

## Music Perception with Cochlear Implants: A Review

Hugh J. McDermott, PhD

The acceptance of cochlear implantation as an effective and safe treatment for deafness has increased steadily over the past quarter century. The earliest devices were the first implanted prostheses found to be successful in compensating partially for lost sensory function by direct electrical stimulation of nerves. Initially, the main intention was to provide limited auditory sensations to people with profound or total sensorineural hearing impairment in both ears. Although the first cochlear implants aimed to provide patients with little more than awareness of environmental sounds and some cues to assist visual speech-reading, the technology has advanced rapidly. Currently, most people with modern cochlear implant systems can understand speech using the device alone, at least in favorable listening conditions. In recent years, an increasing research effort has been directed towards implant users' perception of nonspeech sounds, especially music. This paper reviews that research, discusses the published experimental results in terms of both psychophysical observations and device function, and concludes with some practical suggestions about how perception of music might be enhanced for implant recipients in the future. The most significant findings of past research are: (1) On average, implant users perceive rhythm about as well as listeners with normal hearing; (2) Even with technically sophisticated multiple-channel sound processors, recognition of melodies, especially without rhythmic or verbal cues, is poor, with performance at little better than chance levels for many implant users; (3) Perception of timbre, which is usually evaluated by experimental procedures that require subjects to identify musical instrument sounds, is generally unsatisfactory; (4) Implant users tend to rate the quality of musical sounds as less pleasant than listeners with normal hearing; (5) Auditory training programs that have been devised specifically to provide implant users with structured musical listening experience may improve the subjective acceptability of music that is heard through a prosthesis; (6) Pitch perception might be improved by designing innovative sound processors that use both temporal and spatial patterns of electric stimulation more effectively and precisely to overcome the inherent limitations of signal coding in existing implant systems; (7) For the growing population of implant recipients who have usable acoustic hearing, at least for low-frequency sounds, perception of music is likely to be much better with combined acoustic and electric stimulation than is typical for deaf people who rely solely on the hearing provided by their prostheses.

---

Department of Otolaryngology, The University of Melbourne, Melbourne, Australia.

This work was supported financially by the Garnett Passe and Rodney Williams Memorial Foundation.

Correspondence: Hugh J. McDermott, PhD, Department of Otolaryngology, The University of Melbourne, 384-388 Albert Street, East Melbourne 3002, Australia; E-mail: [hughm@unimelb.edu.au](mailto:hughm@unimelb.edu.au).

©2004 Westminster Publications, Inc., 708 Glen Cove Avenue, Glen Head, NY 11545, U.S.A.

## 1. Introduction

Over two decades ago, when cochlear implants began to emerge as a practical treatment for deafness, expectations of their performance were generally modest. Suitable candidates for implantation were restricted to adults with profound or total hearing loss in both ears, who obtained minimal or no benefit from the use of the best available acoustic hearing aids but who had previously had sufficient hearing to learn and understand spoken language. Early devices were considered to be essentially aids to speech-reading (lip-reading), rather than unique hearing systems that could enable most users to understand speech in the absence of visual cues.

With the continuing development of implant technology, and the growing knowledge in relevant fields such as psychophysics, signal processing, and functional neural excitation, expectations of outcomes have increased steadily. Currently, people with some usable acoustic hearing are receiving cochlear implants and obtaining substantial benefit from them and an increasing number of people have received an implant in both ears.

The current population of implant users presently numbers over 60,000 worldwide, and a large proportion of them can understand most speech and recognize many other types of sound, at least in favorable listening conditions. These advances have led some implant recipients, especially those for whom performing or listening to music was particularly important before their hearing deteriorated, to attempt to use their implants to regain the experience of musical enjoyment.

Unfortunately, existing cochlear implant systems often provide inadequate auditory information about complex musical sounds for their users to enjoy fully that type of listening experience. A number of researchers have been investigating this problem, and several technical improvements to implant systems are now under development that may deliver better performance for listening to music in the future.

One of the earliest published reports of an experimental cochlear implant is that of Djourno and Eyries (1957). In an operation on a man left totally deaf as a consequence of bilateral cholesteatomas, a single-electrode stimulator was placed on the auditory nerve. Stimulation with low-rate electric pulse trains elicited sensations the patient described as “the song of a cicada or

cricket, or the turning of a roulette wheel.” Stimulation at higher rates (above 100 Hz) caused a “sharp tonal sound.” Although the patient could not understand more than a very few words with the device, he did appreciate the ability to hear various environmental noises. Whether he tried listening to music is not reported, but if so, it seems unlikely that he could have derived much enjoyment from it.

A multiple-electrode device was implanted by Simmons (1966) in the right auditory nerve of a man who was totally deaf in that ear and profoundly deaf in the left ear. An extensive series of psychophysical experiments was carried out. Loudness was found to be related to stimulus intensity and pitch to stimulation pulse rate, with increases in rate over the range of 100 to 400 Hz producing consistent increases in the perceived pitch. Of importance was that sensations of different pitch (or timbre) were associated with the separate activation of each of the six electrodes. By activating one electrode with pulse trains of varying rate, melodic pitch changes seemed to be perceived by the patient. For example, he was able to identify a few well-known tunes, such as *Jingle Bells* and *Mary had a Little Lamb*, but not always reliably.

Over a decade later, Bilger (1977) reported the results of a large number of psychophysical tests conducted with 12 users of single-electrode cochlear implants. He found that most subjects could discriminate changes in frequency of electric stimulation at low frequencies (125 and 250 Hz), but not at higher frequencies (1000 and 2000 Hz). Subjective pitch was consistently associated with the frequency of stimulation only for the lower range of frequencies. However, perception of the duration and temporal pattern of stimulation was adequate for subjects to discriminate changes of rhythm in short sequences of stimuli. The loudness perceived was related to the stimulus intensity. Identification of melodies or perception of musical instrument sounds was not assessed.

Moore and Rosen (1979) briefly reported melody recognition by one totally deaf patient implanted with a single electrode placed on the surface of the cochlea. Although the test conditions were not tightly controlled, the subject appeared to be able to identify each of 10 well-known tunes by hearing alone with little difficulty. This finding suggested that musical pitch information could be conveyed by changes in the temporal pattern of

stimuli, without corresponding changes in other parameters such as the cochlear place at which maximal excitation occurs. In normal hearing, temporal and spatial characteristics of excitation in the cochlea are closely interrelated, and both depend on the frequency of acoustic stimuli.

The ability to recognize tunes was investigated in a preliminary experiment with a multiple-electrode implant described by Eddington *et al.* (1978). With one subject, stimuli were presented on a single electrode with the frequency controlled to correspond to the musical notes of five commonly known tunes. Rhythm cues were eliminated. The subject spontaneously identified only three of the tunes. This seemingly poor performance was ascribed to a presumably inconsistent relationship between the stimulus frequency and the pitch perceived.

Few, if any, of these relatively early reports describe musical perception with cochlear implant systems configured in the way that they may have been used in the everyday lives of their recipients. Instead, those experiments were generally conducted under artificial conditions in which controlled stimuli were delivered directly to the auditory nerves of the subjects via the implanted electrodes. Sound processors, which were initially developed specifically to enable implant users to understand speech, were not used in the experiments. More recent publications have addressed the question of how well implant recipients perceive music when listening with the sound processors they normally use. However, to gain an adequate understanding of these studies, it is helpful first to review the design and function of modern implant systems.

## 2. Cochlear Implant Technology

The basic functional principle underlying cochlear implants is that useful hearing sensations can be elicited in a sensorineurally deaf ear by stimulating auditory neurons directly with controlled electric currents. Many different designs of cochlear implants have been described, including both commercial and experimental systems, but all designs have general features in common. All implant systems pick up sound signals with a microphone that is usually packaged in an enclosure worn on the user's pinna, as with conventional behind-the-ear hearing aids.

An electric signal corresponding to the variation of pressure associated with air-borne sound waves is conveyed from the microphone to an electronic signal processor. The processor is designed to convert selected features of acoustic signals into a pattern of electric nerve stimuli that will evoke appropriate hearing sensations in the implant user. Considerable flexibility is available to designers of sound processing circuits and algorithms. This has led to the development and evaluation of many distinct processing schemes; for a detailed review, see Loizou (1998). The schemes most commonly used in current practice are described briefly later.

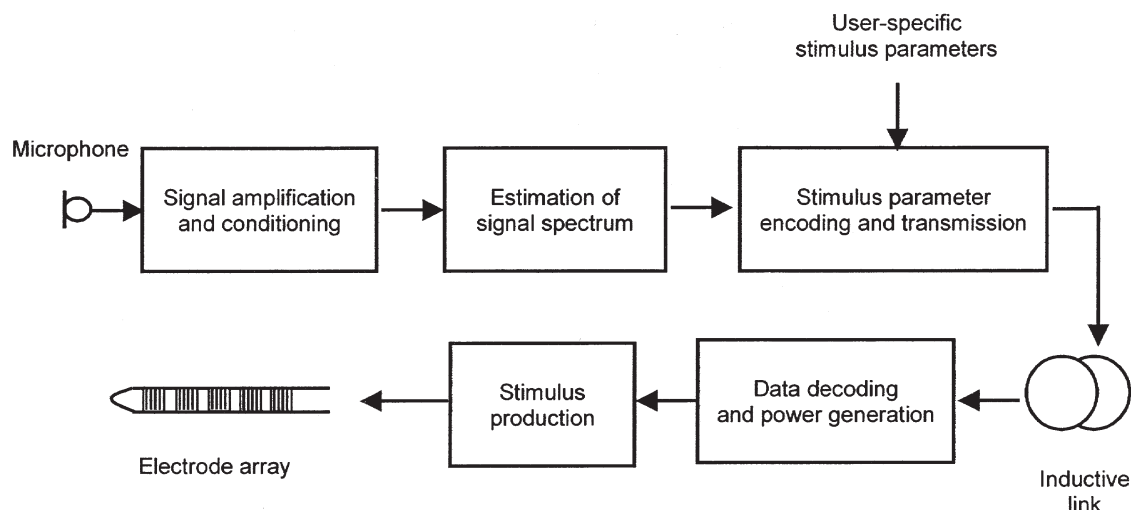
### 2.1. Implanted Devices

In most cases, the output of the sound processor consists of a digital code specifying the parameters of the electric stimuli to be delivered to the implanted electrode array. The code is usually conveyed to the implanted device via an inductive link (see Figure 1). The link, which is composed of two coils of wire separated by the skin overlying the implant, also serves to provide electric power to the implanted electronics.

An integrated circuit in the implant demodulates the signal obtained from the subcutaneous coil and decodes the information transmitted by the sound processor. This specifies the amplitude and temporal parameters of the stimulus to be generated and the electrodes that are to conduct the stimulus current. The output of most existing stimulators is a precisely controlled current that is delivered to the active electrodes as a series of symmetric, biphasic pulses (see Figure 2).

Some implants are capable of generating continuously varying currents as an alternative to discrete pulses. Such "analog" stimuli can be used to represent some details of the waveform of the sound signal with different processing techniques from those needed to represent a similar input signal with trains of rectangular pulses.

In some early implant designs, the primary objective was to deliver stimuli to the entire surviving population of auditory neurons by using a single electrode placed near the neural elements. The electric circuit was completed through a second electrode that was often located at a remote site. A single active electrode was attractive, mainly because of the relative simplicity of both the surgery and the stimulator electronics; however, it has since been established that being able



**Figure 1.** Schematic diagram showing the main functional blocks of a cochlear implant hearing prosthesis. In a typical multiple-channel system, sound signals are picked up by a microphone and passed to an amplifier, where they may undergo preprocessing such as filtering or compression. Next, the short-term spectrum of the signals is estimated. Suitable parameters of the electric stimulation to represent the spectrum are then calculated. These depend on a unique set of values that is determined for the individual implant user during device fitting and programming. The output of the sound processor comprises a digital code that is transmitted across the skin to the implant via a pair of coupled coils. An implanted receiver decodes the data transmitted by the external processor to obtain the parameters of the required pattern of stimulation. These parameters control a stimulator circuit that delivers electric currents to the array of intracochlear electrodes.

to stimulate different sectors of the neural population with some degree of independence has advantages.

In the normal cochlea an orderly relationship exists between the frequency of an audible sound and the location of maximal excitation of auditory neurons (Greenwood, 1990). Relatively high frequencies produce the most activity in neurons that innervate hair cells near the base of the cochlea, whereas lower frequencies activate neurons that innervate hair cells located at more apical positions. This tonotopic organization applies not only to hair cells but also to the cell bodies and dendrites of auditory neurons. Therefore, even in cases of profound sensorineural deafness in which few or no hair cells survive, cochlear implants may still take advantage of the tonotopic organization of the residual auditory neurons by means of an array of electrodes.

Such arrays generally comprise a number of discrete electrodes mounted on a carrier that is

surgically inserted into the cochlea through or near the round window. When an array is located deeply inside the cochlea, electrodes near the tip preferentially stimulate neurons that, with normal acoustic hearing, would have responded best to low-frequency sounds, whereas electrodes nearer the cochlear base stimulate neurons associated with higher-frequency sounds.

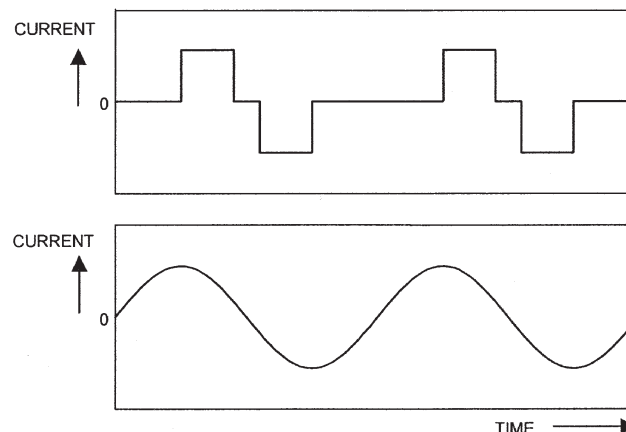
Multiple electrodes can be configured to deliver stimulating currents to the auditory neurons in different ways. The three main configurations available with existing devices are known as monopolar, bipolar, and common ground. As illustrated in Figure 3 (left), the monopolar configuration comprises an active electrode that is located in or close to the cochlea, and one or more separate electrodes that are located further away. These “indifferent” electrodes usually have a larger surface area than the active electrode and may serve as the current return path for many discrete active electrodes. Generally, single-channel im-

plants stimulate by using the monopolar configuration. In multiple-electrode implants employing monopolar stimulation, it is important that the active electrodes be located close to the neural population so that, ideally, stimulation on each electrode excites a spatially distinct set of neurons and consequently elicits a perceptually discriminable auditory sensation.

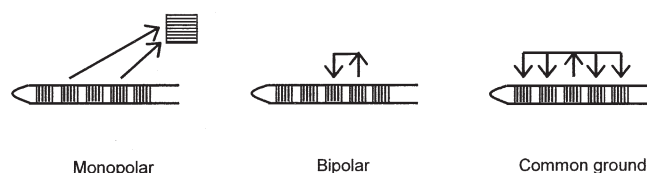
In principle, the spatial separation of the stimulating current paths in multiple-electrode implants can be improved by using bipolar stimulation. In this configuration, currents are passed between two electrodes, both of which are located relatively close to the auditory neurons (Figure 3, center). Several variations on the bipolar configuration may provide practical benefits in some conditions:

- In one variation, the separation between the two active electrodes (the “spatial extent”) can be increased, for example by activating pairs of electrodes that are separated by one or more inactive electrodes on the array. This usually results in a reduction of the current required to produce an audible sensation (i.e., the threshold current).
- Another variation involves arranging the electrodes spatially to direct the current flow in the cochlea more closely around a radial, rather than longitudinal, path. This is intended to increase the electrodes’ spatial selectivity and reduce thresholds by comparison with alternative configurations.
- In the third type of electrode configuration, the common-ground mode, one active electrode is selected, and many or all of the remaining intracochlear electrodes are used together as the return path for the stimulating current (Figure 3, right). In several respects, the common-ground arrangement is intermediate between the bipolar and monopolar configurations. Some of the perceptual effects of stimulating with different electrode configurations are discussed briefly later.

Typically, multiple-electrode implants deliver stimulating currents to the active electrodes in a sequence of temporally nonoverlapping pulses, whereas earlier single-channel devices used a continuously varying waveform. However, in some designs it is possible for analog waveforms



**Figure 2.** Examples of two general forms of electric stimulus that may be generated by cochlear implants. The ordinate shows current, whereas the abscissa shows time. At the top is a sequence of two biphasic pulses. Each pulse comprises two short intervals during which a constant current is delivered to the active electrodes. The current has equal magnitude in the two phases, but opposite directions. The phases may be separated by a brief time during which no current flows. The lower panel shows a so-called “analog” stimulus, in which the electrode current varies continuously in time.



**Figure 3.** Three types of electrode configuration used in multiple-channel implants. The illustration at the left shows the monopolar mode, in which current from the active electrodes on the intracochlear array flows to a single “ground” electrode, which is located remotely. The center illustration shows the bipolar mode, in which current passes between two active electrodes located nearby on the array. The illustration at the right shows the “common ground” mode, in which current from one intracochlear electrode flows to most or all of the remaining electrodes on the array.

(or rectangular pulses) to be delivered simultaneously to several electrodes.

Simultaneous stimulation via multiple electrodes may, in theory, have beneficial perceptual effects, particularly because it should enable the normal patterns of the auditory neurons' responses to acoustic signals to be emulated more closely. Unfortunately, in past experiments with cochlear implants, simultaneous stimulation has frequently been found to produce complicated side effects. For example, the complex summation of currents within the cochlea from multiple active electrodes can result in reduced spatial selectivity of the neural excitation and poorer control of perceived loudness.

## 2.2. Sound Processors

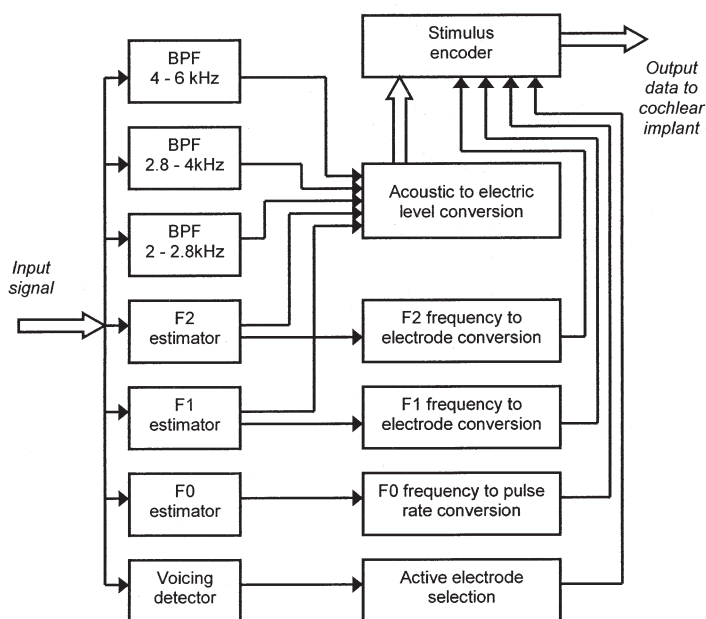
Sound-processing techniques for cochlear implants can be classified into three broad categories: feature-extracting strategies, spectrum-estimating pulsatile schemes, and analog stimulation schemes.

### 2.2.1. Feature-Extracting Strategies

The feature-extraction approach to the design of sound processors is now obsolete, but remains of interest because it was based, in part, on principles derived from psychophysical experiments into the perception of pitch, including musical pitch (discussed later). A series of processing schemes developed mainly during the 1980s at the University of Melbourne and Cochlear Limited (formerly Nucleus Ltd) culminated in the "Multipeak" (or "MPEAK") strategy (Patrick *et al.*, 1990; Patrick and Clark, 1991) which was implemented in the "Mini Speech Processor" (MSP).

A block diagram of the MPEAK scheme appears in Figure 4. The input signal from the microphone was analyzed with the assumption that it usually contained speech. Three acoustic features of the signal were extracted: the fundamental frequency (F0), and the frequencies of the first two formants (F1 and F2), which convey much of the information available in the signal about the identity of vowels and other voiced speech sounds. The frequencies of F1 and F2 were converted to positions of two active electrodes selected from the 22-electrode array according to the tonotopic principle.

In addition, three bandpass filters and envelope detectors estimated the amplitude of the incoming signal within three higher frequency re-



**Figure 4.** Functional block diagram of the MPEAK speech-processing strategy. Input signals from a microphone are analyzed to extract or estimate the parameters of a small number of acoustic features that are important for conveying information about speech. These parameters include voicing, the fundamental frequency (F0), and the frequencies and amplitudes of the first two formants (F1 and F2). In addition, the levels of signals in three higher-frequency bandpass filters (BPFs) are determined. As explained in the text, a subset of these parameters is selected for stimulation in each period. The signal levels are converted into appropriate current levels of electric stimulation. The digital data transmitted to the implant specify the currents to be delivered by the active electrodes. The resulting stimulation pattern comprises groups of four sequential pulses delivered at an overall rate dependent on the estimated F0 whenever the acoustic input signal is judged to contain voiced speech.

gions. These filters were assigned to specific electrodes at the basal end of the electrode array. They were included mainly to improve the processing of certain consonant sounds, such as unvoiced fricatives.

With MPEAK, sequential pulsatile stimulation was generated at a rate that depended on whether voicing was detected in the input signal. If a voiced signal was present, the estimate of F0 was used to control the stimulation frame rate. Within each frame, four pulses were generated representing F1, F2, and the lower two of the

three high frequency bands. If no voicing was detected, a stimulation rate of about 250 Hz was used, and the four pulses presented in each period represented F2 and each of the three high frequency bands.

The amplitudes of the selected acoustic features were converted into appropriate current levels determined for each electrode in each implant user. The minimum current level (i.e., the “T-level”) corresponded approximately to the threshold of audibility, and the maximum level (i.e., the “C-level”) evoked a sensation of comfortable loudness.

The rationale for generating pulsatile stimulation at an overall rate determined by the fundamental frequency of voiced speech was justified by psychophysical findings that showed implant users could perceive a pitch related to the pulse rate for rates ranging from about 100 to 300 Hz (discussed later). This range corresponds approximately to the F0 frequency range for many (but not all) speakers, and for other complex sounds such as those produced by certain musical instruments. Similar reasoning supported the development of an early single-channel extra-cochlear prosthesis (Fourcin *et al.*, 1979), which was intended primarily as a speech-reading aid. That device also applied stimulation to the auditory nerve at a frequency derived from an estimate of F0.

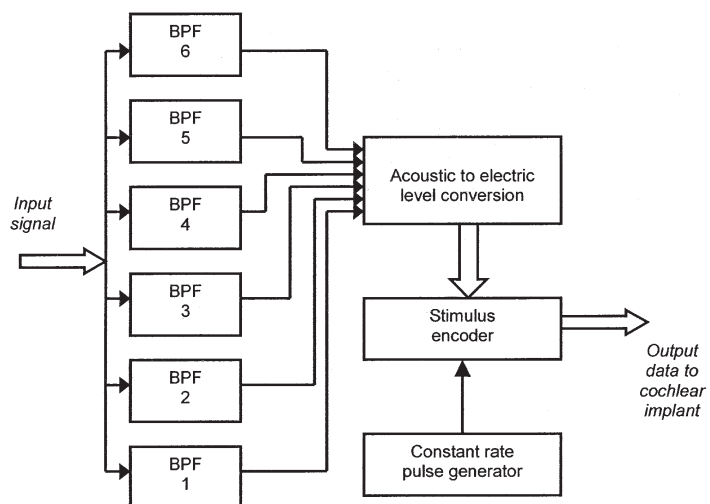
The feature-estimating schemes, such as MPEAK, could provide many implant users with enough information to enable the recognition of most speech sounds, but they had several inherent weaknesses. In particular, it was technically difficult to obtain accurate estimates of the relevant parameters of speech signals in a real-time processor that needed to function reliably in unfavorable conditions, such as situations with high levels of background noise. Estimating F0 in noisy or reverberant situations or in conditions where several different sources of F0 are present simultaneously is especially difficult. Another shortcoming is that strategies that extract or emphasize acoustic features specific to speech signals may not provide optimal processing of nonspeech sounds, including music and environmental noises.

### 2.2.2. Spectrum-Estimating Pulsatile Schemes

Considerations such as these eventually led to the abandonment of the feature-extraction approach to cochlear implant sound processing. For most current users of multiple-electrode devices, spectrum-estimating pulsatile schemes are the preferred choice. The three most widely used

schemes are known as “Continuous Interleaved Sampling” (CIS), “SPEAK,” and “Advanced Combination Encoder” (ACE). Each of these sound-processing schemes is designed to present information about prominent spectral features of sound signals, but it is not assumed that those spectral features are necessarily associated with speech. More important, a stimulation pulse rate is applied that is independent of any parameters of the input signal. Pulses are generally delivered to the active electrodes in a sequential cycle at a constant, relatively high rate.

A block diagram of a typical six channel CIS processor is shown in Figure 5 (Wilson *et al.*, 1991). The short-term spectrum of incoming signals is estimated by means of a bank of bandpass filters. At the output of each filter, the envelope of the waveform is estimated. These envelope signals are sampled at regular times, and their amplitudes are converted to appropriate stimulation current levels.



**Figure 5.** Functional block diagram of the Continuous Interleaved Sampling (CIS) sound-processing strategy. In this example, six bandpass filters (BPFs) are used to enable the short-term (or instantaneous) levels in each of six partially overlapping frequency bands to be estimated. The filters have center frequencies that are typically spaced regularly along a logarithmic scale. The levels at the outputs of the filters are converted into appropriate current levels of electric stimulation. The digital data transmitted to the implant specify the currents to be delivered by each of the electrodes. The resulting stimulation pattern comprises a series of interleaved pulses delivered at a constant rate.

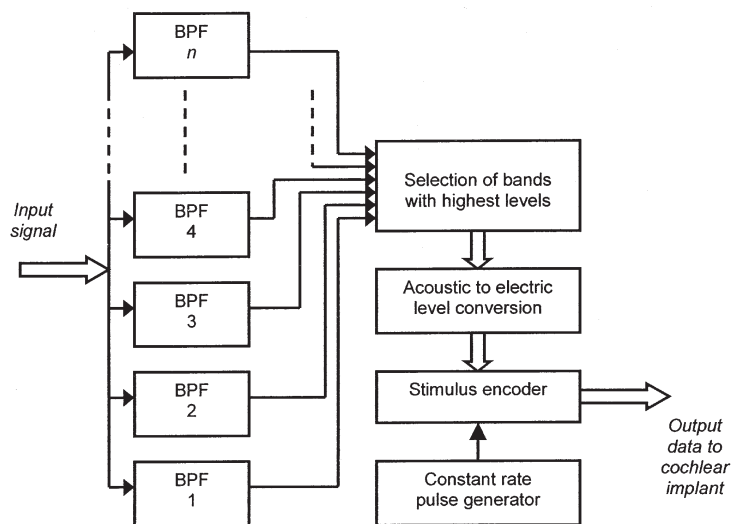
In the implant, brief electric pulses are delivered by electrodes corresponding to the filters at a rate equal to the sampling rate. In most existing implementations of CIS, the pulse rate is on the order of several thousand pulses per second per channel. Both commercial device manufacturers and independent researchers have implemented and evaluated numerous variations of the CIS scheme; for example, different numbers of filters and electrodes have been used, and alternative techniques have been investigated for converting the filters' outputs into levels of electric stimulation. However, the essential functional principles of the CIS scheme have been retained. Some alternative sound-processing schemes for multiple-electrode implants, such as SPEAK and ACE, generally have a larger number of bandpass filters than CIS schemes.

A block diagram of a typical ACE processor for a Nucleus (Cochlear, Lane Cove, NSW Australia) implant is shown in Figure 6 (Vandali *et al.*, 2000). The estimation of the short-term spectrum of incoming signals is performed by a bank of 20 filters having partially overlapping bandpass characteristics. The envelope of the signal at the output of each filter is estimated. In each stimulation period, the amplitudes of the envelopes are compared, and the subset of filter channels with the highest short-term amplitudes is identified. The number of channels selected in this process is limited, for example to 6 or 10. The amplitudes at these channel outputs are converted to appropriate levels of stimulation as in other non-simultaneous pulsatile schemes.

The overall stimulation pulse rate is approximately constant and is usually much higher for ACE than for SPEAK. The SPEAK scheme employs a pulse rate of about 250 Hz per channel, mainly because it was derived from the earlier "Spectral Maxima Sound Processor" (SMSP), in which only relatively low pulse rates were practical for technical reasons (McDermott *et al.*, 1992). The overall pulse rate for ACE may be at least 14.4 kHz, depending on the capabilities of the implanted stimulator, but otherwise ACE and SPEAK are functionally similar.

### 2.2.3. Analog Stimulation Schemes

As mentioned previously, analog stimulation schemes are presently used less often than pulsatile schemes in multiple-electrode prostheses. In one such scheme, "Simultaneous Analog Stimulation" (SAS) (Kessler, 1999), the short-



**Figure 6.** Functional block diagram of the Advanced Combination Encoder (ACE) sound-processing strategy. A relatively large number of bandpass filters (typically 20) are used to estimate the short-term spectrum of the input signal. The filters have partially overlapping frequency responses covering a wide bandwidth (e.g., 200 Hz–10 kHz). The levels at the outputs of the filters are compared so that only the subset comprising the highest levels is passed to the following stages of processing. The subset typically includes the 10 highest levels, and only the 10 corresponding electrodes in the cochlear implant are activated. The selected signal levels are converted into appropriate current levels of electric stimulation. The digital data transmitted to the implant specify the currents to be delivered by the active electrodes. The resulting stimulation pattern comprises a series of interleaved pulses delivered at a constant overall rate.

term spectrum of incoming signals is estimated by means of a bank of bandpass filters. However, in contrast to pulsatile strategies, SAS uses the waveform at the output of each filter, rather than the signal envelopes, as the basis for stimulation. Each filtered waveform is compressed in a manner analogous to the conversion of envelope amplitudes to current levels in pulsatile stimulation strategies. This process ensures that the levels of stimulation result in comfortable loudness and adequate audibility of most sounds for each implant user.

In the implant, the compressed waveforms are delivered simultaneously as continuously



varying currents to the active electrodes. The SAS scheme is closely related to the earlier CA strategy that was used successfully with the now-obsolete Ineraid multiple-electrode cochlear implant (Eddington, 1980). The performance of numerous other sound-processing schemes, including the ones outlined above, in enabling implant users to understand speech has been investigated and reported extensively, and will not be reviewed here.

### 3. Music Perception with Cochlear Implants

Music is difficult to define. A purely phenomenologic “I know it when I hear it” definition is unsatisfactory, because few people would be able to agree about all types of music. Strictly objective definitions are also problematic, because it is hard to imagine *any* type of sound that could not form part of a piece of music, given an appropriate context—environmental noises and synthetic sounds are common elements in certain musical genres. Even a definition that would clearly separate singing (music) from speech (not music) is elusive; utterances in tonal languages, in particular, may sound musical to some listeners, especially those who are unable to comprehend the meaning.

Nevertheless, much of the published research on how cochlear implant users perceive music apparently rests on the assumption that music can be characterized as an organized sequence of sounds that have a small number of fundamental features, including rhythm, melody, and timbre. Additional attributes of sounds, such as harmony and the overall loudness, also contribute to the structure of music. Each of these properties can be described, at least approximately, in terms of physical parameters of acoustic signals. For example, the loudness of a sound is related to its intensity, and rhythm is conveyed in most musical styles by moderately rapid variations in loudness.

Beyond these objective characteristics of sounds, however, are diverse phenomena that are also important in the experience of listening to music. These include subjective quality, mood, and the situational context. For instance, a person’s emotional response to music heard at a crowded dance party organized to celebrate a significant event, such as a birthday, is undoubtedly

a very different experience from that of listening alone to a high-quality recording of a cerebral work by J. S. Bach. However, it is reasonable to assume that these divergent aspects of the musical experience are common to listeners regardless of the functional mode of their hearing. Because they are not specific to users of cochlear implants, they will not be discussed further here.

#### 3.1. Perception of Rhythm

Temporal patterns in musical sounds that impart a distinctive rhythm generally occur at the approximate frequency range of 0.2 to 20 Hz. Acoustic features that change less rapidly than this are associated with overall variations in loudness (often called “dynamics” in music). Higher frequency components of acoustic signals carry pitch information, which will be discussed later.

The ability of listeners with cochlear implants to discriminate rhythms has been investigated by several researchers. A standardized test used by Gfeller and Lansing (1991; 1992) is known as the “Primary Measures of Music Audiation” (PMMA), and was developed mainly for use with children (Gordon, 1979). The rhythm subtest of the PMMA comprises pairs of short sequences of sounds, each recorded with unvarying pitch and timbre. The sequences in each pair are either identical or different in rhythmic pattern and are randomly presented to the listener, who is instructed to indicate whether the pair of sequences is the same or different. The score that would be obtained with uniformly random responses is 50%.

Performance on this test was assessed with 18 adult users of cochlear implants (Gfeller and Lansing, 1991). Ten subjects were users of a now-superseded feature-extraction speech processing strategy developed for the Nucleus 22-electrode device manufactured by Cochlear Limited. This strategy was a predecessor of MPEAK that extracted and presented information about the speech features F0, F1, and F2, but not the three higher frequency bands described earlier and outlined in Figure 4 (Dowell *et al.*, 1987). The remaining eight subjects were users of the Ineraid prosthesis and a four-channel Compressed Analog sound processor; this system is also now obsolete. The mean score for the group was 88% and individual scores ranged from about 80% to 95%. There appeared to be little difference in performance related to the type of prosthesis each subject used.

A subsequent study also used the PMMA (in a slightly modified form) to examine any differences in rhythm perception between two sound-processing schemes used by 17 recipients of the Nucleus 22-electrode device (Gfeller *et al.*, 1997). The two schemes were MPEAK and the earlier F0/F1/F2 strategy. The mean score for both schemes on the PMMA rhythm subtest was approximately 84% correct. This is not only close to the corresponding finding from the previously mentioned study (Gfeller and Lansing, 1991) but also very close to the average score obtained by a control group of 35 subjects with normal hearing (Gfeller *et al.*, 1997).

Results of assessments of rhythmic pattern recognition and reproduction by eight implant users were compared with corresponding results obtained by seven normally hearing subjects in a study reported by Schulz and Kerber (1994). The recognition task required listeners to identify patterns representative of four common musical rhythms (such as those associated with a waltz or a tango), whereas the reproduction task required listeners to repeat by tapping several distinctive rhythmic patterns comprising three or five beats. Average scores for each group of subjects were at least 80% correct for both types of assessment; however, no statistical analyses were presented to determine whether any significant differences in performance existed between the two groups.

More recently, Leal *et al.* (2003) assessed rhythm perception by 29 recipients of the Nucleus 24-electrode device. Twenty subjects used the ACE sound-processing scheme, while the remainder used the SPEAK scheme. The test was essentially similar to the PMMA rhythm subtest but with fewer test items. However, in addition to a discrimination test (i.e., indicating whether pairs of sound sequences had the same or different rhythm), an identification test was also done in which subjects were asked to indicate where in each sequence the rhythmic change occurred. Individual results were classified into "good" and "poor" categories depending on whether the subjects' scores were greater or less than two criterion scores, which were 90% and 75% correct for the rhythm discrimination and identification tests, respectively. On this basis, 24 of the 29 subjects obtained good performance on the rhythm discrimination test, whereas only 12 of the subjects obtained good performance on the rhythm identification test. No differences in performance related to the use of the two sound-processing schemes were reported.

In another test of rhythm perception, 3 implant users and 4 subjects with normal hearing were instructed to identify one of seven distinct rhythmic patterns (Kong *et al.*, 2004). Two of the implanted subjects used the Nucleus 22-electrode device and the SPEAK sound-processing scheme, while the third used the Clarion I system (Advanced Bionics, Sylmar, CA) with the CIS scheme. The normally hearing subjects obtained scores near 100% correct on the test. One Nucleus implant recipient obtained similar near-perfect scores, but the remaining 2 subjects had scores that were 10 to 25 percentage points lower. Interestingly, scores for each subject were very similar for 4 subtests in which the same test materials were presented at different overall speeds (60–150 beats per minute).

In a related study, the performance in discriminating differences in tempo was assessed with five implant users, including two Ineraid recipients using the CIS scheme (Kong *et al.*, 2004). On average, their results were very close to those obtained with the same four normally hearing subjects mentioned above. The change in tempo that was discriminable by the implant users was approximately 4 to 6 beats per minute across the range of tempi (60–120 beats per minute) presented in the tests.

A study with an unusual experimental procedure compared the temporal integration characteristics of 11 single-channel cochlear implant users with those of matched normally hearing listeners (Szelag *et al.*, 2004). The subjects were asked to accentuate mentally a rhythmic pattern within a sequence of regular tone bursts presented at several burst rates between one and five beats per second. The implant users' ability to integrate rhythmic patterns subjectively at beat rates below three beats per second was significantly poorer than that of the participants with normal hearing. At higher beat rates, the performance of the two subject groups was similar.

Although these findings appear somewhat inconsistent with those cited above, which generally reported that performance on rhythmic pattern perception tasks was similar for implant users and normally hearing listeners, it seems possible that the differences may be related mainly to the experimental methods employed in each study to determine subjects' perceptual abilities. In particular, it is difficult to interpret the results of the subjective experiment described by Szelag (2004) directly in terms of musical rhythm perception.

### 3.2. Perception of Melody

#### 3.2.1. Tune Identification

What enables people to recognize melodies? First, there is the question of which tunes are sufficiently familiar to a listener such that he or she would be able to name them on hearing them. This ability depends on a range of highly variable factors, such as the individual's musical training and listening experience, the social culture within which that experience has been gained, and the person's memory of both the tunes and their titles. Recognition is also likely to be affected by the situational context in which the music is heard. For example, in the Western musical tradition, *Happy Birthday* is rated amongst the most familiar melodies for the general population (Gfeller *et al.*, 2002a; Looi *et al.*, 2003), and it is immediately recognizable by nearly everyone in the appropriate circumstances regardless of the intonation of the notes, the correctness of the rhythm, or the acoustical quality of the listening situation. Thus, the ability to perceive accurately fundamental features of musical sounds, such as pitch and temporal patterns, is not always a prerequisite for melody recognition.

As previously summarized, the performance of most cochlear implant users in formal tests of rhythm perception is reported to be similar to that of listeners with normal hearing. This observation leads to the expectation that implant users would be able to recognize melodies that have a distinctive rhythmic pattern more readily than melodies that are less rhythmic. This was confirmed in a study involving eight users of a single-channel cochlear implant with an analog sound-processing scheme (Schulz and Kerber, 1994).

In that study, only four different tunes (well-known children's songs) were presented in several different musical arrangements. Not surprisingly, the normally hearing subjects obtained an average recognition score of close to 100% correct across all of the melodies. The average score for the implant users was much lower (about 50% correct). When the implant users' results were divided equally between those for rhythmically structured songs and those for songs without a distinctive rhythm, a score difference of about 15 percentage points was found in favor of the rhythmic tunes.

A similar pattern of results was reported for a study in which 12 well-known tunes were presented to 49 multiple-channel cochlear implant

recipients and 18 normally hearing subjects (Gfeller *et al.*, 2002a). The implant users listened to the test materials through their own sound-processing devices, which were programmed with either the ACE, CIS, or SPEAK schemes. The overall average melody recognition score for the implant users was approximately 19% correct, whereas the corresponding score for the subjects with normal hearing was about 83%. For each subject group, the average score for melodies classified as rhythmic was approximately 12 percentage points higher than the score for arrhythmic melodies. No significant differences in performance were found for the implant users that could be related to the type of sound-processing scheme implemented in the device.

Kong *et al.* (2004) published further evidence supporting the relative importance of rhythm information for melody recognition by implant users. In their experiments, 6 multiple-channel cochlear implant users were asked to identify 12 familiar songs heard with and without rhythmic cues. Six subjects with normal hearing, who also participated in the study, obtained near-perfect identification scores when the melodies were presented in both conditions. However, the average score for the implant users was only about 63% correct when rhythmic cues were available. When the rhythmic cues were eliminated by equalizing the duration of each note and the silent intervals between notes, the implant users' average performance was reduced to chance levels.

An earlier study with 8 users of the 22-electrode Nucleus implant system, programmed with the now-superseded MPEAK strategy for 7 subjects and the SPEAK scheme for the remaining subject, investigated recognition of familiar melodies in several musical contexts (Fujita and Ito, 1999). On average, subjects obtained higher scores for closed-set identification of songs when played with words than when played with only an instrumental sound. Discrimination of melodies that lacked rhythmic or verbal cues was relatively poor, with average performance at chance levels.

The notion that songs may be identified more readily when they contain meaningful words seems reasonable in terms of the generally satisfactory performance obtained by many users of recent cochlear implant systems for understanding speech, even in moderately noisy conditions. Recognition of just a few of the words in a well-known song may be sufficient for many listeners to name it correctly.

Identification of a small set of familiar songs was investigated with 29 recipients of the 24-electrode Nucleus system using either the ACE or SPEAK schemes (Leal *et al.*, 2003). The test material was presented with and without sung words, and presumably contained at least some items that also had distinctive rhythmic patterns. Either seven or eight melodies were presented to each subject, depending on their familiarity with the available material. When the melodies were played by an orchestra without verbal cues, only one subject could identify more than half of them in a closed-set procedure. However, 28 of the subjects could identify at least half of the songs when the words were sung with an orchestral accompaniment.

A more recent study (Looi *et al.*, 2004) included results from an experiment in which 15 Nucleus implant users listened to 10 familiar melodies presented without verbal cues or accompaniment. However, the melodies were played with normal rhythmic content as well as appropriate pitch sequences. Six subjects used the SPEAK sound-processing scheme, while the remainder used ACE. They were asked to identify each tune from a closed set. Overall, the averaged results showed that the implant users only correctly recognized about half the tunes, whereas normally hearing listeners who performed the same task scored nearly 100% correct. An analysis of the individual responses of the implant users to each melody suggested that both rhythmic and pitch information probably contributed to the subjects' recognition performance.

### 3.2.2. Melodic Pattern Recognition

The task of discriminating between different pitch contours is related to melody identification, but is generally more difficult because of the reduced number of auditory cues available in the test material. In a typical melodic pattern recognition experiment, listeners are asked to label two pitch sequences as the same or different. The notes forming each pair of sequences are presented with identical rhythms, and no coincident verbal cues such as sung words are presented. Thus, discrimination relies on the listener's ability to perceive a pattern of changes in pitch. However, neither the absolute nor the relative pitch of each note needs to be perceived accurately for discrimination of the two sequences to be possible. For example, detection of an overall pitch contour, such as perception of a generally rising or

falling pitch across each entire sequence of notes, may be sufficient for a listener to discriminate the sequences.

Results from an experiment in which implant users were asked to determine if a musical scale was played ascending or descending were reported by Dorman *et al.* (1991). The subjects were 16 users of the Ineraid multiple-channel implant and the CA sound-processing scheme. Most of them were unable to discriminate these pitch sequences reliably. In contrast, eight users of a single-channel device did obtain higher-than-chance scores on a similar test (Schulz and Kerber, 1994). In that study, subjects with normal hearing tested with the same procedure obtained average scores close to 100% correct. These scores were about 15 to 30 percentage points higher than those of the implant users.

A subtest of PMMA has also been used to assess implant users' ability to discriminate pitch patterns. The test procedure is similar to that of the PMMA rhythm subtest described earlier. For the so-called tonal subtest, the material comprises pairs of short sequences of notes that have identical rhythm. The pattern of note pitches within each pair of sequences is either the same or different, and the listener is asked to label each pair accordingly. This procedure was carried out with 8 users of the Ineraid implant with the CA sound-processing scheme, and 10 users of the Nucleus 22-electrode implant with the F0/F1/F2 feature-extraction strategy (Gfeller and Lansing, 1991). As noted previously, both of these systems are now obsolete.

The average score obtained by all subjects on the tonal subtest was 78% correct. Interestingly, this was 10 percentage points lower than the subjects' mean score for the rhythm subtest obtained in the same study. When the PMMA has been conducted with normally hearing subjects, scores reported for the tonal subtest tend to be higher than for the rhythm subtest (Gfeller and Lansing, 1991).

A further study using a modified version of the PMMA compared performance on the tonal subtest between two sound-processing schemes formerly used with the Nucleus 22-electrode device (MPEAK and the F0/F1/F2 strategy) (Gfeller *et al.*, 1997). The participants included 17 implant users and 35 normally hearing subjects. The mean score for the implant users was approximately 77% correct and did not differ significantly between the two sound-processing strategies.

In contrast, the average scores for listeners with normal hearing were about 91% correct on the same test.

Another study that compared perception of pitch patterns between two sound-processing schemes used a set of isolated stimuli that varied in the way the fundamental frequency (F0) changed over time (McKay and McDermott, 1993). The stimuli were voiced phonemes produced with a rising, steady, or falling F0 contour. Four users of the Nucleus 22-electrode device were tested when listening with either the MPEAK strategy or an experimental prototype of the SPEAK scheme (i.e., the SMSP scheme mentioned earlier). As outlined previously, the MPEAK strategy converted an estimate of F0 directly into the stimulation pulse rate, whereas the SPEAK scheme employs a constant rate of stimulation. Despite this functional difference, the ability of the subjects to identify the pitch contours in the experiment was similar, on average, for both types of sound processor.

More recently, an assessment similar to the PMMA tonal subtest was conducted with 29 users of the Nucleus system and either the ACE or SPEAK schemes (Leal *et al.*, 2003). The test contained 12 pairs of pitch sequences. About two-thirds of the subjects obtained discrimination scores of at least 90% correct on this test. In a related test, the same subjects were asked to describe whether the pitch in each sequence became higher or lower, and to indicate where within the sequence the pitch change occurred. The mean score for all subjects on this test was about 73% correct. However, because there appear to be no reports of similar tests having been conducted with normally hearing listeners, these findings are rather difficult to interpret.

---

### 3.3. Perception of Timbre

---

#### 3.3.1. *Timbre Recognition*

One standard definition of timbre is “that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar” (ASA, 1960). Less strictly, timbre can be described as the quality that characterizes differences in tone (or “tone color”) that are apparent when musical notes are played with the same pitch and loudness on several different instruments. The definition can be generalized to

include the perceptual effects of a wide range of properties of acoustic signals (Pratt and Doak, 1976; Grey, 1977). The principal properties are the frequency spectrum and the amplitude envelope of sounds, including changes in those attributes over time, although other characteristics, such as the spatial configuration of sound sources, may also be relevant. However, most published studies on the perception of timbre by users of cochlear implants seem to have focused on the ability of listeners to identify or discriminate the sounds of different musical instruments.

In the study of Schulz and Kerber (1994), eight users of an analog single-channel implant system were asked to identify the instrument playing a melody from a closed set of five alternatives. Even though a small number of different instruments were used in the test, the subjects obtained an average identification score of only about 35% correct. In contrast, listeners with normal hearing scored approximately 90% on the same test.

Gfeller *et al.* (2002b) reported results from an instrument identification test carried out with 51 implant recipients using a variety of device types and sound-processing schemes. Twenty normally hearing listeners also completed the test. The sound stimuli were recordings of eight different instruments playing the same brief sequence of notes. Subjects selected each of their responses from a set of 16 possible alternatives. The implant users obtained an average score of 46.6% correct on the test. This result was significantly lower than the mean score of 90.9% obtained by the subjects with normal hearing. Furthermore, the confusions present in the implant users’ responses displayed a diffuse pattern, whereas the errors made by the normally hearing subjects were more often confusions between instruments within the same family (i.e., brass, woodwind, percussion, or strings), rather than across instrument families.

Recordings of only three different instruments were used in an identification test reported by Leal *et al.* (2003). The same melody was played in a similar pitch range and in a similar style on each of the instruments (i.e., trombone, piano, and violin). Subjects were asked to name the instrument after hearing each recording. Twenty of the 29 users of the Nucleus implant system (with either the ACE or SPEAK sound-processing schemes) who participated in the study identified all three instruments correctly. All except one of the remaining subjects could identify two of the instruments.

In a recent study, 10 recipients of Nucleus implants, all users of the SPEAK sound-processing scheme, were asked to identify 16 different musical instruments in a closed-set procedure (McDermott and Looi, 2004). Recognition scores varied widely, both among subjects and across instrument types.

Figure 7 is a confusion matrix that shows the results for all subjects. It includes the instrument sounds that were tested and the subjects' responses. The instruments were divided equally into a percussive and a nonpercussive group. Overall, the average score for identification of all instruments by the implant users was approximately 44% correct. In contrast, subjects with normal hearing obtained a mean score of 97% on the same test.

As shown in Figure 7, some instruments, such as the drums or xylophone, were identified correctly much more often by the implant users than other instruments, such as the organ or flute. Not surprisingly, more confusion occurred among instruments within the same group (i.e., percussive or nonpercussive) than between groups. For example, the organ was recognized least often out of all 16 instruments, but most of the subjects' in-

correct responses named the violin; none of the incorrect responses included the tambourine or drums. This pattern of results confirms the relative salience for implant listeners of temporal envelope or rhythmic cues in musical sounds in comparison with other timbre or pitch cues.

### 3.3.2. Timbre Appraisal

In assessments of timbre appraisal, as distinct from recognition, subjects may be asked to describe the quality of musical instrument sounds by using adjectives such as "beautiful," or "clear," or to assign ratings, usually numbers, to the sound quality. The rating scales are typically based on one or more subjective descriptors, such as "pleasantness," or "naturalness."

A numerical rating scale that asked listeners to indicate how much the sound quality of 25 different instruments appealed to them was used in the study of Schulz and Kerber (1994). The participants included eight users of an analog single-channel implant system and seven subjects with normal hearing. Although the average quality rating of the implant users for all instruments was significantly lower than that of the normally hearing subjects, the pattern of ratings across the dif-

	Flute	Clarinet	Violin	Double Bass	Trumpet	Organ	Female Singer	Male Singer	Piano	Bells	Xylophone	Harp	Guitar	Bass Drum	Drums	Tambourine
Flute	<b>12</b>	19	13	1	25		2	3		1	4					
Clarinet	4	<b>22</b>	23	1	12	1	3	1	3	1	3	2	1	3		
Violin	4	9	<b>40</b>	2	7		2	1	3		3	4	1	1	3	
Double Bass	1	8		<b>45</b>		1			4			4		12	5	
Trumpet	2	10	3		<b>49</b>	1	7	3		1	1	1			2	
Organ	3	12	25	1	8	<b>8</b>	2	4	4	5	1	2	3	2		
Female Singer							<b>26</b>	<b>52</b>		1		1				
Male Singer		1	5	1	1	1	12	<b>59</b>								
Piano	3	1	2			7			<b>48</b>	4	2	3	5	3	2	
Bells	3	3	3		3	2			3	<b>38</b>	13		1	5		6
Xylophone	2	1			7	1			5	10	<b>53</b>	1				
Harp			2	8		9			30	1	2	<b>21</b>	2	5		
Guitar	1	2	5	2	4	4			16	1		1	<b>21</b>	7	16	
Bass Drum	1			6	1	1			18		4	2	1	<b>27</b>	18	1
Drums		2		1			1			1	4		4	5	<b>59</b>	3
Tambourine					2					11	14			7	17	<b>29</b>
TOTAL	36	90	121	68	119	36	55	123	134	75	104	42	39	77	122	39

**Figure 7.** Confusion matrix showing the results of an experiment in which 10 implant users identified 16 musical instruments in a closed-set procedure. The instrument sounds that were presented are shown in the left column. They were divided into two equal groups: eight were percussive (lower half), and the rest were non-percussive (upper half). Each subject heard each instrument sound a total of eight times. Their responses are shown in each cell of the matrix. The maximum possible score is 80 (10 subjects  $\times$  8 repetitions). Responses in cells that form the main diagonal of the matrix are correct identifications of the instruments; these are shown in bold type. The total number of times each instrument was named in the subjects' responses is shown in the row at the bottom of the matrix.

ferent instrument types was similar for the two subject groups. It seems likely that this pattern represented idiosyncratic variations in the listeners' liking of each musical instrument rather than a characteristic of the mode of hearing applicable to the subjects in each group or functional details of the sound-processing scheme used by the implant recipients.

Gfeller and Lansing (1991) applied a questionnaire, the "Musical Instrument Quality Rating" form, to obtain simple descriptions of the perceived quality of nine instruments. As mentioned previously, 10 of the 18 subjects who participated in the study were users of the Nucleus F0/F1/F2 feature-extraction speech processing strategy, while the remaining 8 subjects were users of the Ineraid CA scheme. Both of these sound-processing techniques have since been superseded. Nevertheless, the proportion of Ineraid users who rated each of the instrument sounds as "beautiful" or "pleasant" was greater than the corresponding proportion of Nucleus users.

Although the number of participants in each of the two groups was small, it seems plausible that the now-obsolete feature-extraction scheme used by the Nucleus implant may have been less effective than alternative or newer sound processors at transmitting some characteristics of acoustic signals that contribute particularly to perceived sound quality.

Few published reports appear to have examined directly the differences between several sound processing schemes when used by the same implant recipients for listening to music. One study with 63 Nucleus implant users compared the perceptual performance of the SPEAK scheme with that of MPEAK (Skinner *et al.*, 1994). Although the experiments described in that report were aimed mainly at investigating differences in speech recognition associated with the use of each sound processor, the study included a questionnaire that enabled the participants to rate the processors subjectively based on listening situations encountered commonly in their everyday lives. One of the situations the questionnaire addressed was listening to music. The results showed that 83.9% of the subjects preferred the SPEAK scheme to the MPEAK strategy. None preferred MPEAK over SPEAK, although 10.7% stated that the two schemes were about the same for listening to music. The responses of the remaining subjects suggested that they were unable to make a definite judgment for that condition.

In a recent comparison of timbre appraisal between implant users and listeners with normal hearing, two types of measures were obtained (Gfeller *et al.*, 2002b). The first was a rating of overall pleasantness on a scale of 0 to 100. The second obtained separate ratings for three perceptual dimensions: *dull-brilliant*, *compact-scattered*, and *full-empty*, using similar numerical scales.

The sound stimuli were recordings of eight different musical instruments representing the brass, woodwind, and strings (including piano) instrument families. The results of the first experiment showed that, on average, implant users gave ratings that were about 17 points lower than the normally hearing listeners. The pattern of ratings across instrumental families was generally similar for the two subject groups, although the implant listeners gave particularly low ratings to the stringed instruments.

The results of the second experiment were consistent with this finding, showing that the implant users rated the strings as poorer in quality (i.e., more scattered, less full, and more dull) on all three of the perceptual dimensions. Furthermore, compared with the normally hearing listeners, they rated the higher-pitched instruments as sounding more scattered and less brilliant.

Ratings of liking and subjective complexity were compared between cochlear implant users and listeners with normal hearing by Gfeller *et al.* (2003). The test stimuli included excerpts of music representing three genres: classical, country-western, and pop. The results showed that, on average, the scores for overall appraisal (i.e., liking) given to classical music by the implant users were significantly lower than those given by the normally hearing listeners. Furthermore, the implant users rated excerpts of country-western and pop music as significantly more complex. Appraisal scores for these two genres and complexity ratings for classical music were generally similar for the implant users and the normally hearing subjects.

#### 4. Effects of Training

Some researchers have investigated whether it might be beneficial to train implant users to improve their perception of certain essential characteristics of music. The rationale is principally

that when implant recipients use their devices in their everyday lives, they are likely to gain more experience in hearing speech and in learning to understand it than in becoming familiar with music. Thus, the improvements expected in speech recognition with use of a hearing prosthesis over time may not be matched by improvements in perception of music.

This is supported by evidence that some implant users obtain less enjoyment from listening to music postimplantation than they recall from the time before their hearing had deteriorated to levels at which a cochlear implant became the most appropriate form of treatment. For example, in one study, Leal *et al.* (2003) reported that only 21% of experienced implant recipients agreed that they enjoyed listening to music and took opportunities to listen to it. In contrast, of the subset of those subjects who were able to describe their listening interests before losing their hearing, 41% agreed with the same statement.

In another study, Gfeller *et al.* (2000b) reported that about one-third of implant users stated that they tended to avoid music because of its aversive sound quality.

Gfeller *et al.* (1999) and Gfeller (2001) described an aural training program that they had developed specifically to provide implant recipients with structured experience in listening to musical sounds. The program was designed to be self-administered using a personal computer and consisted of 48 sessions (or "lessons"), scheduled 4 times per week. The training materials included musical stimuli containing predetermined pitch sequences and recorded sounds of different instruments. The training procedures included tasks that were designed to help implant listeners to discriminate, identify, or accurately describe the materials.

In a study that investigated the effects of applying this training program in an attempt to improve the perceptual abilities of implant users when listening to music, appraisal ratings were obtained for a set of complex songs representing various musical genres (Gfeller *et al.*, 2000a). Appraisal was measured on two scales: liking and complexity. The scales had endpoints of 0 and 100, with larger numbers indicating better liking and greater perceived complexity.

Twenty-four recipients of the Nucleus 24-electrode cochlear implant system participated. They were divided into a control group and a training group; only the latter completed the 12-

week training program. Results for all subjects showed an average liking rating of about 56, and an average complexity rating of about 41. The training program appeared to produce small but significant positive effects, with an increase in liking of approximately 6 points (on the 0–100 scale), and a reduction in perceived complexity of approximately 4 points. The authors of the study argued that a lower rating of complexity was associated with better appreciation of musical sounds.

The study also assessed recognition of melodies (Gfeller *et al.*, 2000a). The results of the experiments showed that the subjects who had participated in the training program obtained an average score increase of approximately 11 percentage points for identifying simple melodies, and about 33 percentage points for complex songs. The average recognition score for the control subjects was only approximately 5% correct for the same tasks.

Although the effects of training were large and statistically significant, the ability of subjects to generalize their learnt ability to recognize melodies was less clear. For example, subjects correctly recognized only a few of the previously unfamiliar simple melody items when they were tested after the completion of the training program. However, the same subjects correctly recognized a larger proportion of unfamiliar complex songs after training.

The same training program was used to determine whether it would improve recognition and appraisal of musical timbre for implant recipients (Gfeller *et al.*, 2002c). The assessments included a test of the subjects' ability to recognize which of 8 instruments was being played from sound recordings; responses were restricted to 16 possible instruments. In addition, two measures of appraisal were obtained: a rating of overall pleasantness, and three separate ratings for specific perceptual dimensions of the sounds, as described previously (Gfeller *et al.*, 2002b).

The results of the instrument identification test showed that completing the training program had the effect of increasing the average recognition score by nearly 20 percentage points. No increase in recognition ability over a similar interval of time was found for the control implant users, whose average score remained at about 33% correct. Overall quality ratings by the participants in the training group also increased significantly, whereas those of the control subjects did not.



However, the specific effects of training on the appraisal ratings for the three separate perceptual dimensions were found to be generally small.

Few reports appear to have been published on the music listening experience and skills of children with cochlear implants, even though a large proportion of implant recipients are children. Of interest is that information obtained from a questionnaire (Gfeller *et al.*, 1998) suggested that many children who were implant users were involved in either formal or informal musical activities. A music training program designed specially for children with implants was described and evaluated by Abdi *et al.* (2001). The program involved children in either perceptual tasks (learning to discriminate rhythms and pitches) or production tasks (learning to play a simple musical instrument). Although a detailed description of a formal evaluation of this program's effectiveness does not seem to have been published, brief reports of the musical development of 14 children participating in the program suggested that it might have been beneficial.

## 5. Psychophysical Studies Relevant to Music Perception

The research studies discussed above may be generally summarized as follows:

First, perception of musical rhythm by users of cochlear implants has been found to be similar to that of listeners with normal hearing. This is not surprising when the results of relevant psychophysical studies are considered. Perception of rhythm in music is related to the perception of the duration of sounds and the gaps between sounds. To perceive rhythm patterns in most types of music adequately, the temporal resolution required for either duration or gaps is probably on the order of tens of milliseconds (ms).

Several psychophysical studies investigating perception of synthetic, nonmusical signals have shown that most implant users have sufficient ability to resolve temporal changes in signals for perceiving musical rhythms. For example, the gap detection threshold for simple signals of moderate loudness has been reported to be usually less than 10 ms (Shannon, 1989; Shannon, 1993), although it may increase beyond 50 ms for signals that are very soft.

Second, users of implant systems typically have great difficulty recognizing melodies, even when the tunes are familiar and are played as a sequence of isolated notes without accompaniment or harmony. If distinctive rhythm patterns are noticeable in the tunes presented in the tests, those patterns appear to provide most of the information that implant recipients use when they identify the melodies. These findings provide evidence of one of the most serious problems that confronts implant users when they listen to music: pitch information is conveyed very poorly. Understanding why this problem is present in existing cochlear implant systems, and what might be done in practice to alleviate it, are substantial topics that are further discussed later.

Third, perception of timbre has generally been reported as much poorer for implant users than for listeners with normal hearing. The major finding is most implant recipients, using only auditory cues, cannot readily identify the musical instrument that is played to them but can sometimes discriminate between instruments when differences in the temporal envelope of the sounds are obvious, for example, distinguishing the sound of a flute from that of a drum. This suggests that information concerning the spectral shape that characterizes musical instrument timbres is represented only crudely in the electric stimuli generated by existing implant systems. This topic is also discussed later in more detail.

Finally, appraisal ratings of musical sounds that indicate the subjective pleasantness of the sounds (i.e., how much listeners like them) have been reported as lower for implant users than with normally hearing listeners. The application of specific music training programs can help improve appraisal ratings. However, it seems likely that subjective judgments of the quality and pleasantness of musical sounds will remain relatively poor unless better information about pitch and timbre can be made available to implant users. Some practical suggestions about how such improvements might be achieved are presented towards the conclusion of this article.

---

### 5.1. Perception of Pitch

---

#### 5.1.1. Introduction

Previous studies have shown that users of multiple-electrode cochlear implants may perceive pitch in two fundamental ways:

- The primary mechanism relies on rapid temporal fluctuations in electric stimuli. The percept associated with such temporal patterns is often called *rate pitch*, although a similar pitch percept exists that is related to modulations in the envelope of a carrier stimulus. Typically the amplitude-modulated carrier consists of a train of pulses presented at a relatively high rate.
- The secondary pitch mechanism depends on the position in the cochlea at which the electric stimulus is delivered. The associated percept is usually called *place pitch*.

However, some researchers have questioned whether varying the place of stimulation elicits a change in the perceived pitch that is able to convey melodic information; instead, they suggest that changes in place mainly affect the perceived timbre (McDermott and McKay, 1997; Moore and Carlyon, 2005). In experimental research with cochlear implant users, it is nearly always impossible to distinguish absolutely between changes in pitch and changes in timbre. Therefore, this distinction will be set aside temporarily in the brief review of relevant psychophysical studies that follows.

First, however, it is important to clarify a few terms that are widely used in the literature published on these topics. In experiments that require subjects to detect whether two sounds differ, or which one of three or more sounds differs from the others, the ability under investigation is *discrimination*. The ability to discriminate sounds does not imply that the sounds differ in some predetermined characteristic, such as pitch or timbre. In practice, subjects may use any perceptible differences between the sounds to perform the experimental task.

However, if subjects are asked to listen to two sounds presented in sequence, and to judge which one has the higher pitch, the procedure is often called *pitch ranking*. The experimental context, or the parameters of the stimuli, assumes that the varying sound quality used by the subjects in the task is, in fact, pitch. Of course, it is possible that some other quality of the signals that changes, such as timbre or even loudness, might also enable subjects to successfully rank the stimuli.

Pitch experiments can also involve *scaling* or *identification*, in which subjects are asked either to assign numbers in an orderly way to the per-

ceived pitch of each of a set of sounds (scaling) or to recognize and label each one of a small number of sounds (identification). Studies that use ranking, scaling, and especially identification provide only limited information about musical pitch. In particular, no information can be obtained from these procedures about the perceived size of musical intervals (e.g., whether two signals that differ in frequency by a factor of two are perceived as spanning one octave).

The ability to perceive interval size accurately is crucial to music appreciation; for instance, melodies will sound out of tune if the intervals are heard incorrectly. Unfortunately, precise judgments of interval size require listeners to have received considerable formal musical training before they received their cochlear implant, and to have retained that knowledge in a form that is applicable to the unnatural signals heard with the device. Consequently, only a few studies have been published that have investigated this most important aspect of musical pitch perception.

#### 5.1.2. Temporal Pitch Mechanisms

The simplest electric stimuli that have been used by psychophysical researchers who investigate auditory perception with cochlear implants include sine waves and regular pulse trains (see Figure 2). Usually these signals are delivered to a single cochlear location. In single-channel implants there is, of course, no way of changing the site of stimulus delivery, but such variations are possible with multiple-electrode devices. To select a single stimulation site in these implants, either one electrode is activated in a monopolar (or common ground) configuration, or two closely spaced electrodes are used in a bipolar mode (see Figure 3). Although stimuli may be delivered to only one cochlear position at a time, the site is often a parameter that is varied systematically in the experiments by the selection of different active electrode positions.

Numerous researchers have reported that varying the rate of a steady pulse train (or the frequency of a sinusoidal stimulus) presented at one cochlear site results in a change in the pitch perceived. Typically, the pitch increases with increasing rate over a range from about 50 to 300 Hz, although the upper limit varies across electrode positions and among implant recipients (Moore and Carlyon, 2005). At lower rates, the signal tends to be perceived as a buzz or fluttering sound that does not seem to have a salient

pitch. With very high rates, the pitch of the percept is affected only slightly by changes in the rate, and may instead be dominated by the location of the active electrode.

As described previously, none of the sound-processing schemes currently used most commonly with cochlear implants (i.e., ACE, CIS, and SPEAK) are designed to vary the rate of stimulation to represent some feature of the acoustic signal. Certain earlier speech-processing strategies, such as MPEAK, used an estimate of the voice F0 to control stimulation rate, but such feature-estimating schemes have since been superseded.

The newer, constant-rate pulsatile stimulation schemes may, however, also present some information about F0 in the temporal domain. Typically, they use a relatively high stimulation rate, and modulate the amplitude of the pulse trains on each active electrode in accordance with the estimated envelope (or amplitude) of the input signal in each of several corresponding frequency bands. The frequency analysis of the input signal is usually performed by a bank of bandpass filters or a digital spectrum estimation technique. In any case, the envelope of the signal in each band generally contains modulations arising from the fundamental frequency of the input signal. Psychophysical studies have been carried out using idealized forms of these stimulation patterns to determine whether pitch information can be derived from amplitude-modulated, high-rate pulse trains (see Figure 8).

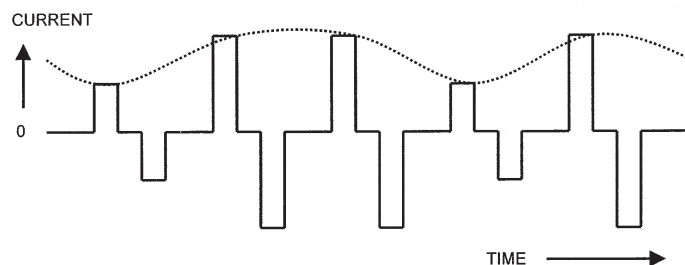
In general, the results of the studies have shown that a pitch is associated with the fre-

quency of the modulation (McKay *et al.*, 1995; McKay, 2004). The range of frequencies that produce a systematic variation of pitch is similar to that found for changes in pulse rate at low rates (about 50–300 Hz). With amplitude modulation, the rate of the carrier pulse train may also affect the pitch perceived. In particular, to avoid anomalies in the relationship between the pitch and the modulation frequency, the carrier rate must be at least four times higher than the modulation frequency (McKay *et al.*, 1994).

Most of the above studies have employed pitch-scaling procedures. They suggest that musical pitch information may be derived from temporal patterns in electric stimuli delivered to a single intracochlear site over a restricted, relatively low range. For most implant users who have participated in these experiments, the range encompasses only approximately the two to three octaves below middle-C on the piano keyboard.

However, the range of frequencies resulting in monotonic pitch changes for implant users provides no information about the smallest detectable frequency difference. In the Western musical scale, the smallest interval used in the construction of melodies is one semitone, which equals one-twelfth of an octave (on a logarithmic frequency scale). Thus, two notes that differ in pitch by one semitone are separated by a frequency ratio of about 5.95%. For instance, under conventional tuning, middle-C has a fundamental frequency of 261.6 Hz and the note C-sharp, one semitone above middle-C, has a fundamental frequency of 277.2 Hz. It may be inferred that, for implant listeners to recognize the pitch changes in melodies correctly, they would need to have an ability to discriminate frequency changes smaller than approximately 6%.

In one early study with two subjects who had received a four-electrode implant in the auditory nerve, rate difference limens (i.e., rate ratios producing just-noticeable perceptual differences) were found to be around 5% for rates below about 350 Hz (Simmons *et al.*, 1981). In a review of 5 more recent studies from which data were obtained with a total of 19 subjects (Pfungst *et al.*, 1994; van Hoesel and Clark, 1997; McKay and McDermott, 1999; McKay *et al.*, 2000; Zeng, 2002), Moore and Carlyon (2005) reported an average rate difference limen of 7.3% at a rate of 100 Hz. However, the results varied greatly among subjects, and were also dependent on the details of the procedure applied in the experiments.



**Figure 8.** Illustration of an amplitude-modulated current pulse train. The carrier is a sequence of biphasic pulses delivered at a constant rate, as in Figure 2 (upper panel). The level of each pulse is determined by the amplitude of a signal waveform, which is shown as the dotted line.

Even the best performance found in these subjects (a rate difference limen of less than 2%) was much poorer than the performance of a typical listener with normal hearing (frequency difference limen of less than 1% for a 100-Hz pure tone). Nonetheless, it would be expected that, if fundamental frequency was converted to a purely temporal code, such as the stimulation pulse rate, many users of cochlear implants should be able to recognize familiar melodies using pitch cues.

In a study involving 17 implant users, Pijl and Schwarz (1995) found an average of 44% of common tunes were recognized correctly when the fundamental frequencies of the notes were converted directly to pulse rates. Initially, the tunes were presented with intact rhythm cues. In a further experiment with three of the subjects, tune identification was tested with modified melodies in which the only information available was conveyed by variations in the pulse rate. With rhythm cues eliminated, these subjects obtained scores close to 100% correct at low pulse rates. Their performance varied slightly according to which electrode was activated, and dropped towards chance levels as the overall stimulation rate was increased. Consistent with other reports, it was found that the maximum rate at which useful pitch information was conveyed varied among the subjects over a range of about 300 to 600 Hz.

With the same three subjects, two of whom had previously received some musical training, an additional experiment investigated whether pulse rate ratios produced the expected musical intervals (Pijl and Schwarz, 1995). The subjects were asked to label pairs of stimuli that comprised steady pulse trains delivered to single electrode positions in terms of conventional musical intervals. Only the pulse rate changed between the two stimuli in each pair. The subjects judged the perceived intervals by using their memory of salient intervals in well-known tunes; for example, the first pitch change in *Twinkle, Twinkle, Little Star* is a fifth, and corresponds to a frequency increase of approximately 50%.

The results showed, on average, that the subjects were very accurate in labeling the intervals, although a trend was noted for them to overestimate the size of the pitch interval as the overall pulse rate was increased. Changing the active electrode position had little effect. However, in another study (Pijl, 1997a), it was found that the ability of two implant users to match two stimuli in pitch, using only variations in pulse rate, deteriorated as

the spatial separation between the two electrodes activated by the two stimuli was increased.

Interestingly, a further study evidently conducted with the same two subjects (Pijl, 1997b) used a similar interval-labeling procedure to assess musical pitch perception when notes played on a piano were presented via the SPEAK sound-processing scheme. In contrast to the results obtained when the stimuli were controlled in pulse rate and delivered directly to fixed electrode positions, almost no pitch information seemed to be available when the subjects listened to these more realistic musical sounds through the processors that they used ordinarily. The poor outcome with the SPEAK scheme in this experiment is most likely explained by its use of a nearly constant (and relatively low) pulse rate, the fact that the electrodes selected for activation vary rapidly in time, and the use of an acoustic input signal that was spectrally and temporally complex.

The perception of musical intervals was also studied in a series of experiments with one implant user who had been trained as a musician and piano tuner before his hearing deteriorated (McDermott and McKay, 1997). The apparently unique background of this subject enabled experiments to be carried out that did not depend on his recollection of common melodies. Instead, he was able both to estimate and to produce musical intervals using standard terminology (i.e., "fifth", "octave", etc).

In experiments investigating temporal pitch information, single electrodes were activated with sequential pairs of stimuli; no conventional sound-processor was used. Either the stimulation pulse rate or the frequency of amplitude modulations applied to a high-rate carrier pulse train was varied to change the pitch perceived. Consistent with previously published results, both types of temporal manipulation of the signals were found to elicit changes of pitch of about the expected size; for example, a fifth was associated with a rate ratio or modulation frequency ratio of approximately 50%.

#### 5.1.3. *Pitch Associated with Place of Stimulation*

As mentioned earlier, when the frequency of a pure tone presented to an acoustically hearing ear is varied, changes occur in both the temporal and spatial patterns of the resulting neural excitation. At very low frequencies, it is likely that information about frequency is conveyed mainly in the temporal patterns of neural activity. In contrast,

for very high frequencies, most such information is probably represented in the spatial configuration of the nerve fibers that are maximally excited. However, for a wide range of intermediate frequencies (approximately 50–5000 Hz), changes in frequency result in closely related changes to both spatial and temporal aspects of the neural excitation pattern.

Unlike acoustic stimulation, electric stimulation generated by a cochlear implant can be controlled, at least in principle, to separate the temporal and spatial patterns of neural activity (McKay *et al.*, 2000). The studies reviewed previously attempted to maintain a constant place of stimulation while experimentally manipulating the temporal patterns.

The converse has also been investigated. Several publications have confirmed that the pitch perceived generally increases when a constant-rate pulse train is presented on one electrode at a time, and the position of the active electrode is moved from an apical to a more basal location in the cochlea (Tong and Clark, 1985; Townshend *et al.*, 1987; McDermott and McKay, 1994; Nelson *et al.*, 1995; Zwolan *et al.*, 1997).

However, considerable variation has been found in the relation between the pitch perceived and the place of stimulation. In some cases “reversals” have been noted (i.e., the pitch decreased as the active electrode was shifted basally), whereas in others, the size of the pitch difference between adjacent electrodes seemed to vary irregularly along the electrode array. Even in the same subjects, changing the stimulation mode among bipolar, monopolar, and common ground configurations also resulted in large differences in the pitch versus place relationships (Busby *et al.*, 1994; Cohen *et al.*, 1996; Collins *et al.*, 1997).

Most of the previous studies used pitch ranking, scaling, or absolute identification procedures. Thus, their results are difficult to interpret in terms of musical pitch. Although the subjects who participated in those studies may have described the sounds perceived with the electric stimuli as varying in pitch, it is possible that their responses were associated with one or more attributes of the percepts that might not have included melodic pitch. For example, a normally hearing (but musically untrained) person, presented with the same note played sequentially on two different instruments, might loosely describe one as having a higher “pitch” than the other. Such a description would presumably be based on differ-

ences in spectral shape, given that the fundamental frequencies of the sounds were identical. A musically trained listener would almost certainly recognize the two sounds as instances of the same musical note, but might describe the tonal quality of one as being “brighter” or “sharper” than the other. In other words, consistent ranking might be possible based on the timbre of sounds that have the same pitch.

On average, most users of cochlear implants would be expected to have somewhat less musical training and experience, because of the severity and duration of their hearing impairment, than people with normal hearing. Furthermore, only a small minority of implant recipients has received extensive formal musical training (Gfeller *et al.*, 2000b). Therefore, it is unlikely that the distinction between musical pitch and timbre would have been clear to each of the participants in the reported psychophysical experiments.

The studies involving changes exclusively to temporal parameters of electric stimuli demonstrated that melodic pitch information could be conveyed by this means, at least over a limited range. Could some musical pitch information also be available from stimuli in which only the site of delivery is changed? In the study of McDermott and McKay (1997) with one musically trained implant user, limited evidence was reported in support of this notion. Stimuli were presented in pairs, with each stimulus being delivered to a different electrode. The first electrode in each pair was selected randomly from among the positions available on the 22-electrode array, and the second electrode was constrained to be at one of the more basal (i.e., higher pitched) locations.

The subject was generally able to label these pairs of stimuli in terms of standard musical intervals. As the spatial separation between the pairs of electrodes increased, the subject systematically judged the intervals as wider, although the relationship between perceived interval size and electrode spacing appeared to be weaker for relatively wide separations.

Changing the rates on the electrodes seemed to have little effect on the subject’s judgments, suggesting that the changes in the place of stimulation were dominating the pitch percept. However, no experiments were carried out to assess whether melodies (without rhythmic or other coincident cues) could be recognized by converting pitch to place of stimulation while excluding temporal information.

#### 5.1.4. *Effects of Simultaneous Variation of Spatial and Temporal Parameters*

When complex acoustic signals, including speech, are converted into patterns of electric stimulation by the sound processors that are most commonly used with the present cochlear implants, both spatial and temporal properties of the output stimuli vary in accordance with relevant aspects of the input signal. As described previously, the sound processing schemes used with multielectrode implants convert each frequency component of a complex input signal into the selection of a corresponding electrode.

The division of the input signal into frequency subbands is performed by means of a bank of bandpass filters or a digital spectrum analysis technique. The subbands are assigned to the electrodes in an orderly way such that input signal components with high frequencies activate electrodes that are located at more basal positions than electrodes activated in response to lower input frequencies.

The stimuli delivered by the active electrodes comprise pulse trains that are generated at a constant, relatively high rate. This carrier rate is assumed to have a minimal effect on the pitch or other characteristics of the sensations perceived by the implant user. However, each pulse train is amplitude-modulated as a function of the level variations present in the envelope (or amplitude) of the corresponding frequency subband extracted from the input signal.

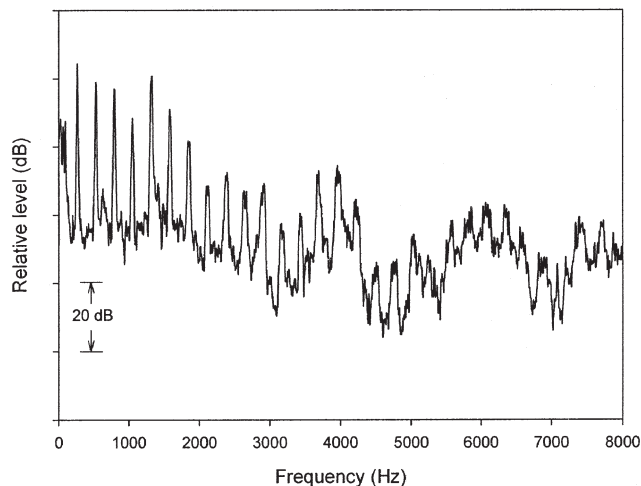
As reviewed briefly above, amplitude modulation of a constant-rate pulse train can elicit a pitch percept, provided that the modulation frequency is low enough (less than about 300 Hz), and the carrier pulse rate is high enough (at least four times the modulation frequency). Therefore, when the pitch of a complex acoustic signal at the input of a sound processor changes, the pattern of electric stimuli produced at the output of the implant is likely to vary in both temporal and spatial domains. The overall effect of these variations on the pitch, timbre, or other auditory characteristics perceived by the implant user is complicated and difficult to predict.

To address this problem further, it is helpful to consider a specific type of complex acoustic signal. One set of sounds that seems particularly suitable for this type of investigation comprises vowels sung at defined fundamental frequencies. Vowel sounds are, of course, repeatedly heard in speech. In Western languages, the voice pitch

(F0) and its variation convey some information about the identity of the speaker, including sex and age, and in certain situations, about the context of the utterance, such as whether it is a statement or a question.

Short-term changes in F0 can provide semantic information directly in a number of non-Western languages; for example, in tonal languages the same syllable uttered with several different F0 contours may have several corresponding meanings. Moreover, because listeners with acoustic hearing perceive changes in F0 as changes in pitch, musical melodies are very commonly produced and heard via the medium of singing. Thus, sung vowels have many important acoustic properties that are common to both music and speech.

Figure 9 shows the spectrum of the vowel /a/ sung by a woman at a pitch corresponding to the musical note middle-C. Because the fundamental frequency of this note is approximately 262 Hz, the spectrum contains numerous narrow peaks of energy located at multiples of this frequency.



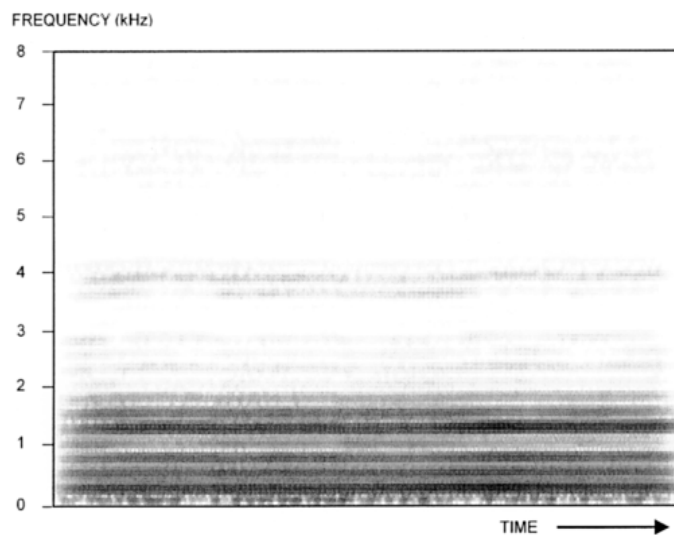
**Figure 9.** Spectrum of the vowel /a/ sung by a woman with a fundamental frequency (F0) of 262 Hz, which corresponds to the musical note middle-C. The abscissa is a linear frequency axis, whereas the ordinate shows the relative level in dB. The spectrum has numerous narrow peaks at frequencies equal to F0 and multiples (harmonics) of F0. The overall level variation superimposed on the peaks (i.e., the “spectral envelope”) is related partly to the same resonances in the vocal tract that create the formants of voiced speech.

These narrow spectral peaks are clearly visible in the spectrogram of Figure 10.

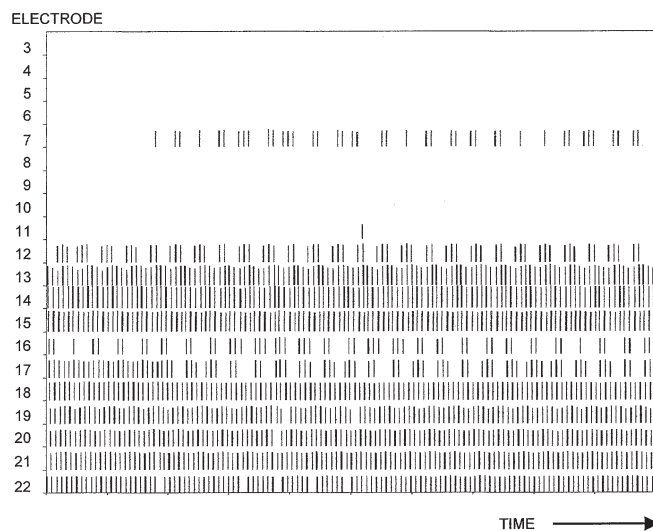
Also noticeable in both figures is an overall variation in the amplitude of the spectral peaks across frequency. This smoother variation, known as the “spectral envelope,” is characterized by a small number of relatively broad peaks that are representative of the formants. The formants are created by resonances in the vocal tract, and vowels in speech can generally be identified if the frequencies of the lowest two formants are perceived reasonably accurately.

The output of a cochlear implant in response to this sung vowel when processed by the ACE scheme is shown in Figure 11. The representation of the output stimuli in that figure is comparable in some ways with the spectrogram of Figure 10. The vertical axis shows the 20 active electrodes, ordered from apical (bottom) to basal (top). This ordering is the same as the frequency axis of the spectrogram, although the ACE processing scheme assigns frequencies to electrodes according to a nonlinear relationship.

The horizontal axis of each figure represents time. Each electric pulse at the output of the im-



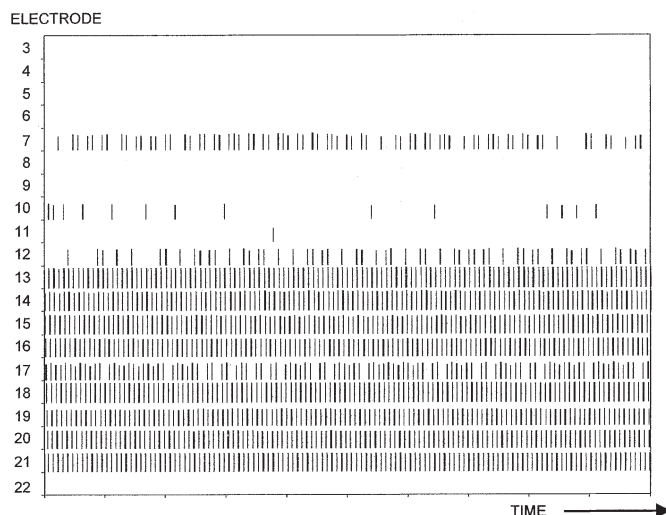
**Figure 10.** Spectrogram of the steady vowel whose spectrum is shown in Figure 9. The abscissa shows time (total duration of 500 ms), whereas the ordinate is a linear frequency axis. The relative level of the signal at each frequency is represented by the darkness of the plot. The dark horizontal bands correspond to the narrow peaks shown in Figure 9 and have frequencies equal to  $F_0$  and its multiples.



**Figure 11.** Output of the ACE sound-processing scheme for an input signal consisting of the vowel /a/ sung by a woman with a fundamental frequency of 262 Hz, as in Figures 9 and 10. The abscissa shows time (total duration of 100 ms), whereas the ordinate shows each of the 20 electrodes activated by the ACE processor. The ACE scheme was programmed to select 10 spectral maxima in each stimulation period. Apical electrodes (activated by low-frequency signals) are at the bottom of the axis, while basal electrodes are at the top. The stimulation delivered by each electrode is shown as a series of short vertical bars. Each bar represents one current pulse, with the height of the bar indicating the relative current level; the range is delimited by the “T-level” and “C-level” for each electrode.

plant is shown as a short vertical line in Figure 11. The height of each line represents the current amplitude of the pulse, delimited by the “T-level” and “C-level” for that electrode. Amplitude modulations in the stimuli are visible which correspond to the  $F_0$  of the input signal. For example, modulations with a period of about 3.8 ms (inverse of 262 Hz) may be seen clearly in the pulse train delivered by electrode 13. In the sound processor, this electrode was assigned to an input frequency band centered on 1.6 kHz.

The same format is used in Figure 12 to show the effect of changing the fundamental frequency of the sung vowel. In the audio recording used to create this figure, the same vowel was sung by the same woman with the  $F_0$  increased to 370 Hz (half of an octave).



**Figure 12.** Output of the ACE sound-processing scheme, as in Figure 11, for an input signal consisting of the vowel /a/ sung by a woman with a fundamental frequency of 370 Hz.

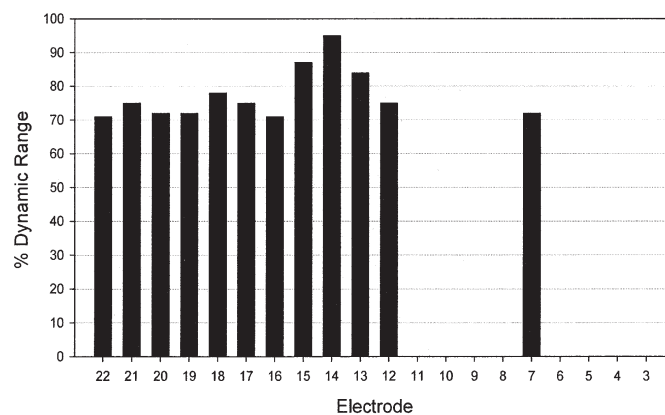
The amplitude modulations in the pulse trains on each electrode are generally shallower than for the lower F0. This is a consequence of the filter design that is used in the ACE processor, which progressively attenuates modulations above about 200 to 300 Hz. As noted previously, for modulation frequencies much higher than this, changes in modulation frequency do not usually result in large perceptual changes for implant users. Nevertheless, modulations with a period of approximately 2.7 ms (inverse of 370 Hz) are visible in the pulse train delivered by electrode 17, which was assigned to a frequency band centered on 920 Hz.

Provided that these amplitude modulations are perceptible to an implant user, and that the modulation frequencies do not exceed the limit for temporal pitch perception of that listener, it should be possible for these two vowels to be ranked correctly in pitch. However, coincident cues are also present in the spatial patterns of stimulation. By averaging the current levels of the stimuli delivered to each electrode over time, a graph showing the spatial pattern can be created.

Graphs for the above two sung vowels are shown in Figure 13 (F0 of 262 Hz) and Figure 14 (F0 of 370 Hz). In these figures, the abscissa rep-

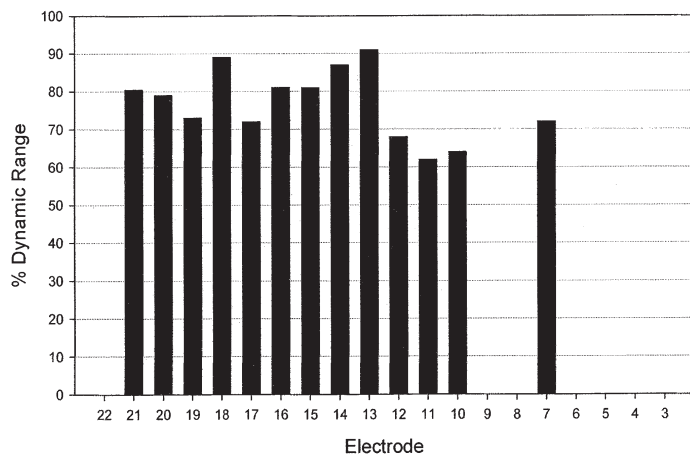
resents the electrode position, with apical electrodes towards the left and basal electrodes towards the right. The ordinate shows the average stimulation level. By comparing the graph for the higher F0 with that for the lower F0, a general shift of the stimulation pattern in the direction of the more basal electrodes can be seen. For example, electrode 22 is active for the lower F0, but not for the higher F0; whereas, electrode 10, which is inactive for the lower F0, becomes active at the higher F0. The implant user would be expected to perceive this shift in the stimulation pattern towards the cochlear base as a pitch (or timbre) increase. Thus, for these two signals, the concurrent spatial and temporal cues providing information about the change in F0 appear to be consistent, at least in the direction of the change.

Experimental results have been obtained with implant users listening to sung vowels. In one recent study (Looi *et al.*, 2004), recordings were used of two vowels (/a/ and /i/) that had been sung by a male and a female singer, both of whom had received formal musical training. For the experiment, a set of test signals was created using the male singer's vowels with F0 values of 98.0, 139, 196, and 277 Hz. Similarly, a set of



**Figure 13.** The distribution of average current levels across electrodes when the vowel /a/ with a fundamental frequency 262 Hz is processed by the ACE scheme. The data plotted were derived from the graph of Figure 11. The ordinate shows the average level as a percentage of the electrical dynamic range on each electrode. The abscissa shows the 20 electrode positions activated by the ACE processor, ranging from apical on the left to basal on the right.





**Figure 14.** The distribution of average current levels across electrodes, as in Figure 13, when the vowel /a/ with fundamental frequency 370 Hz is processed by the ACE scheme. The data plotted were derived from the graph of Figure 12.

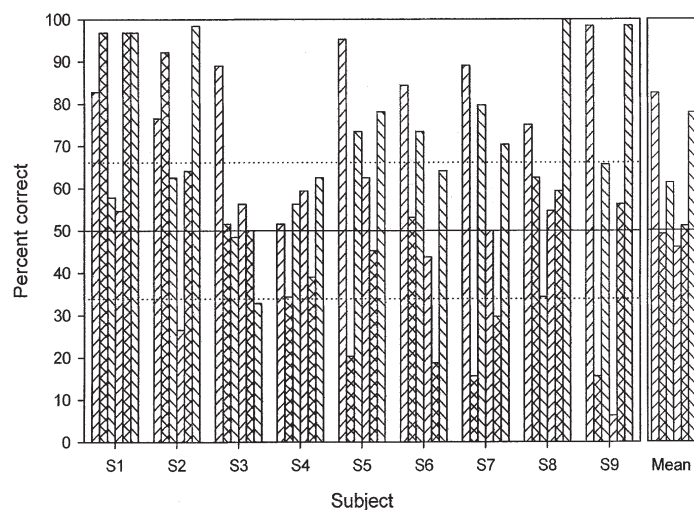
stimuli recorded by the female singer was selected that had F0 values of 262, 370, 523, and 740 Hz. For each singer and each vowel, the signals were presented in sequential pairs that had F0 values separated by half of one octave.

The subjects were 15 users of Nucleus multiple-channel cochlear implants, with sound processors programmed with either the ACE or the SPEAK strategies. The procedure required subjects to identify which stimulus in each pair had the higher pitch. Averaged across subjects and signals, the score for this test was only 62% correct. Although this was significantly higher than the chance score of 50%, it was much lower than the scores from normally hearing listeners who participated in the same study. Not surprisingly, the average score of the latter subjects was close to 100% correct, confirming that most listeners with normal hearing easily perceive the pitch change encompassed by this relatively large musical interval.

Of interest is a moderate correlation that was found between the implant users' performance in the pitch-ranking experiment and their ability to identify familiar melodies played from a small closed set. This finding was interpreted as implying that melody recognition was assisted by pitch discrimination, even though there were coincident cues, such as distinctive rhythms, in the

small number of familiar tunes presented to the subjects in the experiments.

Poor performance in pitch ranking of sung vowels was also found in another study that involved nine users of the SPEAK sound-processing scheme. The same set of test signals was used. The results from this experiment are presented in Figure 15. In that graph, the scores are shown separately for each of the three fundamental frequencies presented using vowels sung by each of



**Figure 15.** Results of an experiment in which nine users of the SPEAK sound-processing scheme were asked to rank the pitch of pairs of sequential sounds having various fundamental frequencies. The sounds were the sung vowels /a/ and /i/. The fundamental frequency ratio between the two sounds in each pair was one-half of one octave. The ordinate shows percentage of correct scores for each of the subjects (abscissa), with the mean across all subjects plotted at the far right. The experimental procedure was a two-alternative forced-choice task, with a score of 50% expected for random responses; this is shown as the solid horizontal line. The two dotted horizontal lines delimit the values required for the scores to be significantly different from chance ( $p < .01$ ). Note that scores below the lower dotted line indicate pitch reversals. For each subject, and for the mean across subjects, the data are plotted as clusters of six columns. In each cluster, the three columns on the left represent data for the male singer; the pairs of fundamental frequencies were 98–139, 139–196, and 196–277 Hz, respectively. The three remaining columns represent data for the female singer, with fundamental frequency pairs of 262–370, 370–523, and 523–740 Hz, respectively.

the two singers. Although the F0 was varied in the experiment, the musical interval separating each of the pairs of sounds that the subjects were asked to rank remained constant at half an octave. The data are plotted for each subject individually, with the group mean scores shown on the right. For each subject, the six clustered columns represent the scores for each of the six intervals, with F0 increasing from left to right. The first three columns represent data obtained using the male singer's vowels, whereas the remaining columns are for the female singer. Each column represents the score averaged across a total of 64 repetitions of the stimuli (32 presentations for each of the two vowels). The mean score for all subjects and stimuli was 61.6%, in close agreement with the result reported by Looi *et al.* (2004).

The individual data plotted in Figure 15 show very wide variation. The solid horizontal line shows the score (50%) expected if the subjects' responses were random. Higher scores are associated with intervals that were ranked in the correct direction; that is, the signal with the higher F0 was judged as higher in pitch, whereas scores below 50% indicate reversals in pitch perception. The two dotted horizontal lines delimit the range of scores that would be expected with random responses. That is, scores represented by columns that lie between these two lines are not significantly different from chance at the 1% level of significance.

It is evident that none of the nine subjects ranked all the pairs of stimuli in the correct direction with scores significantly better than chance. All of the subjects apparently gave random responses for some of the stimulus pairs. Of even greater concern is the finding that many instances of reversed rankings occurred that were statistically significant. For example, three of the nine subjects (S5, S7, and S9) gave responses that were statistically significant but reversed for the vowel pair sung by the male singer with F0 values of 139 and 196 Hz. When the data were averaged across all subjects, the scores for this vowel pair and the two vowel pairs with the lowest fundamental frequencies sung by the female singer were not significantly different from chance. The best ranking scores were obtained, on average, for the vowel pairs with the lowest (male singer, 98.0–139 Hz) and the highest (female singer, 523–740 Hz) values of F0.

The finding that perception of the pitch of complex sounds is generally poor for implant users

and varies greatly among listeners can be explained in terms of three broad aspects. First, as noted earlier, the ability of an implant user to extract pitch information from temporal patterns in electric stimuli is highly variable. In some cases, implant users can discriminate simple temporal differences, such as a small change in the rate of a steady pulse train, over a much wider range of rates than other implant users. The reasons for the observed large individual differences in the upper rate limit of temporal pitch discrimination are unclear, but probably include factors such as the number and condition of auditory neurons susceptible to stimulation by the implanted electrodes.

In the study described above, the highest average score obtained by the subjects was for the vowel pair having the two lowest fundamental frequencies. This is probably because the higher of these two values of F0 (139 Hz) was below the temporal pitch limit for all the implant users who participated in the experiment. Presumably, temporal pitch cues would have been available in the amplitude modulations of the electric stimuli produced in response to these acoustic signals.

However, the ability of implant users to extract pitch information from the frequency of amplitude modulations superimposed on constant-rate pulse trains may also depend on whether those modulations are aligned consistently across electrode positions for each input signal. For the sung vowels, the amplitude modulations representing F0 are present in the signal across a wide range of frequencies. The electric stimulation pattern shown in Figure 13 for the vowel sung with a fundamental frequency of 262 Hz contains F0-related modulations on most, if not all of the electrodes that are most active. The stimulation delivered by those electrodes was determined by the levels in the bandpass filters of the sound processor that covered a frequency range of approximately 120–2100 Hz.

A close inspection reveals that although the period of the modulations is the same on different electrodes, the phase of the modulations is not always aligned precisely in time across electrodes. For example, the modulation peaks on electrode 19 do not occur at exactly the same time as the peaks on electrode 17. If implant users were able to extract pitch information from the modulations on each active electrode independently, then phase misalignments such as these would have little perceptual effect. The temporal pitch information available from each electrode's stimula-

tion pattern would be identical; all the amplitude modulations have a frequency equal to  $F_0$ .

On the other hand, if the pitch percept is determined by a combination of stimuli integrated spatially across electrode positions, then phase misalignments could reduce or eliminate this type of temporal pitch information. For instance, if the amplitude modulations on electrodes 19 and 17 from the above example were, in effect, summed at each time instant, then the combined modulation pattern would be much shallower than the modulation present on either electrode by itself. The phase shift between the modulation patterns on these two electrodes would result in the peaks from one electrode almost canceling the valleys from the other electrode.

Unfortunately, evidence from psychophysical studies suggests that in certain conditions, implant users do combine temporal patterns across electrode positions such that phase shifts among amplitude-modulated stimuli have a substantial perceptual effect. In one study (McKay and McDermott, 1996), two stimuli, each a pulse train with a carrier frequency of 500 Hz modulated at 100 Hz, were delivered to two electrode positions. Both the phase shift between the modulation waveforms and the positions of the two active electrodes were varied.

The responses of the subjects who participated in the experiments indicated that changes in the phase shift were detectable when the electrodes were relatively closely spaced. The electrode separation had to exceed a distance that varied among the subjects before the perceptual effect of the phase shift was negligible. Beyond this distance, which ranged from 2.25–7.0 mm among the subjects, the temporal patterns on the two electrodes were perceived independently. This finding was supported by estimates of the pitch perceived by the subjects when the phase shift was constant but the electrode separation was varied. The perceived pitch was associated with the combined temporal pattern when the electrodes were close together, but was determined by the individual temporal pattern when the electrodes were relatively far apart.

These results are consistent with the assumption that neighboring electrodes stimulate partially overlapping populations of auditory neurons. One important implication is that when complex signals are heard through sound processors, the pitch information available from amplitude modulations in the pulse trains produced by

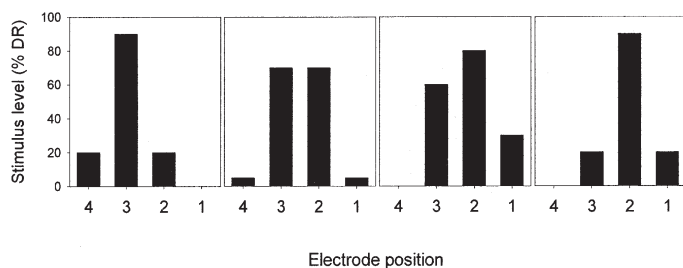
the active electrodes may be affected by the relative phase of the modulations. In particular, such temporal pitch information might be reduced if large phase shifts exist between the modulations present in stimuli delivered by electrodes that are closely spaced. These conditions may occur frequently in practice. It is common for phase shifts in complex signals to be present in amplitude modulations across a wide range of frequencies.

In many listening situations, phase shifts can result from the natural modification of the signal as it propagates from the source to the microphone of the implant system, encountering various resonances and reflections on its path. For sung vowels, filtering associated with the resonances in the vocal tract that create the formants might also result in phase shifts across frequency. Some potential improvements to the design of cochlear implant sound processors that might alleviate problems such as these are discussed briefly later.

A second possible explanation for the generally poor ability of the implant user to perceive musical pitch may be found in the way spatial patterns of stimulation are related to the spectral content of complex acoustic signals. As reviewed previously, a number of psychophysical studies have shown that shifting the place of stimulation from an apical to a more basal electrode position usually results in a sensation of increasing pitch or “sharpness.”

Such studies have nearly always involved activating only one electrode at a time using a pulse train with constant rate and amplitude. However, when a sound processor receives a complex acoustic signal, it typically produces activity on many electrodes concurrently, as illustrated in Figures 11 to 14. It is important to note that this is generally true even when the input signal is a pure tone.

Figure 16 shows a simulated distribution of average stimulus levels across four adjacent electrodes when a pure tone is processed by a generic multiple-channel sound processor (e.g., ACE, CIS, or SPEAK). The panel at the left shows the spatial distribution of stimulation when the tone’s frequency equals the center frequency of the bandpass filter assigned to electrode 3. Because the bandpass filters have partially overlapping frequency responses, some activity also occurs on the two neighboring electrodes (i.e., electrodes 2 and 4). In the adjacent panel to the right, the output level distribution is shown for a slightly in-



**Figure 16.** Illustration of how the stimulus level distribution across nearby electrodes varies when the frequency of a pure tone input to a sound processor is increased. A typical multiple-channel sound-processing scheme such as ACE, CIS, or SPEAK is assumed. In each panel, the abscissa shows electrode positions, with apical electrodes to the left and basal electrodes to the right. Average stimulus levels are shown as a percentage of electrical dynamic range on the ordinate. In the leftmost panel, the input tone has a frequency equal to the center frequency of the bandpass filter assigned to electrode 3. The three panels to the right show the effect of increasing the tone's frequency. The rightmost panel shows the stimulus level distribution when the frequency of the tone is equal to the center frequency of the filter assigned to electrode 2.

creased tone frequency. This frequency coincides with the crossover point between the filters assigned to electrodes 2 and 3, so the stimulation levels on these electrodes are equal. A further increase in tone frequency results in a shift of the stimulation pattern to more basal electrodes, until the tone is centered in the filter assigned to electrode 2 (rightmost panel).

Evidence from a psychophysical experiment shows that the percepts associated with stimulation patterns like those illustrated in Figure 16 can be ranked in an orderly sequence by implant users (McDermott and McKay, 1994). In effect, the pitch perceived seems to be related to the spatial centroid of the distributed stimulation pattern; as the centroid moves to more basal positions, the pitch is described as increasing.

Unfortunately, however, this perceptual relationship is probably less reliable than the relationship between perceived pitch and temporal patterns of stimulation at a fixed location. The place-related pitch depends on many factors, including the placement, geometry, and configuration of the electrodes, the intracochlear current paths arising from activation of the selected elec-

trodes, and the number, density, and location relative to the electrodes of the auditory neurons.

The third set of effects that warrants consideration when attempting to explain poor pitch perception is more subjective than physical. In particular, the amount of musical knowledge, training, and experience of the listener can affect the results of pitch-ranking experiments. This has been shown to apply to subjects with normal hearing, and it would be expected to be at least as important for cochlear implant users.

For example, a series of experiments described by Pitt (1994) showed that musicians and nonmusicians (with normal hearing) differed in their judgments of stimuli that varied in pitch and/or timbre. It was found that trained musicians rarely confused changes in pitch with changes in timbre in any of the experimental conditions. In contrast, nonmusicians frequently reported that only timbre had changed when, in fact, both pitch and timbre had changed simultaneously, or that both pitch and timbre had changed when only the timbre of the stimulus had changed.

Furthermore, the nonmusician subjects reported no perceptual change in the signal more often than the musicians when the pitch had changed while the timbre had not. It is reasonable to assume that similar confusions might have been made at least occasionally by the implant users who participated in the pitch perception studies discussed previously, since few of them would have had any musical training. This might partly explain the relatively low scores they obtained in the pitch-ranking experiments, particularly given the inherent uncertainty about whether pitch, timbre, or some combination of pitch and timbre was perceived as varying when the temporal and spatial parameters of the electric stimuli were manipulated.

The sung vowels described previously provide an interesting example of this type of uncertainty. As already discussed, temporal and spatial cues are both present in the stimulation patterns when these signals are processed by a typical cochlear implant sound-processing scheme, and these cues may not always provide consistent information. Studies have also demonstrated that conflicting cues may exist in the original acoustic signals; in particular, the frequencies of vowel formants can be affected by the fundamental frequency (F0) of the speaker (Loizou *et al.*, 1998; Maurer and Landis, 1995).

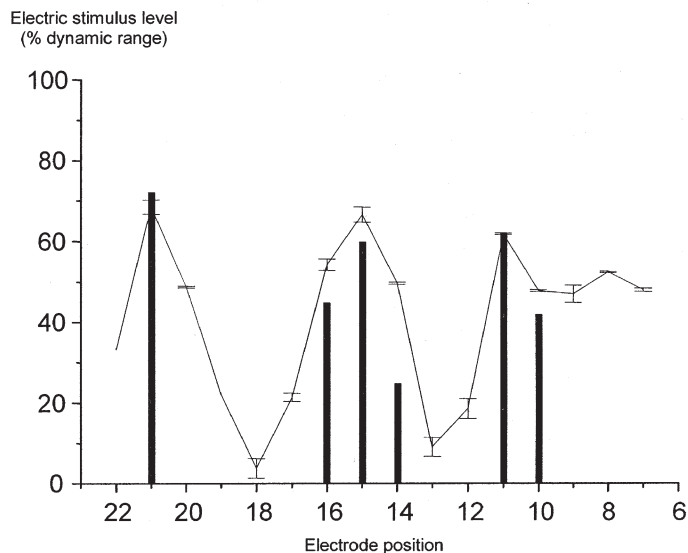
Increases in F0 may, under certain conditions, be accompanied by decreases in some formant frequencies. This has important implications for the studies in which implant users ranked the pitch of sung vowels. When the F0 of the sung vowels was increased, it is possible that some subjects detected this change via temporal stimulation patterns, whereas other subjects may have perceived the change in the spatial distribution of the electric stimuli as more salient. While the former group of implant users would have ranked the pitch change correctly, subjects in the latter group could sometimes have perceived a pitch change in the reverse direction, if the centroid of the spatial stimulation pattern moved towards a more apical location in accordance with an inconsistent change in formant frequencies.

## 5.2. Perception of Timbre

As discussed earlier, timbre can be used to describe various aspects of sounds, including dynamic characteristics; however, the perception by cochlear implant users of spectral shape is particularly significant. Discrimination among spectral shapes, even in the absence of other variations, enables implant users to identify vowels in speech and many other types of sounds.

Spectral shapes representative of a number of musical instrument sounds were studied in one published experiment (McKay, 2004). Stimuli were constructed to simulate the output of the SPEAK sound-processing scheme when brief, steady portions of each of these sounds were processed. An example is shown in Figure 17 for a stimulation pattern derived from the sound of the violoncello. The simulated SPEAK scheme activated six electrodes to represent the frequency bands that had the highest levels when this sound was processed (vertical columns).

A forward-masking experiment was conducted to investigate in detail how implant users perceived this relatively sparse representation of the steady spectral shape. The stimulus had a 200-ms duration and was presented at a comfortable loudness. After a silent interval of 4 ms, a probe stimulus with a 20-ms duration was presented on a single electrode. The probe electrode position was varied systematically, and its threshold of audibility was determined. The forward-masked thresholds were compared with the thresholds of the probe presented alone.



**Figure 17.** Results of a forward-masking experiment investigating the perception of spectral shape by a multiple-channel cochlear implant user (McKay, 2004). The masking stimulus was chosen to simulate the output of the SPEAK sound-processing scheme for an input signal derived from the sound of the violoncello. In the experimental version of SPEAK, six electrodes were activated representing the frequency bands containing the highest levels; the corresponding stimulation levels are displayed as the vertical columns at electrode positions shown along the abscissa. The forward-masked thresholds, plotted as a proportion of the electrical dynamic range of each electrode on the ordinate, are shown as points connected by a line. The error bars represent plus and minus one standard error of the mean.

The elevation of thresholds caused by the masking stimulation pattern are plotted as a function of electrode position for one subject in Figure 17 (solid line). It can be seen that the spatial variation of masked thresholds generally approximates the spatial pattern of the masker stimulus; however, thresholds were elevated to some extent on electrodes that were not activated by the masker (e.g., electrodes 17 and 20). Furthermore, there were several electrodes at positions more basal than any of the activated masker electrodes for which thresholds were elevated (electrodes 7, 8, and 9).

These results suggest that relatively fine details of the spectral shape of an acoustic signal are probably smeared perceptually for implant listeners. The relatively crude spatial resolution implied

by these findings may be the result of spread of stimulating currents around the active electrodes, poor neural survival, or other factors. In any case, the combination of the rather coarse spectral analysis applied by most existing implant sound processors and the perceptual smearing revealed by psychophysical studies such as the one just outlined provides at least a partial explanation for the generally unsatisfactory performance attained by implant users when they identify complex sounds, including those of musical instruments.

## 6. Enhancing Music Perception by Improving Sound Processor Design

Many of the studies reviewed in this paper have found that perception of pitch and timbre (particularly in terms of spectral shape) are not adequate to enable most implant users to appreciate music fully. Although it appears that discrimination of rhythm is often satisfactory, it is likely that loudness perception in general could be improved. For example, the input dynamic range of existing sound processors is usually much narrower than the range of levels present in speech (Zeng *et al.*, 2002), and the overall range of levels in music is presumably even wider.

The conversion of acoustic input levels into levels of electric stimulation most likely results in loudness perception for complex sounds that is very different for implant users than in normally hearing listeners. A new sound-processing scheme has been developed recently that attempts to reduce this difference (McDermott *et al.*, 2003). The processor incorporates a model of loudness perception for normal acoustic hearing (Moore and Glasberg, 1997) and a loudness model for electric stimulation (McKay *et al.*, 2003).

The stimuli produced by the processor are determined in real time by these models, so that the distribution of loudness across frequencies in the acoustic signal is represented by a corresponding distribution of loudness across electrode positions in the cochlear implant. In addition, the total loudness perceived by the implant user closely approximates the total loudness that would be perceived by a normally hearing listener for the same acoustic signal. A perceptual experiment confirmed that this processing scheme did normalize the loudness of a number of complex acoustic signals that differed in level and bandwidth

(McDermott *et al.*, 2003). It is plausible that a loudness normalization technique such as this might enhance implant users' experience of listening to music.

Sound-processing algorithms could potentially be modified in a number of ways to provide implant recipients with more information about the pitch of sounds. One practical possibility is to improve the design of the filter-banks typically used to analyze the short-term spectrum of acoustic signals. One experimental design modification described recently (Geurts and Wouters, 2004) exploits the place-pitch cues discussed previously by providing finer spatial resolution in the representation of the fundamental frequency of complex sounds. On the other hand, temporal resolution could also be enhanced.

A comparatively simple modification of existing processing schemes would reduce or eliminate the phase shifts across electrode positions that seem to degrade implant users' ability to extract consistent pitch cues from the amplitude modulations related to the fundamental frequency. Furthermore, the depth of these amplitude modulations could be artificially expanded to enhance the salience of this type of pitch cue.

A more radical modification of existing processing schemes would apply the fundamental frequency directly to control the rate of stimulation. Although this technique was used in the MPEAK strategy and several of its predecessors, it became obsolete as a consequence of the improvements in speech understanding that were gained when constant-rate schemes such as ACE, CIS, and SPEAK were introduced. Nevertheless, psychophysical studies have shown convincingly that controlling the rate or frequency of electric stimulation produces a change in the perceived pitch, and the relationship between pitch and rate is similar to that between pitch and tone frequency in normal acoustic hearing. Therefore, the rate of stimulation seems to be a particularly suitable parameter to control in order to produce pitches that represent musical notes appropriately.

In practice, a fundamental frequency estimator is required that functions reliably in real time, and in realistic listening conditions (e.g., with noise and reverberation present). The estimated F0 could be used to modulate simultaneously the stimuli presented on all active electrodes. The preliminary results of ongoing experiments in our laboratory evaluating a modified ACE-like scheme that presents F0 in this manner suggest that it

might provide better musical pitch information to implant users, at least over a restricted range of fundamental frequencies.

To overcome the inherent limitations of temporal pitch encoding in cochlear implants, techniques may be needed to provide perceptual information specifically about the “fine structure” of acoustic signals. The fine structure contains rapidly varying components of sounds that are not present in the envelope levels that are relied upon exclusively in most existing implant sound-processing schemes. In recent experiments with normally hearing subjects, the fine structure has been shown to be far more important for pitch perception than envelope cues (Smith *et al.*, 2002).

How information about fine structure can be conveyed most successfully to implant users in practice is not yet clear. It is possible that novel processing schemes that attempt to simulate some of the properties of the mechanical traveling wave that propagates along the basilar membrane of the normal cochlea might be effective. Improvements in the design of electrode arrays to increase the number of discrete sites of stimulation and to place electrodes closer to auditory neurons might also be beneficial, especially if the amount of spatial overlap resulting from stimulation on adjacent or nearby electrodes can be reduced.

Two additional avenues that are worth investigating further in implant design are the use of continuous analog stimulation waveforms and the delivery of electric stimuli to multiple neural sites simultaneously. Techniques such as these may facilitate the development of innovative spatio-temporal signal coding schemes that convey better pitch information to implant users.

Finally, a new potential for major advances is arising currently from the use of combined acoustic and electric stimulation in suitable implant recipients. The perceptual improvements that have accompanied developments in cochlear implant systems over at least the past 20 years have encouraged an increasing number of people with usable acoustic hearing to receive an implant. These people usually have some hearing sensitivity at relatively low frequencies, but little or no hearing at higher frequencies. In some research centers, they may now be implanted with a device that has a shorter electrode array than the array used in conventional implants, or they may receive a conventional device in which the

electrode is not fully inserted during surgery (Kiefer *et al.*, 1998; von Ilberg *et al.*, 1999; Skarzynski *et al.*, 2003).

In a number of cases, hearing threshold levels are affected only slightly as a consequence of the implantation procedure. Thus, these people can continue to hear low-frequency components of sounds postoperatively, usually with the help of acoustic hearing aids, but also obtain information about high-frequency signals via the electric stimulation. Not surprisingly, their low-frequency hearing assists greatly with pitch perception (Turner *et al.*, 2004) and would be expected to enhance their enjoyment and appreciation of music in general. As cochlear implants become more widely accepted as a safe and effective treatment for many people who have partial, rather than profound or total hearing impairment, it seems certain that the size of this new population of implant recipients will grow steadily. Such people will most probably experience much better perception of music than the typical implant user of today.

### Acknowledgments

Many colleagues have contributed substantially to this work. Thanks are due, in particular, to Bob Carlyon, WaiKong Lai, Valerie Looi, David MacFarlane, Colette McKay, Thomas Stainsby, Catherine Sucher, David Tsang, and Andrew Vandali. We also thank the subjects who volunteered to participate in the reported studies.

### References

- Abdi S, Khalessi MH, Khorsandi M, Gholami B. Introducing music as a means of habilitation for children with cochlear implants. *Int J Pediatr Otorhi* 59:105-113, 2001.
- ASA. American standards acoustic terminology. New York: Acoustical Society of America, 1960.
- Bilger RC. Psychoacoustic evaluation of present prostheses. *Ann Otol Rhinol Laryngol* 86 Suppl 38:92-140, 1977.
- Busby PA, Whitford LA, Blamey PJ, et al. Pitch perception for different modes of stimulation using the Cochlear multiple-electrode prosthesis. *J Acoust Soc Am* 95:2658-2669, 1994.

- Cohen LT, Busby PA, Clark GM. Cochlear implant place psychophysics. 2. Comparison of forward masking and pitch estimation data. *Audiol Neurootol* 1:278-292, 1996.
- Collins LM, Zwolan T, Wakefield GH. Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects. *J Acoust Soc Am* 101:440-455, 1997.
- Djourno A, Eyries C. Prosthèse auditive par excitation électrique à distance du nerf sensoriel à l'aide d'un bobinage inclus à demeure. *Presse Med* 65:1417, 1957.
- Dorman MF, Basham K, McCandless G, Dove H. Speech understanding and music appreciation with the Ineraid cochlear implant. *The Hearing Journal* 44:34-37, 1991.
- Dowell RC, Seligman PM, Blamey PJ, Clark GM. Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis. *Ann Otol Rhinol Laryngol* 96 Suppl 128:132-134, 1987.
- Eddington DK. Speech discrimination in deaf subjects with cochlear implants. *J Acoust Soc Am* 68:885-591, 1980.
- Eddington DK, Dobelle WH, Brackmann DE, et al. Auditory prosthesis research with multiple channel intracochlear stimulation in man. *Ann Otol Rhinol Laryngol* 87 Suppl 53:5-39, 1978.
- Fourcin AJ, Rosen SM, Moore BCJ, et al. External electrical stimulation of the cochlea: Clinical, psychophysical, speech perceptual, and histological findings. *Br J Audiol* 13:85-87, 1979.
- Fujita S, Ito J. Ability of Nucleus cochlear implantees to recognize music. *Ann Otol Rhinol Laryngol* 108:634-640, 1999.
- Geurts L, Wouters J. Better place-coding of the fundamental frequency in cochlear implants. *J Acoust Soc Am* 115:844-852, 2004.
- Gfeller K. Aural rehabilitation of music listening for adult cochlear implant recipients: Addressing learner characteristics. *Music Therapy Perspectives* 19:88-95, 2001.
- Gfeller K, Christ A, Knutson J, et al. The effects of familiarity and complexity on appraisal of complex songs by cochlear implant recipients and normal hearing adults. *J Music Ther* 40:78-112, 2003.
- Gfeller K, Christ A, Knutson JF, et al. Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *J Am Acad Audiol* 11:390-406, 2000b.
- Gfeller K, Lansing C. Musical perception of cochlear implant users as measured by the "Primary Measures of Music Audiation": An item analysis. *J Music Ther* 29:18-39, 1992.
- Gfeller K, Lansing CR. Melodic, rhythmic, and timbral perception of adult cochlear implant users. *J Speech Hear Res* 34:916-920, 1991.
- Gfeller K, Turner C, Mehr M, et al. Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants International* 3:29-53, 2002a.
- Gfeller K, Witt S, Adamek M, et al. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *J Am Acad Audiol* 13:132-145, 2002c.
- Gfeller K, Witt SA, Kyung-Hyun K, et al. Preliminary report of a computerized music training program for adult cochlear implant recipients. *J Acad Rehab Audiol* 32:11-27, 1999.
- Gfeller K, Witt S, Spencer LJ, et al. Musical involvement and enjoyment of children who use cochlear implants. *The Volta Review* 100:213-233, 1998.
- Gfeller K, Witt S, Stordahl J, et al. The effects of training on melody recognition and appraisal by adult cochlear implant recipients. *J Acad Rehab Audiol* 33:115-138, 2000.
- Gfeller K, Witt S, Woodworth G, et al. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann Otol Rhinol Laryngol* 111:349-356, 2002b.
- Gfeller K, Woodworth G, Robin DA, et al. Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear Hear* 18:252-260, 1997.
- Gordon EE. Primary measures of music audiation. Chicago: GIA Publications, 1979.
- Greenwood DD. A cochlear frequency-position function for several species—29 years later. *J Acoust Soc Am* 87:2592-2605, 1990.
- Grey JM. Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am* 61:1270-1277, 1977.
- Kessler DK. The CLARION multi-strategy cochlear implant. *Ann Otol Rhinol Laryngol* 108 Suppl 177:8-16, 1999.
- Kiefer J, von Ilberg C, Reimer B, et al. Results of cochlear implantation in patients with severe to profound hearing loss: Implications for patient selection. *Audiology* 37:382-395, 1998.
- Kong Y-Y, Cruz R, Jones JA, et al. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear* 25:173-185, 2004.
- Leal MC, Shin YJ, Laborde M-L, et al. Music perception in adult cochlear implant recipients. *Acta Otolaryngol* 123:826-835, 2003.
- Loizou PC. Mimicking the human ear: An overview of signal-processing strategies for converting sound into electrical signals in cochlear implants. *IEEE Signal Processing Magazine* September:101-130, 1998.
- Loizou PC, Dorman MF, Powell V. The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor. *J Acoust Soc Am* 103:1141-1149, 1998.



- Looi V, McDermott HJ, McKay CM, et al. Pitch discrimination and melody recognition by cochlear implant users. In: Proceedings of the VIII International Cochlear Implant Conference, Indianapolis, Indiana. Elsevier, (in press), 2004.
- Looi V, Sucher CM, McDermott HJ. Melodies familiar to the Australian population across a range of hearing abilities. *Austr NZ J Audiol* 25:75-83, 2003.
- Maurer D, Landis T. F0-dependence, number alteration, and non-systematic behaviour of the formants in German vowels. *Intern J Neuroscience* 83:25-44, 1995.
- McDermott HJ, Looi V. Perception of complex signals, including musical sounds, with cochlear implants. In: Proceedings of the VIII International Cochlear Implant Conference, Indianapolis, IN. Elsevier, (in press), 2004.
- McDermott HJ, McKay CM. Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. *J Acoust Soc Am* 96:155-162, 1994.
- McDermott HJ, McKay CM. Musical pitch perception with electrical stimulation of the cochlea. *J Acoust Soc Am* 101:1622-1631, 1997.
- McDermott HJ, McKay CM, Richardson LM, et al. Application of loudness models to sound processing for cochlear implants. *J Acoust Soc Am* 114:2190-2197, 2003.
- McDermott HJ, McKay CM, Vandali AE. A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *J Acoust Soc Am* 91:3367-3371, 1992.
- McKay CM. Psychophysics and electrical stimulation. In Zeng F-G, Popper A.N. and R. R. Fay, ed. *Springer Handbook of Auditory Research: Auditory prostheses*. New York: Springer-Verlag, 286-333, 2004.
- McKay CM, Henshall KR, Farrell RJ, et al. A practical method of predicting the loudness of complex electrical stimuli. *J Acoust Soc Am* 113:2054-2063, 2003.
- McKay CM, McDermott HJ. Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP). *Ear Hear* 14:350-367, 1993.
- McKay CM, McDermott HJ. The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites. *J Acoust Soc Am* 100:1081-1092, 1996.
- McKay CM, McDermott HJ. The perceptual effects of current pulse duration in electrical stimulation of the auditory nerve. *J Acoust Soc Am* 106:998-1009, 1999.
- McKay C, McDermott H, Carlyon R. Place and temporal cues in pitch perception: are they truly independent? *Acoustic Research Letters Online* 1:25-30, 2000.
- McKay CM, McDermott HJ, Clark GM. Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. *J Acoust Soc Am* 96:2664-2673, 1994.
- McKay CM, McDermott HJ, Clark GM. Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth. *J Acoust Soc Am* 97:1777-1785, 1995.
- Moore BC, Carlyon RP. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: C. J. Plack and A. J. Oxenham, ed. *Springer Handbook of Auditory Research: Pitch Perception*. New York: Springer-Verlag, (in press), 2005.
- Moore BG, Glasberg BR. A model of loudness perception applied to cochlear hearing loss. *Auditory Neuroscience* 3:289-311, 1997.
- Moore BC, Rosen SM. Tune recognition with reduced pitch and interval information. *Q J Exp Psychol* 31:229-240, 1979.
- Nelson DA, Van Tasell DJ, Schroder AC, et al. Electrode ranking of "place pitch" and speech recognition in electrical hearing. *J Acoust Soc Am* 98:1987-1999, 1995.
- Patrick JF, Clark GM. The Nucleus 22-channel cochlear implant system. *Ear Hear* 12 Suppl 1:3-9, 1991.
- Patrick JF, Seligman PM, Money DK, et al. Engineering. In: Clark G.M, Tong YC, and Patrick JF (eds). *Cochlear Prostheses*. Edinburgh: Churchill Livingstone, 99-124, 1990.
- Pfingst BE, Holloway LA, Poopat N, et al. Effects of stimulus level on nonspectral frequency discrimination by human subjects. *Hear Res* 78:197-209, 1994.
- Pijl S. Pulse rate matching by cochlear implant patients: Effects of loudness randomization and electrode position. *Ear Hear* 18:316-325, 1997a.
- Pijl S. Labeling of musical interval size by cochlear implant patients and normally hearing subjects. *Ear Hear* 18:364-372, 1997b.
- Pijl S, Schwarz DW. Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. *J Acoust Soc Am* 98:886-895, 1995.
- Pitt MA. Perception of pitch and timbre by musically trained and untrained listeners. *J Exp Psychol Human* 20:976-986, 1994.
- Pratt RL, Doak PE. A subjective rating scale for timbre. *Journal of Sound and Vibration* 45:317-328, 1976.
- Schulz E, Kerber M. Music perception with the MED-EL implants. In: Hochmair-Desoyer IJ and Hochmair ES (eds). *Advances in Cochlear Implants*. Vienna: Manz, 326-332, 1994.
- Shannon RV. Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. *J Acoust Soc Am* 85:2587-2592, 1989.
- Shannon RV. Psychophysics. In: R. S. Tyler, ed. *Cochlear implants: audiological foundations*. San Diego: Singular, 357-388, 1993.
- Simmons FB. Electrical stimulation of the auditory nerve in man. *Arch Otolaryngol* 84:2-54, 1966.

- Simmons FB, White RL, Walker MG, et al. Pitch correlates of direct auditory nerve electrical stimulation. *Ann Otol Rhinol Laryngol Suppl* 90:15-18, 1981.
- Skarzynski H, Lorens A, Piotrowska A. A new method of partial deafness treatment. *Medical Science Monitor* 9:CS26-30, 2003.
- Skinner MW, Clark GM, Whitford LA, et al. Evaluation of a new Spectral Peak coding strategy for the Nucleus 22-channel cochlear implant system. *Am J Otol* 15 Suppl 2:15-27, 1994.
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416:87-90, 2002.
- Szelag E, Kolodziejczyk I, Kanabus M, et al. Deficits of non-verbal auditory perception in postlingually deaf humans using cochlear implants. *Neurosci Lett* 355:49-52, 2004.
- Tong YC, Clark GM. Absolute identification of electric pulse rates and electrode positions by cochlear implant patients. *J Acoust Soc Am* 77:1881-1888, 1985.
- Townshend B, Cotter N, Van Compernelle D, White RL. Pitch perception by cochlear implant subjects. *J Acoust Soc Am* 82:106-115, 1987.
- Turner CW, Gantz BJ, Vidal C, et al. Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing. *J Acoust Soc Am* 115:1729-1735, 2004.
- van Hoesel RJ, Clark GM. Psychophysical studies with two binaural cochlear implant subjects. *J Acoust Soc Am* 102:495-507, 1997.
- Vandali AE, Whitford LA, Plant KL, et al. Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system. *Ear Hear* 21:608-624, 2000.
- von Ilberg C, Kiefer J, Tillein J, et al. Electric-acoustic stimulation of the auditory system. *ORL* 61:334-340, 1999.
- Wilson BS, Finley CC, Lawson DT, et al. Better speech recognition with cochlear implants. *Nature* 352:236-238, 1991.
- Zeng F-G. Temporal pitch in electric hearing. *Hear Res* 174:101-106, 2002.
- Zeng F-G, Grant G, Niparko J, et al. Speech dynamic range and its effect on cochlear implant performance. *J Acoust Soc Am* 111:377-386, 2002.
- Zwolan TA, Collins LM, Wakefield GH. Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects. *J Acoust Soc Am* 102:3673-3685, 1997.