

Effects of modulation wave shape on modulation frequency discrimination with electrical hearing

David M. Landsberger

Department of Auditory Implants and Perception, House Ear Institute, 2100 West 3rd Street, Los Angeles, California 90057,

Department of Otolaryngology, The University of Melbourne, 384-388 Albert St. East Melbourne 3002 Australia, and School of Life and Health Sciences,

*Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom
dlandsberger@hei.org*

Abstract: Amplitude modulations of pulsatile stimulation can be used to convey pitch information to cochlear implant users. One variable in designing cochlear implant speech processors is the choice of modulation waveform used to convey pitch information. Modulation frequency discrimination thresholds were measured for 100 Hz modulations with four waveforms (sine, sawtooth, a sharpened sawtooth, and square). Just-noticeable differences (JNDs) were similar for all but the square waveform, which often produced larger JNDs. The results suggest that a sine, sawtooth, and sharpened sawtooth waveforms are likely to provide similar pitch discrimination within a speech processing strategy.

© 2008 Acoustical Society of America

PACS numbers: 43.66.Ts, 43.66.Fe, 43.66.Hg, 43.66.Nm [QJF]

Date Received: February 1, 2008 **Date Accepted:** April 17, 2008

1. Introduction

Cochlear implantees using commercial speech processing strategies are often able to understand speech well enough to conduct conversations in quiet situations. These high levels of performance are obtained even though most clinical speech processing strategies do not provide clear fundamental frequency (F0) information. For both CIS (Continuously Interleaved Sampling; Wilson *et al.*, 1991) and ACE (Advanced Combination Encoder; Vandali *et al.*, 2000) processing strategies, F0 cues are primarily provided by amplitude modulations within a channel, which can modulate at F0 if the corresponding filter bandwidth is sufficiently broad (Shamma and Klevin, 2000). It has been shown (Shannon *et al.*, 1995; Faulkner *et al.*, 2000) that F0 is not required for recognition of vowels or consonants, which may partly explain the relatively high speech understanding in quiet despite the limited F0 coding in these strategies.

Implant patients have difficulty with tasks that are facilitated by F0 information. Patients have difficulty discriminating between questions and statements (Green *et al.*, 2005), identifying a speaker (Fu *et al.*, 2004), recognizing tones in tonal languages (Fu *et al.*, 2004), and understanding a speaker in the presence of a speech masker (Stickney *et al.*, 2004). To improve performance in F0 related tasks, novel speech processing strategies have been developed to specifically encode F0 in the modulations of the outputs of the electrodes. Geurts and Wouters (2001) created a speech processing strategy (F0 CIS) in which modulation depths were increased at F0. However, F0 discrimination for synthesized vowels was similar for standard CIS and F0 CIS. Laneau *et al.* (2006) tested a strategy (F0mod) that was a modification of ACE in which the envelopes of each channel were modulated by F0 at 100% modulation depth. Performance with F0mod was better than that with ACE for familiar melody recognition and musical note discrimination when F0 was below 250 Hz. However, F0mod was not tested on speech. Green *et al.* (2005) implemented a strategy where outputs of electrodes were modulated at F0 using a sharpened sawtooth (see illustration in Fig. 1). Compared to standard CIS, listeners were better able to discriminate between rising and falling pitch, and between questions and

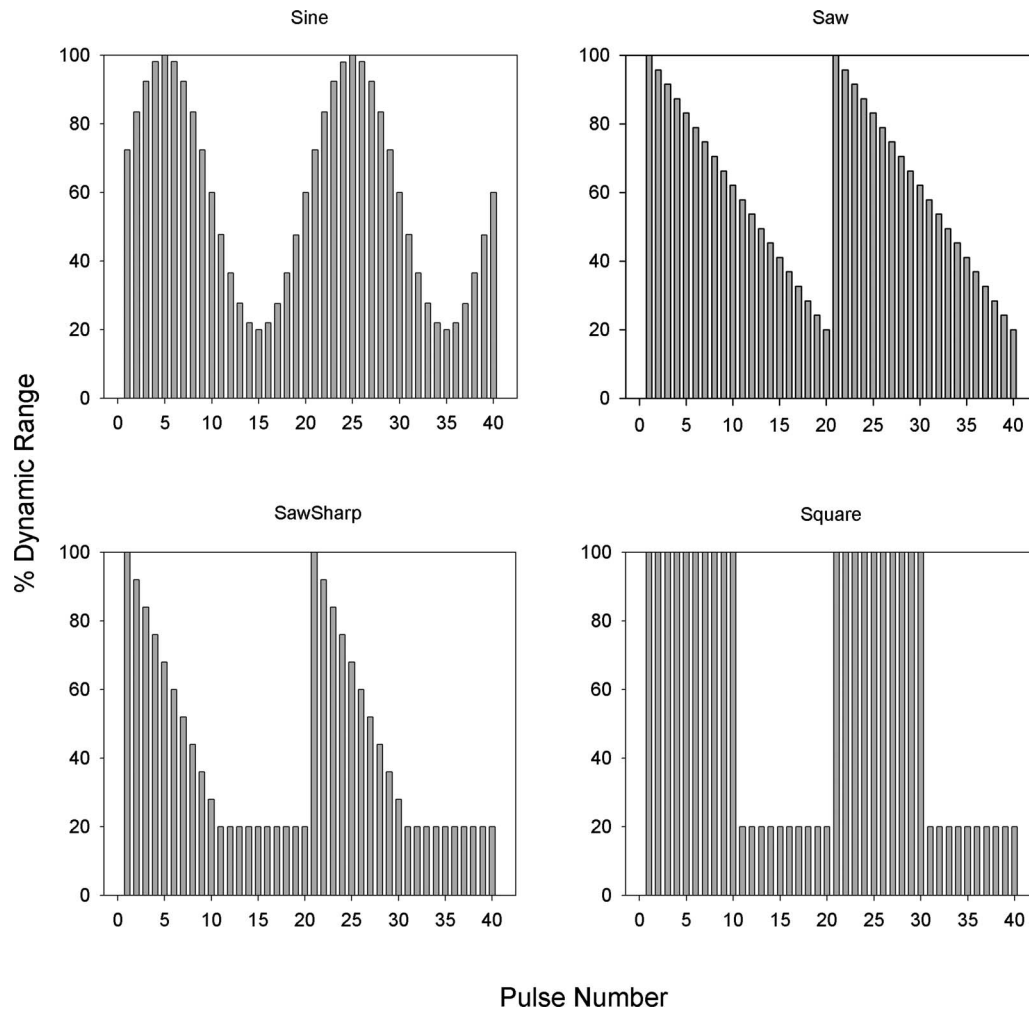


Fig. 1. Illustration of the four experimental modulation waveforms.

statements with the experimental strategy; however, vowel recognition was significantly poorer with the experimental strategy. These results are consistent with [Chatterjee and Peng \(2008\)](#) who showed that modulation frequency discrimination on a single electrode was strongly correlated with speech intonation. [Vandali et al. \(2005\)](#) tested several speech processing strategies (both experimental and commercial) and found that the experimental strategies that modulated outputs of the electrodes with F0 information enabled better pitch discrimination in a sung vowel task compared to performance with ACE. However, with the exception of Multi-Channel Envelope Modulation (MEM), all of the experimental strategies resulted in speech comprehension that was poorer than that with ACE. Only the MEM strategy, which modulated the outputs of all electrodes by the envelope of the input signal (and thereby had inherent modulations at F0) provided an improvement in the sung vowel pitch task without detriment to speech performance. Although commercial speech processing strategies do not provide optimal F0 information via channel modulations, the limited modulation pitch information is useful. For example, [Laneau et al. \(2004\)](#) showed that removing amplitude modulations above 10 Hz in a speech processing strategy significantly reduced F0 discrimination.

Each of these experimental strategies was developed to improve F0 pitch perception by modulating the current output from the electrodes. However, when designing a speech pro-

cessing strategy, there are many parameters to evaluate. Some of these parameters include which (or how many) electrodes to modulate, what is the appropriate modulation depth, and what waveform should be used for the modulations. Geurts and Wouters (2001) showed that presenting amplitude modulations on multiple electrodes improved listeners' ability to detect and discriminate modulations. Assumptions about the appropriate waveform have been made, but not adequately tested. Green *et al.* (2004) compared the ability of cochlear implant subjects to perform a pitch glide discrimination with either standard CIS, or modified CIS strategies which provided modulations on all electrodes at F0. The F0 modulations were presented either as sawtooth or sharpened sawtooth waveforms. While the modified strategies improved pitch glide discrimination relative to standard CIS, no differences in speech performance were detected between the sawtooth and sharpened sawtooth waveforms. Further experiments with cochlear implantees (Green *et al.*, 2004, 2005) used only the sawtooth waveform for modulating F0.

In the current study, just-noticeable differences (JNDs) for modulation frequency discrimination were measured for different modulation waveforms, namely: sine, sawtooth (saw), sharpened sawtooth (sawsharp), and square waveforms (see Fig. 1). Most previous studies regarding amplitude modulation JNDs with cochlear implants have used sinusoidal modulation (e.g., Geurts and Wouters, 2001; Chatterjee and Peng, 2008); thus, the data from the present study can be readily compared to that from previous studies. Naturally occurring sounds modulate sinusoidally; sinusoidal modulation therefore may be the optimal waveform for cochlear stimulation, and may produce the smallest modulation frequency JNDs. However, if modulation frequency pitch is determined by the interpulse interval in the neural firing pattern, then the sharp onset of the sawtooth and sharpened sawtooth may produce the smallest JNDs. Alternatively, if modulation frequency pitch is determined by neurons that detect the transitions between the onset and offset of a stimulus, the square waveform may produce the smallest JNDs.

2. Methods

2.1 Subjects

Eight subjects with the Nucleus CI24 implant participated in the present study. Subjects used either the SPEAK (Seligman and McDermott, 1995; Whitford *et al.*, 1995) or ACE strategy in their clinical speech processors. The data for five of the subjects was collected at the University of Melbourne in Australia, while the data for the remaining subjects were collected at Aston University in England. The Australian subjects had all participated in previous psychophysical experiments while the English subjects had no previous research experience.

2.2 Stimuli

All stimuli consisted of monopolar (MP1+2) biphasic pulse trains presented on a single electrode using a SPEAR speech processor (HearWorks Ptv. Ltd., 2003) that was controlled by custom-written software. Stimulation consisted of pulses presented to electrodes 20, 12, or 6, which are located in the apical, medial, and basal regions of the cochlea, respectively. Stimuli had a pulse phase duration of 26 μ s, an interphase gap of 8.4 μ s, and a duration of 500 ms.

The stimuli contained modulations that were either sinusoidal (sine), sawtooth (saw), a sharpened sawtooth (sawsharp), or square (square) waveforms. The modulation depth was fixed at 80% of the dynamic range (DR). The peak of the modulation corresponded to 100% DR and minimum of the modulation corresponded to 20% DR. The carrier rate of the modulated stimuli was fixed at 20 times the modulation rate. As a result, every modulation cycle was encoded with exactly 20 samples. This ensured that, for the saw and sawsharp waveforms, the sharp onset of the waveform was sampled accurately every cycle. In a pilot study, subjects were easily able to discriminate between 100 and 100.5 Hz modulated stimuli (saw waveform) when the carrier was fixed at 2000 pps, most likely because of aliasing artifacts present in the 100.5 Hz stimulus. Thus, in the present study, 100 Hz modulation was coded with a 2000 pps carrier and 100.5 Hz modulation was coded with a 2010 pps carrier.

2.3 Procedure

On each experimental electrode, the DRs were estimated for the 2000 pps (unmodulated) carrier. To measure the loudest comfortable level (*C* level), subjects listened to the stimulus at increasing amplitudes until they reported that the sound was at the loudest comfortable level. Threshold levels (*T* levels) were measured using a one-up, one-down 4 interval forced choice (IFC) task. This procedure was repeated until ten reversals were obtained. The last six reversals were averaged to estimate the 50% detection threshold (Levitt, 1971). The DR was defined as the difference between *C* and *T* levels, in Nucleus current levels.

The loudness of two stimuli were balanced using a one-up, one-down, 2IFC task. One of the stimuli had a fixed amplitude and the other stimulus was adjusted according to subject response by 1 Nucleus current step (0.18 dB). Ten reversals were measured, and the last six reversals were averaged to estimate the level of equal loudness. The loudness balancing procedure was then repeated, with the previous reference as the probe and the previous probe as the reference. The loudness balancing procedure was repeated twice and all of the loudness estimates were averaged. The loudness balance procedure was the same used for subjects MM, BK, and DC in a previous study (Landsberger and McKay, 2005). For each electrode, 100 Hz modulated stimuli with a 2000 pps carrier were loudness balanced to 100 pps unmodulated stimuli at *C* level.

For each of the four waveforms, a 4IFC adaptive procedure (one-up, one-down) was used to measure the modulation frequency JND at each of the three electrode locations. During the procedure, three intervals contained stimuli that were modulated at 100 Hz (the reference stimuli) and one contained a stimulus that was modulated at a frequency above 100 Hz (the target stimulus). To prevent the use of loudness differences as a cue for discrimination, the amplitudes of the target and the three reference stimuli were jittered up or down by a maximum of 4 Nucleus current levels (0.7 dB). Subjects were instructed to select the interval that was different from the others in any way other than loudness. The modulation frequency of the target stimulus was adjusted according to subject response by 10%. A total of ten reversals were measured, and the last six reversals were averaged to estimate the JND corresponding to 50% discrimination. Within each experimental block, the JND was measured once for all four waveforms and three electrodes. The order of stimulus presentation was randomized for each block. A total of five blocks were tested per subject.

To ensure that the ± 4 Nucleus current level (0.7 dB) jitter was sufficient to mask any loudness cues generated by the change in carrier rate, the loudness of each of the reference stimuli was balanced to a target stimulus modulated at the frequency minimally discriminable from 100 Hz (i.e., the JND). If an adjustment of more than 1 Nucleus current level (0.18 dB) was required to make the two stimuli equally loud, the modulation frequency JNDs were remeasured with the level-adjusted target stimulus.

A test was performed to verify that the detected changes for each stimulus were not based on the difference in carrier rate. Subjects were asked to discriminate an unmodulated pulse train at the reference stimulus carrier rate (2000 pps) from an unmodulated pulse train at the carrier rate of the target stimulus representing the largest measured difference limen on the same electrode, regardless of waveform. Thus, if the maximum modulation frequency JND was 10%, carrier rate discrimination was measured between 2000 and 2200 pps stimuli. Current levels were randomly jittered by ± 4 Nucleus current levels (0.7 dB). Subjects were asked to choose which sound was different, ignoring differences in loudness. Carrier rate discrimination was measured using a 4IFC task, and each comparison was repeated ten times for each electrode. While it was unlikely that there were perceptual differences between the carrier rates (Landsberger and McKay, 2005), it was important to verify this assumption to reduce concerns of carrier rate as a confounding variable.

3. Results

Figure 2 displays the modulation frequency JNDs for each of the four waveforms. There was no obvious advantage observed for any of the experimental waveforms, as performance across the

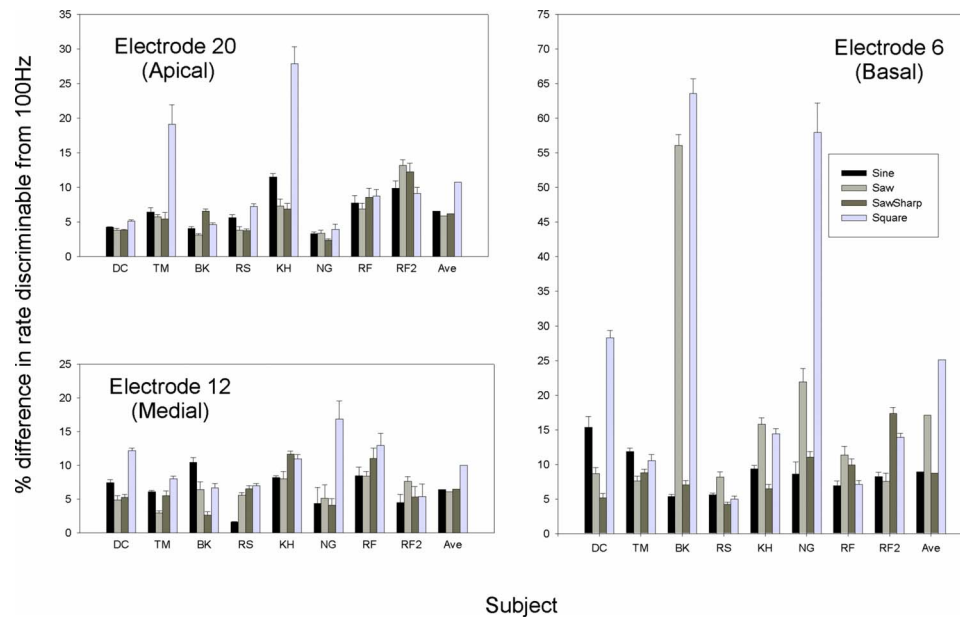


Fig. 2. (Color online) Modulation frequency JNDs (in percent difference re: 100 Hz) for individual subjects. The three panels show data for the three experimental electrode locations. The error bars show one standard error of the mean.

waveforms was relatively similar. The size of the JNDs was generally between 5 and 10%, consistent with previous sinusoidal modulation frequency discrimination data in cochlear implant users (Geurts and Wouters, 2001; Chatterjee and Peng, 2008). The optimal waveform for individual subjects often varied with the electrode. For example, for subject BK, the smallest JNDs were produced by the sawtooth on electrode 20, the sawsharp on electrode 12, and the sine on electrode 6. In general, the square waveform produced JNDs similar to those with the sine, saw, and sawsharp waveforms; about half of the JNDs with the square waveform were below 10%. However, for most subjects, on at least one of the electrodes, the JNDs measured with the square wave were much larger than those with the other waveforms; 25% of the JNDs with the square waveform were above 20%. These outlier JNDs with the square waveform varied inconsistently across subjects, in terms of electrode location. Note the outlier for subject BK with the saw waveform on electrode 6.

A two-way repeated measures analysis of variance found significant effects for both electrode location [$F(2, 42)=4.025, p=0.042$; power of analysis: 0.486] and modulation waveform [$F(3, 42)=5.862, p=0.005$; power of analysis: 0.849]. However, no significant interaction was observed between electrode location and waveform [$F(6, 42)=1.613, p=0.168$; power of analysis: 0.214]. Presumably, the main effect for waveform was due to the sometimes poorer and more variable performance with the square wave. A pairwise *post-hoc* test (Holm-Sidak) was performed to compare all waveforms. Significant differences were found between the square and the sine waveforms, and between the square and the sawsharp waveforms.

A binomial test was unable to detect a perceptual difference in the carrier discrimination task. Subjects were able to identify the different carrier rate between two and four out of ten times. Given a chance level of 0.25, a subject would have to identify the different carrier rate six or more times to obtain a significantly different result ($\alpha = 0.05$).

4. Discussion

In the present study, modulation frequency JNDs were similar for all experimental waveforms. The size of the JNDs was similar to that found in previous cochlear implant studies in which

sinusoidal modulation was presented to a single electrode (Geurts and Wouters, 2001; Chatterjee and Peng, 2008), or presented to multiple electrodes within a speech processing strategy (Geurts and Wouters, 2001). The data are consistent with the finding of Green *et al.* (2004) that glide discrimination was not significantly different for sawtooth and sharpened sawtooth waveforms. Note that in the present study, the minimum of the amplitude modulations were presented at 20% DR, and not at threshold as in the Green study.

In most cases, the size of the JNDs measured with the square waveform was similar to the JNDs measured with the other waveforms. However, there were a number of outliers with the square waveform across subjects and electrodes. It is unclear why performance with the square waveform was more prone to outliers than were the other waveforms. The outliers were unlikely to be a result of noisy data because the variability for the outlying conditions was relatively small. The outliers also did not seem to follow an obvious pattern, as they were observed for most subjects, but at different electrode locations. Thus, it is unlikely that the outliers were due to individual subject differences or differences in electrode location. Fu (2002) has shown that modulation detection thresholds are strongly correlated with speech recognition in cochlear implants. Pfingst *et al.* (2007) have argued that stimulating at sites within the cochlea which are poor at modulation detection might actually make speech perceptual tasks more difficult. It is possible that, within a speech processing strategy, modulating with a square wave may reduce speech comprehension.

The data suggest that when designing a speech processing strategy in which modulations are used to convey F0, a sine, saw, and sawsharp waveform are interchangeable. Note that in the present study, modulation frequency JNDs were only measured for 100 Hz stimuli with 80% modulation depth. It is unclear whether modulation frequency JNDs may differ among the waveforms at other base frequencies or modulation depths. The choice in waveform may be driven by concerns other than frequency discrimination. For example, it is unclear whether the waveform shapes may interact differently when presented on multiple electrodes within a speech processing strategy. Speech comprehension has often been reduced when modifying a speech processing strategy to include modulations at F0 across electrodes (Geurts and Wouters, 2001; Green *et al.*, 2005; Vandali *et al.*, 2005). It is possible that different modulation waveforms may change the magnitude of effect on speech comprehension. If there is no perceptual difference between the waveforms, the choice of waveform may depend on engineering issues. Using a waveform with a sharp attack would require additional algorithmic steps to ensure that there are no aliasing artifacts. Additionally, the amount of current required to achieve a fixed loudness would likely be higher for narrower waveforms (e.g., sawsharp) than for sine waveforms. As such, the choice of waveform might affect battery life in the speech processor. These engineering issues aside, the present data suggests that sine, saw, or sharpened waveforms may provide similar benefits for speech processing strategies that encode pitch via amplitude modulations.

Acknowledgments

The author would like to thank the subjects for giving their time and effort to this project. Additional gratitude is extended to John Galvin and three anonymous reviewers for their helpful comments. Funding for this project was provided by an NIDCD fellowship.

References and links

- Chatterjee, M., and Peng, S. C. (2008). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," *Hear. Res.* **235**, 143–156.
- Faulkner, A., Rosen, S., and Smith, C. (2000). "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," *J. Acoust. Soc. Am.* **108**, 1877–1887.
- Fu, Q. J. (2002). "Temporal processing and speech recognition in cochlear implant users," *NeuroReport* **13**, 1635–1639.
- Fu, Q. J., Chinchilla, S., Nogaki, G., and Galvin, J. J., 3rd (2005). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," *J. Acoust. Soc. Am.* **118**, 1711–1718.
- Fu, Q. J., Hsu, C. J., and Horng, M. J. (2004). "Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant users," *Ear Hear.* **25**, 501–508.

- Geurts, L., and Wouters, J. (2001). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *J. Acoust. Soc. Am.* **109**, 713–726.
- Green, T., Faulkner, A., and Rosen, S. (2004). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," *J. Acoust. Soc. Am.* **116**, 2298–2310.
- Green, T., Faulkner, A., Rosen, S., and Macherey, O. (2005). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," *J. Acoust. Soc. Am.* **118**, 375–385.
- HearWorks Pty. Ltd. (2003). "Spear 3 Research System" <http://www.hearworks.com.au/spear>, Last viewed 6/20/08.
- Landsberger, D. M., and McKay, C. M. (2005). "Perceptual differences between low and high rates of stimulation on single electrodes for cochlear implantees," *J. Acoust. Soc. Am.* **117**, 319–327.
- Laneau, J., Wouters, J., and Moonen, M. (2004). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," *J. Acoust. Soc. Am.* **116**, 3606–3619.
- Laneau, J., Wouters, J., and Moonen, M. (2006). "Improved music perception with explicit pitch coding in cochlear implants," *Audiol. Neuro-Otol.* **11**, 38–52.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Pfingst, B. E., Xu, L., and Thompson, C. S. (2007). "Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants," *J. Acoust. Soc. Am.* **121**, 2236–2246.
- Seligman, P., and McDermott, H. (1995). "Architecture of the Spectra 22 speech processor," *Ann. Otol. Rhinol. Laryngol. Suppl.* **166**, 139–141.
- Shamma, S., and Klein, D. (2000). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," *J. Acoust. Soc. Am.* **107**, 2631–2644.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (2000). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," *Ear Hear.* **21**, 608–624.
- Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W. D., and McDermott, H. J. (2005). "Pitch ranking ability of cochlear implant recipients: A comparison of sound-processing strategies," *J. Acoust. Soc. Am.* **117**(5), 3126–3138.
- Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., and Clark, G. M. (1995). "Evaluation of the Nucleus Spectra 22 processor and new speech processing strategy (SPEAK) in postlinguistically deafened adults," *Acta Oto-Laryngol.* **115**, 629–637.
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," *Nature (London)* **352**, 236–238.