven (Full Professor since 2005) where he teaches 5 physics and audiology courses. His research focuses on audiology, the auditory system and auditory prostheses. He is author of about 240 articles in international peer-reviewed journals, associate editor of 3 international journals, president of the European Federation of Audiology Societies EFAS, president of the Belgian Audiology Society B-Audio, secretary general of the International Collegium of Rehabilitative Audiology ICRA.

REFERENCES

[1] G. O'Donoghue, "Cochlear implants–science, serendipity, and success." *N. Engl. J. Med.*, vol. 369, no. 13, pp. 1190–3, Sep. 2013.

[2] P. Loizou, "Mimicking the human ear," *Signal Process. Mag. IEEE*, 1998.

[3] M. Dorman and B. Wilson, "The design and function of cochlear implants," *Am. Sci.*, 2004.

[4] F.-G. Zeng, S. Rebscher, W. Harrison, X. Sun, and H. Feng, "Cochlear implants: system design, integration, and evaluation." *IEEE Rev. Biomed. Eng.*, vol. 1, pp. 115–42, Jan. 2008.

[5] B. Moore, *An Introduction to the Psychology of Hearing*, 6th ed. Bingley, UK: Emerald, 2013.

[6] A. Botros, R. Banna, and S. Maruthurkkara, "The next generation of Nucleus() fitting: a multiplatform approach towards universal cochlear implant management." *Int. J. Audiol.*, vol. 52, no. 7, pp. 485–94, Jul. 2013.

[7] M. Milczynski, J. E. Chang, J. Wouters, and A. van Wieringen, "Perception of Mandarin Chinese with cochlear implants using enhanced temporal pitch cues." *Hear. Res.*, vol. 285, no. 1-2, pp. 1–12, Mar. 2012.

[8] C. J. Plack, R. R. Fay, A. J. Oxenham, and A. N. Popper, Eds., *Pitch*, ser. Springer Handbook of Auditory Research. New York: Springer-Verlag, 2005, vol. 24.

[9] R. P. Carlyon and J. M. Deeks, "Limitations on rate discrimination." *J. Acoust. Soc. Am.*, vol. 112, no. 3 Pt 1, pp. 1009–25, Sep. 2002.

[10] W. A. Yost, "Discriminations of interaural phase differences." *J Acoust Soc Am*, vol. 55, no. 6, pp. 1299–1303, 1974.

[11] G. B. Henning, "Detectability of interaural delay in high-frequency complex waveforms." *J Acoust Soc Am*, vol. 55, no. 1, pp. 84–90, 1974.

[12] L. R. Bernstein and C. Trahiotis, "Lateralization of low-frequency, complex waveforms: the use of envelope-based temporal disparities." *J Acoust Soc Am*, vol. 77, no. 5, pp. 1868–1880, 1985.

[13] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants." *Nature*, vol. 352, no. 6332, pp. 236–8, Jul. 1991.

[14] L. A. J. Reiss, C. W. Turner, S. A. Karsten, and B. J. Gantz, "Plasticity in human pitch perception induced by tonotopically mismatched electro-acoustic stimulation." *Neuroscience*, vol. 256, pp. 43–52, Jan. 2014.

[15] P. J. Boyd, "Potential benefits from deeply inserted cochlear implant electrodes." *Ear Hear.*, vol. 32, no. 4, pp. 411–27, 2011.

[16] C.-H. Chouard, C. Fugain, B. Meyer, and H. Lacombe, "Long-term results of the multichannel cochlear implant," *Ann. N. Y. Acad. Sci.*, vol. 405, no. 1 Cochlear Pros, pp. 387–411, Jun. 1983.

[17] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise." *J Acoust Soc Am*, vol. 95, no. 2, pp. 1085–1099, 1994.

[18] C. M. McKay, H. J. McDermott, A. E. Vandali, and G. M. Clark, "Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant," *Aust. J. Oto-Laryngology*, vol. 6, no. 5, pp. 354–359, 1991.

[19] P. Seligman and H. McDermott, "Architecture of the Spectra 22 speech processor," *Ann. Otol. Rhinol. Laryngol. Suppl.*, vol. 104, no. (Suppl. 166), pp. 139–141, 1995.

[20] M. W. Skinner, L. K. Holden, T. A. Holden, R. C. Dowell, P. M. Seligman, J. A. Brimacombe, and A. L. Beiter, "Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant," *Ear Hear.*, vol. 12, no. 1, pp. 3–22, 1991.

[21] J. F. Patrick, P. A. Busby, and P. J. Gibson, "The development of the Nucleus Freedom Cochlear implant system." *Trends Amplif.*, vol. 10, no. 4, pp. 175–200, Dec. 2006.

[22] M. W. H. Skinner Laura K.; Whitford, Lesley A.; Plant, Kerrie L.; Psarros, Colleen; Holden, Timothy A., M. W. Skinner, L. K. Holden, L. a. Whitford, K. L. Plant, C. Psarros, and T. a. Holden, "Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speechcoding strategies in newly implanted adults," *Ear Hear.*, vol. 23, no. 3, pp. 207–223, Jun. 2002.

[23] C. Zierhofer, "Electrical nerve stimulation based on channel specific sampling sequences," pp. US 6,594,525 B1, Jul. 2003.

[24] J. Müller, S. Brill, R. Hagen, A. Moeltner, S.-J. Brockmeier, T. Stark, S. Helbig, J. Maurer, T. Zahnert, C. Zierhofer, P. Nopp, and I. Anderson, "Clinical trial results with the MED-EL fine structure processing coding strategy in experienced cochlear implant users." *ORL. J. Otorhinolaryngol. Relat. Spec.*, vol. 74, no. 4, pp. 185–98, Jan. 2012.

[25] L. Magnusson, "Comparison of the fine structure processing (FSP) strategy and the CIS strategy used in the MED-EL cochlear implant system: speech intelligibility and music sound quality." *Int. J. Audiol.*, vol. 50, no. 4, pp. 279–87, Apr. 2011.

[26] D. Riss, J.-S. Hamzavi, A. Selberherr, A. Kaider, M. Blineder, V. Starlinger, W. Gstoettner, and C. Arnoldner, "Envelope versus fine structure speech coding strategy: a crossover study." *Otol. Neurotol.*, vol. 32, no. 7, pp. 1094–101, Sep. 2011.

[27] W. Nogueira, L. M. Litvak, B. Edler, J. Ostermann, and A. Büchner, "Signal Processing Strategies for Cochlear Implants

Using Current Steering," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–21, Nov. 2009.

[28] B. H. Bonham and L. M. Litvak, "Current focusing and steering: modeling, physiology, and psychophysics." *Hear. Res.*, vol. 242, no. 1-2, pp. 141–53, Aug. 2008.

[29] J. B. Firszt, L. K. Holden, R. M. Reeder, and M. W. Skinner, "Speech recognition in cochlear implant recipients: comparison of standard HiRes and HiRes 120 sound processing." *Otol. Neurotol.*, vol. 30, no. 2, pp. 146–52, Feb. 2009.

[30] G. S. Donaldson, P. K. Dawson, and L. Z. Borden, "Within-subjects comparison of the HiRes and Fidelity120 speech processing strategies: speech perception and its relation to place-pitch sensitivity." *Ear Hear.*, vol. 32, no. 2, pp. 238–50, 2011.

[31] W. R. Drennan, J. H. Won, K. Nie, E. Jameyson, and J. T. Rubinstein, "Sensitivity of psychophysical measures to signal processor modifications in cochlear implant users." *Hear. Res.*, vol. 262, no. 1-2, pp. 1–8, Apr. 2010.

[32] C. M. McKay and H. J. McDermott, "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites." *J. Acoust. Soc. Am.*, vol. 100, no. 2 Pt 1, pp. 1081–92, Aug. 1996.

[33] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*. Springer Science & Business Media, 2007.

[34] W. Nogueira, A. Büchner, T. Lenarz, and B. Edler, "A Psychoacoustic NofM-Type Speech Coding Strategy for Cochlear Implants," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 18, pp. 3044–3059, 2005.

[35] A. Buechner, A. Beynon, W. Szyfter, K. Niemczyk, U. Hoppe, M. Hey, J. Brokx, J. Eyles, P. Van de Heyning, G. Paludetti, A. Zarowski, N. Quaranta, T. Wesarg, J. Festen, H. Olze, I. Dhooge, J. Müller-Deile, A. Ramos, S. Roman, J.-P. Piron, D. Cuda, S. Burdo, W. Grolman, S. R. Vaillard, A. Huarte, B. Frachet, C. Morera, L. Garcia-Ibáñez, D. Abels, M. Walger, J. Müller-Mazotta, C. A. Leone, B. Meyer, N. Dillier, T. Steffens, A. Gentine, M. Mazzoli, G. Rypkema, M. Killian, and G. Smoorenburg, "Clinical evaluation of cochlear implant sound coding taking into account conjectural masking functions, MP3000." *Cochlear Implants Int.*, vol. 12, no. 4, pp. 194–204, Nov. 2011.

[36] C. M. McKay, K. R. Henshall, R. J. Farrell, and H. J. McDermott, "A practical method of predicting the loudness of complex electrical stimuli," *J. Acoust. Soc. Am.*, vol. 113, no. 4pt1, pp. 2054–2063, 2003.

[37] B. C. Moore and B. R. Glasberg, "A revised model of loudness perception applied to cochlear hearing loss," *Hear. Res.*, vol. 188, pp. 70–88, 2004.

[38] H. J. McDermott, C. M. McKay, L. M. Richardson, and K. R. Henshall, "Application of loudness models to sound processing for cochlear implants," *J. Acoust. Soc. Am.*, vol. 114, no. 4pt1, pp. 2190–2197, 2003.

[39] H. J. McDermott, C. M. Sucher, and C. M. McKay, "Speech perception with a cochlear implant sound processor incorporating loudness models." *Acoust. Res. Lett. Online*, vol. 5, no. 4, pp. 1–7, 2004.

[40] A. Varsavsky and H. J. McDermott, "Application of real-time loudness models can improve speech recognition for cochlear implant users." *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 81–7, Jan. 2013.

[41] T. Francart, A. Lenssen, and J. Wouters, "The effect of interaural differences in envelope shape on the perceived location of sounds (L)." *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 611–4, Aug. 2012.

[42] B. Delgutte, "Auditory neural processing of speech," in *Handb. Phonetic Sci.*, Hardcastle WJ and L. J, Eds. Oxford: Blackwell Publishers, 1997, pp. 507–538.

[43] M. S. Lewicki, "Information theory: A signal take on speech (and references therein)." *Nature*, vol. 466, no. 7308, pp. 821–2, Aug. 2010.

[44] L. Geurts and J. Wouters, "Enhancing the speech envelope of continuous interleaved sampling processors for cochlear implants." *J. Acoust. Soc. Am.*, vol. 105, no. 4, pp. 2476–84, Apr. 1999.

[45] A. E. Vandali, "Emphasis of short-duration acoustic speech cues for cochlear implant users." *J. Acoust. Soc. Am.*, vol. 109, no. 5 Pt 1, pp. 2049–61, May 2001.

[46] A. Bhattacharya, A. Vandali, and F.-G. Zeng, "Combined spectral and temporal enhancement to improve cochlear-implant speech perception." *J. Acoust. Soc. Am.*, vol. 130, no. 5, pp. 2951–60, Nov. 2011.

[47] R. Koning and J. Wouters, "The potential of onset enhancement for increased speech intelligibility in auditory prostheses." *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. 2569–81, Oct. 2012.

[48] ——, "Speech onset enhancement improves intelligibility in adverse listening conditions in cochlear implants," *Hear. Res.*, vol. submitted, 2014.

[49] S. a. Shamma, M. Elhilali, and C. Micheyl, "Temporal coherence and attention in auditory scene analysis." *Trends Neurosci.*, vol. 34, no. 3, pp. 114–23, Mar. 2011.

[50] L. Geurts and J. Wouters, "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants." *J. Acoust. Soc. Am.*, vol. 109, no. 2, pp. 713–26, Feb. 2001.

[51] C. M. McKay, H. McDermott, and G. M. Clark, "Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth." *J Acoust Soc Am*, vol. 97, no. 3, pp. 1777–1785, 1995.

[52] T. Green, A. Faulkner, and S. Rosen, "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants." *J Acoust Soc Am*, vol. 116, no. 4 Pt 1, pp. 2298–2310, 2004.

[53] T. Green, A. Faulkner, S. Rosen, and O. Macherey, "Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification." *J. Acoust. Soc. Am.*, vol. 118, no. 1, pp. 375–85, Jul. 2005.

[54] A. E. Vandali, C. M. Sucher, D. J. Tsang, C. M. McKay, J. W. Chew, and H. McDermott, "Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies." *J Acoust Soc Am*, vol. 117, no. 5, pp. 3126–3138, 2005.

[55] J. Laneau, J. Wouters, and M. Moonen, "Improved music perception with explicit pitch coding in cochlear implants." *Audiol Neurootol*, vol. 11, no. 1, pp. 38–52, 2006.

[56] A. E. Vandali and R. J. M. van Hoesel, "Development of a temporal fundamental frequency coding strategy for cochlear implants." *J. Acoust. Soc. Am.*, vol. 129, no. 6, pp. 4023–36, Jun. 2011.

[57] ——, "Enhancement of temporal cues to pitch in cochlear implants: Effects on pitch ranking." *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 392–402, Jul. 2012.

[58] M. Milczynski, J. Wouters, and A. van Wieringen, "Improved fundamental frequency coding in cochlear implant signal processing." *J Acoust Soc Am*, vol. 125, no. 4, pp. 2260–2271, 2009.

[59] T. Francart, A. Osses, and J. Wouters, "Speech perception with F0mod, a cochlear implant pitch coding strategy," *Int. J. Audiol.*, vol. In press, 2014.

[60] R. J. M. van Hoesel, G. L. Jones, and R. Y. Litovsky, "Interaural time-delay sensitivity in bilateral cochlear implant users: effects of pulse rate, modulation rate, and place of stimulation." *J. Assoc. Res. Otolaryngol.*, vol. 10, no. 4, pp. 557–67, Dec. 2009.

[61] R. J. van Hoesel and R. S. Tyler, "Speech perception, localization, and lateralization with bilateral cochlear implants." *J Acoust Soc Am*, vol. 113, no. 3, pp. 1617–1630, 2003.

[62] B. Laback, I. Zimmermann, P. Majdak, W.-D. Baumgartner, and S.-M. Pok, "Effects of envelope shape on interaural envelope delay sensitivity in acoustic and electric hearing." *J. Acoust. Soc. Am.*, vol. 130, no. 3, pp. 1515–29, Sep. 2011.

[63] V. A. Noel and D. K. Eddington, "Sensitivity of bilateral cochlear implant users to fine-structure and envelope interaural time dif-

ferences." *J. Acoust. Soc. Am.*, vol. 133, no. 4, pp. 2314–28, Apr. 2013.

[64] A. Ihlefeld, A. Kan, and R. Y. Litovsky, "Across-frequency combination of interaural time difference in bilateral cochlear implant listeners." *Front. Syst. Neurosci.*, vol. 8, p. 22, Jan. 2014.

[65] T. Francart, A. Lenssen, A. Buechner, T. Lenarz, and J. Wouters, "Effect of channel synchrony on interaural time difference perception with bilateral cochlear implants," *Ear Hear.*, vol. submitted, 2014.

[66] T. Francart, J. Brokx, and J. Wouters, "Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation," *J Assoc Res Otolaryngol*, vol. 10, no. 1, pp. 131–141, Mar. 2009.

[67] T. Francart, A. Lenssen, and J. Wouters, "Sensitivity of bimodal listeners to interaural time differences with modulated single- and multiple-channel stimuli," *Audiol Neurootol*, vol. 16, no. 2, pp. 82–92, 2011.

[68] ——, "Modulation enhancement in the electrical signal improves perception of interaural time differences with bimodal stimulation." *J. Assoc. Res. Otolaryngol.*, vol. 15, no. 4, pp. 633–47, Aug. 2014.

[69] D. W. Grantham, D. H. Ashmead, T. A. Ricketts, D. S. Haynes, and R. F. Labadie, "Interaural Time and Level Difference Thresholds for Acoustically Presented Signals in Post-Lingually Deafened Adults Fitted with Bilateral Cochlear Implants Using CIS Processing." *Ear Hear*, vol. 29, no. 1, pp. 33–44, 2008.

[70] P. Senn, M. Kompis, M. Vischer, and R. Haeusler, "Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors." *Audiol. Neurootol.*, vol. 10, no. 6, pp. 342–52, 2005.

[71] B. Laback, S.-M. Pok, W.-D. Baumgartner, W. A. Deutsch, and K. Schmid, "Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors." *Ear Hear.*, vol. 25, no. 5, pp. 488–500, Oct. 2004.

[72] O. U. R. Qazi, B. van Dijk, M. Moonen, and J. Wouters, "Understanding the effect of noise on electrical stimulation sequences in cochlear implants and its impact on speech intelligibility." *Hear. Res.*, vol. 299, pp. 79–87, May 2013.

[73] S. N. Garadat, T. A. Zwolan, and B. E. Pfingst, "Using temporal modulation sensitivity to select stimulation sites for processor MAPs in cochlear implant listeners." *Audiol. Neurootol.*, vol. 18, no. 4, pp. 247–60, Jan. 2013.

[74] M. Mc Laughlin, T. Lu, A. Dimitrijevic, and F.-G. Zeng, "Towards a closed-loop cochlear implant system: application of embedded monitoring of peripheral and central neural activity." *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 443–54, Jul. 2012.

[75] J. A. Undurraga, R. P. Carlyon, J. Wouters, and A. van Wieringen, "The polarity sensitivity of the electrically stimulated human auditory nerve measured at the level of the brainstem." *J. Assoc. Res. Otolaryngol.*, vol. 14, no. 3, pp. 359–77, Jun. 2013.

19