Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees[a)]

Johan Laneau[b)] and Jan Wouters
Lab. Exp. ORL, K.U.Leuven, Kapucijnenvoer 33, B 3000 Leuven, Belgium

Marc Moonen
ESAT-SCD, K.U.Leuven, Kasteelpark Arenberg 10, B 3001 Leuven, Belgium

(Received 9 February 2004; revised 24 September 2004; accepted 1 October 2004)

The effect of the filter bank on fundamental frequency ($F0$) discrimination was examined in four Nucleus CI24 cochlear implant subjects for synthetic stylized vowel-like stimuli. The four tested filter banks differed in cutoff frequencies, amount of overlap between filters, and shape of the filters. To assess the effects of temporal pitch cues on $F0$ discrimination, temporal fluctuations were removed above 10 Hz in one condition and above 200 Hz in another. Results indicate that $F0$ discrimination based upon place pitch cues is possible, but just-noticeable differences exceed 1 octave or more depending on the filter bank used. Increasing the frequency resolution in the $F0$ range improves the $F0$ discrimination based upon place pitch cues. The results of $F0$ discrimination based upon place pitch agree with a model that compares the centroids of the electrical excitation pattern. The addition of temporal fluctuations up to 200 Hz significantly improves $F0$ discrimination. Just-noticeable differences using both place and temporal pitch cues range from 6% to 60%. Filter banks that do not resolve the higher harmonics provided the best temporal pitch cues, because temporal pitch cues are clearest when the fluctuation on all channels is at $F0$ and preferably in phase. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1823311]

PACS numbers: 43.66.Ts, 43.66.Fe, 43.66.Hg [GDK] Pages: 3606–3619

I. INTRODUCTION

Although cochlear implants (CI) have helped many deaf people in their daily lives by providing a means for oral communication, the sound that CI patients perceive is very different from the sounds normal-hearing subjects perceive. Cochlear implant patients have, for instance, problems discriminating pitch (Gfeller and Lansing, 1992; Pijl, 1997; Geurts and Wouters, 2001). Geurts and Wouters found just-noticeable differences (jnd) of fundamental frequency ($F0$) from 4% up to 13% in four Laura cochlear implant subjects (Geurts and Wouters, 2001). These jnd's are very poor compared to normal-hearing subjects who typically perceive differences of less than 0.5% (Moore and Glasberg, 1990). However, pitch perception is very important for intonation, speaker identification, music perception, and speech perception in tonal languages. In an attempt to improve the $F0$ discrimination for cochlear implant recipients, this study assessed the effects of the filter bank design on the discrimination of the fundamental frequency of speech-like harmonic sounds. The filter banks tested in the present experiment differ in cutoff frequency distribution in the low frequencies, amount of overlap between filters, and filter shape.

Pitch cues can be elicited using two independent mechanisms in cochlear implants: place pitch cues and temporal pitch cues (Tong *et al.*, 1982; Shannon, 1983; Townshend *et al.*, 1987; McKay *et al.*, 2000). Temporal pitch cues arise when the repetition rate of stimulation on one channel changes. The temporal pitch sensation rises with increasing rate up to 300 Hz but saturates at higher rates (Shannon, 1983). Place pitch cues arise when the site of stimulation is changed while keeping the stimulation rate constant, with more basal stimulation eliciting higher pitches.

It has been shown that both temporal and place pitch cues enable CI recipients to some degree to perceive $F0$ differences of synthetic harmonic sounds with currently used sound processors (Geurts and Wouters, 2001), and in normal-hearing subjects with acoustic models of CI processors (Green *et al.*, 2002). Geurts and Wouters (2001) measured the discrimination of $F0$ of synthetic steady-state vowels in subjects implanted with the 8-channel Laura implant. Their results indicated that even without temporal pitch cues, only preserving place pitch cues, some discrimination of $F0$ was possible although the jnd's were very large. Without temporal pitch cues the jnd's were on average around 50% but varied over conditions and subjects from 5% to more than 100%. With the inclusion of temporal information, the jnd's decreased significantly, as would be expected. Green *et al.* (2002) examined the discrimination of CI simulations of frequency glides in sawtooth waves and synthetic diphthongs in normal-hearing subjects. They studied the influence of both place and temporal pitch in modulated noise-band vocoder CI simulations, and concluded that both pitch mechanisms contribute to the discrimination of $F0$. Subjects were able to integrate both cues as the inclusion of both pitch cues produced the highest discriminative performance. However, when the original sounds were diphthongs, the inherent variation of formant structures perturbed the place pitch cue.

The current sound processing in cochlear implants has a

[a)]Portions of this work were presented in "Pitch perception in cochlear implants with different filter bank designs," Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun, Mexico, 2003.

[b)]Electronic mail: johan.laneau@med.kuleuven.ac.be

 © 2004 Acoustical Society of America

TABLE I. Relevant information about each of the subjects who participated in the experiments. The last column shows the results for a phoneme recognition task with monosyllabic consonant–vowel–consonant (CVC) words of the NVA list (Wouters *et al.*, 1994).

Subject	Duration of profound deafness (y)	Etiology	Implant type	Speech processor	Clinical speech processing strategy	Age (y)	Implant experience (y;m)	Phoneme recognition (%)
S1	11	Hereditary	Nucleus CI24R(ST)	ESPrit	Speak	19	1;11	74%
S2	3	Unknown	Nucleus CI24R(ST)	ESPrit	Speak	21	3;2	86%
S3	30	Progressive	Nucleus CI24R(CS)	SPrint	ACE	54	0;8	84%
S4	32	Progressive	Nucleus CI24R(CS)	ESPrit 3G	ACE	49	0;6	52%

negative effect on $F0$ discrimination, since CI subjects appear to be able to estimate musical intervals more accurately for pulse trains than for real musical sounds processed by their speech processors (Pijl, 1997). Part of this limited $F0$ discrimination ability in CI recipients using their speech processors may be due to the limited transmission of the fine temporal information in current speech processors, since most current speech processors only extract the slowly varying envelope and use this to modulate a constant rate pulse train (Smith *et al.*, 2002).

Because of this limited use of fine temporal pitch cues in current CI sound processors, previous studies have tried to improve the transmission of temporal pitch cues in order to improve the $F0$ discrimination in cochlear implants (Jones *et al.*, 1995; Geurts and Wouters, 2001; Barry *et al.*, 2002). However, no study resulted in a significant improvement of $F0$ discrimination. Explicit presentation of the fundamental frequency by stimulating the most apical channel with a pulse train at the rate of $F0$ did not improve the perception of suprasegmental cues such as intonation (Jones *et al.*, 1995). Barry *et al.* (2002) assessed the effect of pulse rate on the perception of tones in Cantonese-speaking children by comparing a group of children using the SPEAK strategy to a group of children using the higher rate ACE strategy. The hypothesis was that higher rates better preserve the temporal code (Wilson, 1997; Rubinstein *et al.*, 1999), but the authors found no significant differences between the strategies. Indeed, the group of children with the ACE strategy performed slightly (but nonsignificantly) worse than the group of SPEAK users. Geurts and Wouters (2001) developed a variation to the CIS sound-processing scheme that enhanced the modulation depth of the beating of the output of the filters, and also did not find a significant difference when compared to the standard CIS in the discrimination of $F0$ of synthetic vowels.

In a recent study, Geurts and Wouters (2004) showed that a new filter bank can improve the place pitch cues for $F0$ discrimination of synthetic vowels. Their newly proposed filter bank had more resolution in the lower frequencies and more filter overlap between different channels compared to the reference filter bank in that study. An important improvement in $F0$ discrimination based only upon place pitch cues was obtained but it remains, however, unclear whether this is due to the increased frequency resolution, the smooth transition between filters, or the increased filter overlap. With the inclusion of temporal cues, the improvement was small or even absent.

The present study investigates which aspect of the filter bank (cutoff frequency distribution, filter shape, or filter overlap) has the strongest effect on $F0$ discrimination for both place pitch cues and temporal pitch cues. It also investigates which physical properties of the stimuli are related to the place pitch cues that facilitate $F0$ discrimination, and which lead to clearer temporal pitch cues.

The effect of four filter banks on the voice-range pitch discrimination is compared in cochlear implant users. In order to separate the effects of place pitch cues and temporal pitch cues on complex pitch perception, two experimental conditions were created. In the first condition, temporal pitch cues were excluded by low-pass filtering the stimulation envelopes at 10 Hz, and consequently only place pitch cues were present. In the second condition both place and temporal pitch cues (up to approximately 200 Hz) are present. Performance on $F0$-discrimination task was measured as a function of relative difference of fundamental frequency, using a constant-stimuli 2I2AFC pitch-ranking experiment. Four Nucleus CI24 cochlear implant subjects took part. The harmonic signals were synthetic stylized vowel-like single-formant stimuli.

II. METHODS

A. Subjects

Four experienced postlingually deafened users of the Nucleus CI24 cochlear implant participated in this study. All subjects performed all tests and were paid for their collaboration. Some relevant details about the subjects can be found in Table I. Subjects S1 and S2 are implanted with the straight version of the Nucleus electrode array CI24R(ST) Nucleus24k and use a SPEAK strategy in daily life. Subjects S3 and S4 are implanted with the perimodiolar version of the Nucleus electrode array CI24R(CS) Nucleus24 Contour and

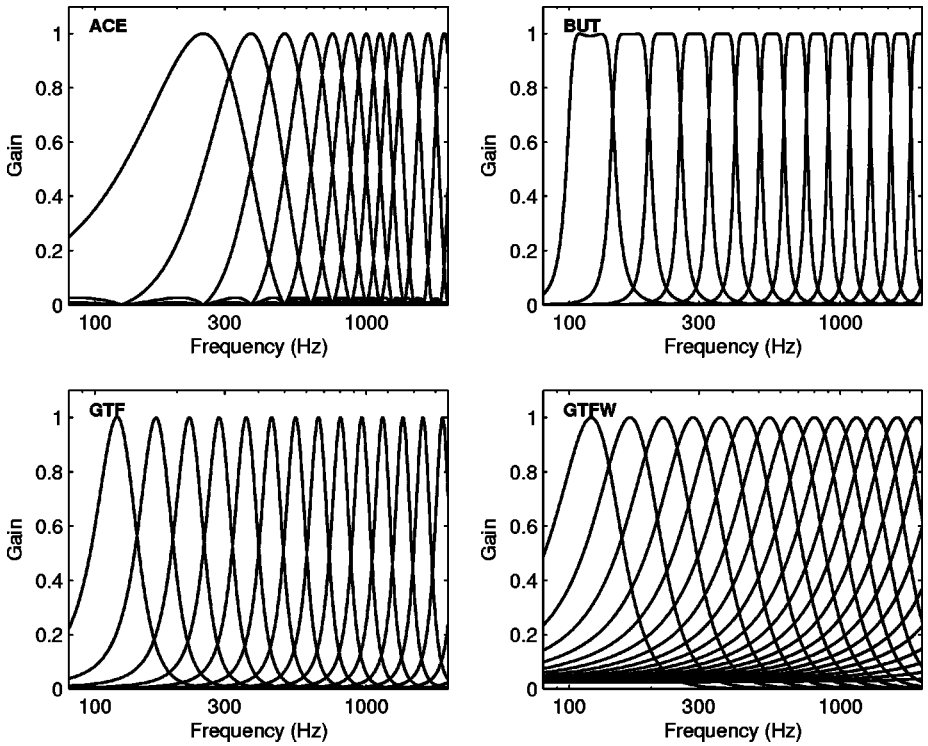

FIG. 1. The frequency response of the different filter banks examined in the present study. The ACE filter bank is the filter bank implemented in the commercial Sprint speech processor. The GTF filter bank is based upon the physiology of the normal ear and has a higer resolution in the lower frequencies. The BUT filter bank is based upon the GTF filter bank but with more rectangular filter shapes. The GTFW is also based upon the GTF filter bank but the bandwidth of the filters of the GTFW filter bank were set not to resolve any overtone. Because the stimuli in the present study contained virtually no energy in higher frequencies, the frequency response is represented up to 2000 Hz for clarity.

use an ACE strategy (900 pps per channel) in daily life. All subjects are relatively good performers with their implants in daily life.

B. Sound processing

Four different filter banks were used to process the stimuli. The frequency response of the four filter banks is depicted in Fig. 1. The filter banks differ in spectral resolution for the low frequencies and in the shape of the individual filters. All filter banks have 22 bands corresponding to the number of channels in the Cochlear Nucleus CI24 using monopolar stimulation mode. One filter bank (ACE) is at present most commonly used clinically for Nucleus CI24, and is included in the present experiment as a reference for current performance of cochlear implant subjects. The three other filter banks are specifically designed for this experiment and resemble more closely the frequency analysis of the normal human ear. All three filter banks have more resolution in the low frequencies compared to the ACE filter bank. The GTF filter bank is based upon a model for the filtering of the normal ear (Patterson *et al.*, 1995). The BUT and GTFW are derived from this filter bank by changing the filters' shape and bandwidth, respectively.

The first filter bank (ACE) is the filter bank of the standard commercial speech-processing strategy (Cochlear Corp: ACE) of the Cochlear SPrint speech processor implemented in MATLAB (Cochlear Ltd, 2002). It consists of a 128-point FFT with a Hamming window, leading to 64 frequency bins. The first two bins, containing the dc and the lowest frequencies, are discarded. The following eight bins are assigned to the first eight channels of the filter bank, respectively. The higher bins are summed in magnitude in groups of two or more to form the rest of the channels, leading to an approximately logarithmic frequency distribution in the higher channels.

The GTF, BUT, and GTFW filter bank all consist of 22 fourth-order IIR bandpass filters. The filters in the GTF filter bank were fourth-order gammatone filters using Slaney's implementation (Slaney, 1993). The center frequencies of the gammatone filters were set corresponding to the middle of 22 equally long sections along the cochlea ranging from 100 to 8000 Hz in total. The transformation from frequency to distance along the cochlea was calculated by Greenwood's formula for a cochlear length of 35 mm and $\alpha = 0.06$ (Greenwood, 1990). The bandwidth of the filter was set to the equivalent rectangular bandwidth (ERB) of the auditory filter corresponding to its center frequency (Glasberg and Moore, 1990).

The BUT filter bank is similar to the GTF filter bank; however, the shape of the filters is more rectangular (a flat frequency response in the passband and 24-dB/octave slopes). The BUT filter bank consists of 22 fourth-order Butterworth filters, of which the cutoff frequencies were equidistantly spaced along the length of the cochlea with cutoff frequencies ranging from 100 to 7800 Hz in total. The highest cutoff frequency was set slightly below the Nyquist frequency to avoid numerical problems. Both the GTF and the BUT filter bank resolve one or two of the lowest harmonics for most of the stimuli in the present study. The GTFW filter bank is based upon the GTF filter bank but the bandwidth of the filters is increased so that all overtones are unresolved. This caused the output of every filter for a harmonic sound to beat at the fundamental frequency. The GTFW was implemented identically to the GTF filter bank except that the bandwidth was set at half the filter's center frequency.

In order to separate the effects of temporal and place pitch cues, the envelopes of the filter outputs were obtained

in two different ways. In one condition, including both temporal and place pitch cues, the envelope contained modulations up to 200 Hz. In the other condition, including place pitch cues but without temporal pitch cues, the envelope only contained modulations up to 10 Hz. The deletion of modulations at frequencies above 10 Hz removes all fluctuations that generate temporal pitch percepts, although listeners can still understand speech, at least when presented in a quiet background (Drullman *et al.*, 1994; Shannon *et al.*, 1995).

In the ACE filter bank, the envelope is extracted by taking the magnitude of the complex-valued output. The bin width of the FFT filter bank limits the maximal modulation frequency after this envelope extractor to approximately 200 Hz. In the condition without temporal pitch cues these envelopes were, in addition to the standard ACE processing, low-pass filtered with a fourth-order Butterworth filter at 10 Hz.

The envelopes for the GTF, BUT, and GTFW were extracted using a half-wave rectifier and a fourth-order Butterworth low-pass filter. The cutoff frequency of the low-pass filter was set at 200 Hz when temporal pitch cues were transmitted and at 10 Hz when temporal pitch cues were removed.

The subsequent processing was identical for all filter banks. First, the signals were resampled from the audio sampling rate (16 kHz) to the channel stimulation rate (900 Hz). Then, only the eight most intense channels per time slot were selected for stimulation. Finally, the amplitude of the signals is compressed using the standard compression function of the Nucleus MATLAB Toolbox (Cochlear Ltd., 2002), in order to map the acoustic input level onto the subjects' electrical dynamic range. The compression function is defined by

$$y(x)=\begin{cases} 0 & \text{for } 0\leqslant x<0.0156 \\ \log\left(1+416.2\dfrac{x-0.0156}{0.5703}\right)\Big/\,6.033 & \text{for } 0.0156\leqslant x<0.5859. \\ 1 & \text{for } 0.5859\leqslant x\leqslant 1 \end{cases} \tag{1}$$

In formula (1) x is the amplitude of the extracted envelopes for each channel and y is the compressed amplitude expressed in proportion of the dynamic range expressed in clinical current units of the subject for that channel. The obtained envelopes were used to modulate pulse trains of 900 pulses per second with biphasic pulses of 25-μs phase width and 8-μs interphase gap. The pulses were delivered in monopolar mode with both external return electrodes active (MP1+2).

C. Stimuli

Original stimuli before sound processing were synthetic harmonic complexes produced by the excitation of a single-formant resonator by a periodic signal, and resemble highly stylized vowels. The generation of the single-formant stimuli (SFS) was based upon the Klatt speech synthesizer (Klatt, 1980). The stimuli were generated by filtering a 500-ms pulse train sampled at 16 kHz and ramped on and off with a raised cosine over 10 ms with a cascade of two filters.

The first filter models the voice source and its output is similar to the typical glottal volume velocity waveform. It is implemented as a second-order low-pass filter with a cutoff frequency of 50 Hz (Klatt, 1980). The second filter is a resonator that shapes the single formant of the stimuli. It is implemented as a second-order bandpass filter, of which the transfer function is defined by

$$H(z)=\frac{A}{1-Bz^{-1}-Cz^{-2}}$$

$$\text{with } \begin{aligned} C&=-\exp(-2.\pi B_w T) \\ B&=2\exp(-\pi B_w T)\cos(2\pi F_r T), \\ A&=1-C-B \end{aligned} \tag{2}$$

where T is the sampling period (62.5 μs) and F_r and B_w are the filter's center frequency and bandwidth, respectively (Klatt, 1980).

The fundamental frequency of the stimuli used in the present study ranged from 133 Hz up to 558 Hz. The center frequencies of the bandpass filter, or formant frequencies ($F1$), were 300, 350, 400, 450, or 500 Hz, and the bandwidth of the formant resonator was fixed at 100 Hz for all stimuli. The stimuli were first equalized in rms and then scaled with a common factor so that most of the signals nearly optimally used the input dynamic range of the compression function [Eq. (1)] but none of the signals was clipped. Figure 2 shows the spectrum of two representative stimuli.

D. Psychophysical procedure

Fundamental frequency discrimination was measured in a two-interval, two-alternative forced choice (2I2AFC) procedure using the method of constant stimuli. The subjects were instructed to indicate which of the two intervals contained the stimulus with higher pitch, while ignoring the loudness cues. The intervals were separated by a 500-ms silent gap and no feedback was provided, in order to force the subjects to base their judgments on a perceptual dimension that corresponded to their intuitive sense of pitch. This further reduced the likelihood of any residual loudness cues affecting their judgments.

Both stimuli in a trial had the same formant frequency. The $F0$ of one stimulus in each 2I2AFC trial was equal to the reference frequency, and that of the other was set at a higher frequency. This reference frequency was 133 or 165 Hz. These frequencies lie in the range of normal adult voice,

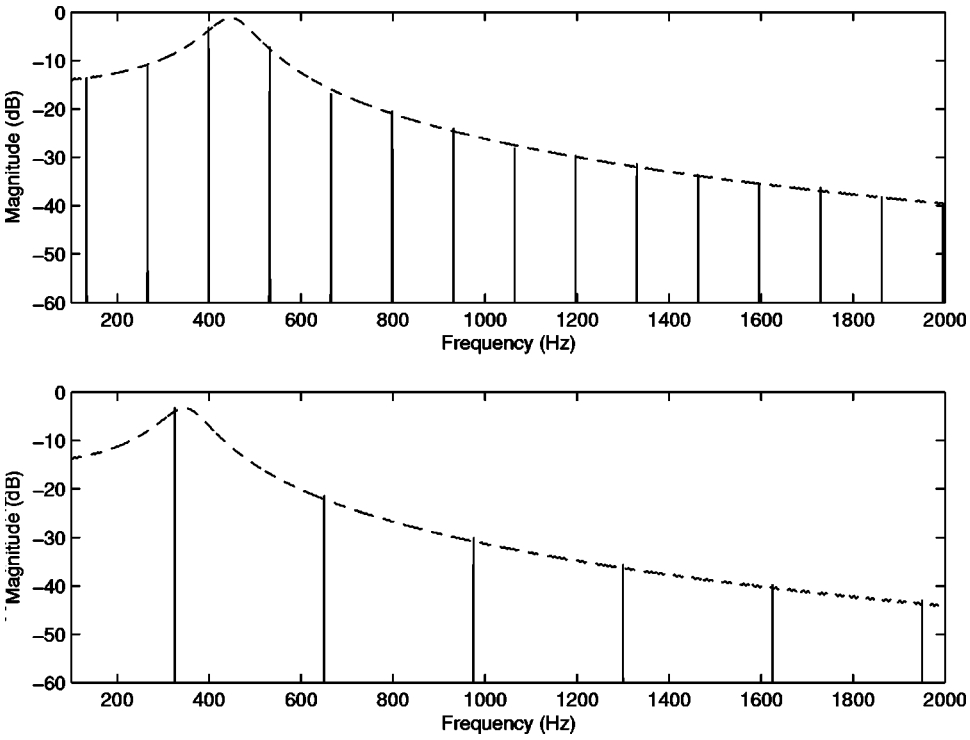

FIG. 2. The spectra of two representative stimuli used in the present study. The upper panel shows the spectrum of a single-formant stimulus with formant frequency 450 Hz and fundamental frequency 133 Hz. The lower panel shows the spectrum of a single-formant stimulus with formant frequency 350 Hz and fundamental frequency 325 Hz. The dotted line in each panel shows the combined frequency response of the two filters that shape the spectrum (see the text). The plot of the frequency response is limited up to 2000 Hz for clarity.

and 133 Hz is set near the first crossover between filters for the IIR filter banks, and 165 Hz is set near the center of the second filter of the IIR filter banks. The $F0$ frequencies of the eight comparison stimuli were 140, 150, 165, 180, 210, 250, 325, and 450 Hz for the reference stimuli with an $F0$ of 133 Hz. For the reference stimuli with an $F0$ of 165 Hz, the eight comparison stimuli had $F0$ frequencies of 174, 186, 205, 223, 261, 310, 403, and 558 Hz. The relative differences between $F0$ are the same for both reference $F0$ conditions and range from 5.2% up to 238%. The formant frequency was kept constant within trials because stimuli with the same $F0$ but with their main energy in different frequency regions can lead to differences in pitch (Ohgushi, 1978).

The trials were presented in blocks. All 16 conditions resulting from the combinations of type of envelope extraction (with or without temporal pitch cues), filter bank (ACE, GTF, BUT, or GTFW), and reference $F0$ (133 or 165Hz), were presented in separate blocks. Each block contained trials with all eight comparison frequencies for all five formant frequencies. These 40 different trials were repeated twice, leading to 80 trials per block. Different formant frequencies were included per block in order to prevent subjects from identifying and learning the reference stimulus based upon other cues. Every block was presented twice to every subject. Care was taken that blocks processed with the ACE, GTF, and BUT filter bank were presented in random order. The blocks processed with the GTFW filter bank were presented at the end of the experiment.

To reduce the influence of loudness cues, amplitude was roved by multiplying the amplitude of all stimuli with a random factor between 0.85 and 1.1, with amplitude expressed in clinical current units and relative to the dynamic range. An exception was made for subject S4 in the condition of the ACE filter bank without temporal pitch cues and the reference $F0$ of 165 Hz, where the subject judged the loudness of the stimuli as uncomfortably loud using the default amplitude roving. In this one case, the amplitude was roved with a factor between 0.77 and 1.

III. RESULTS

The proportion of times the comparison stimulus was judged higher than the reference stimulus, pooled over all formant frequencies, is shown in the absence and presence of temporal pitch cues by Figs. 3 and 4, respectively. The top four rows of panels in each figure show the results for a single subject, and the mean data are shown in the bottom, fifth row of panels. The left column in every figure shows the results for the condition with a reference $F0$ of 133 Hz and the right column shows the results for the condition with a reference $F0$ of 165 Hz. The proportions depicted in Figs. 3 and 4 are based upon 20 presentations per data point.

A repeated measures analysis of variance (ANOVA) was performed on the proportions of correctly ranked trials using five factors. The independent variables were presence of temporal pitch cues (absent or present), reference $F0$ (133 or 165 Hz), filter bank (ACE, GTF, BUT, or GTFW), relative $F0$ difference (eight levels), and formant frequencies (five different $F1$). The main effects of presence of temporal pitch cues ($p=0.002$) and relative $F0$ difference ($p=0.007$) were significant. The other main factors were not significant: reference $F0$ ($p=0.678$), filter bank ($p=0.251$), and formant frequency ($p=0.122$). The latter result allows averaging over the different formant frequencies in the presentation of the data in Figs. 3 and 4. Only the interaction term between filter bank and presence of temporal pitch cues was significant ($p=0.015$). No other interaction terms were significant, but there were significant intersubject differences ($p<0.001$).

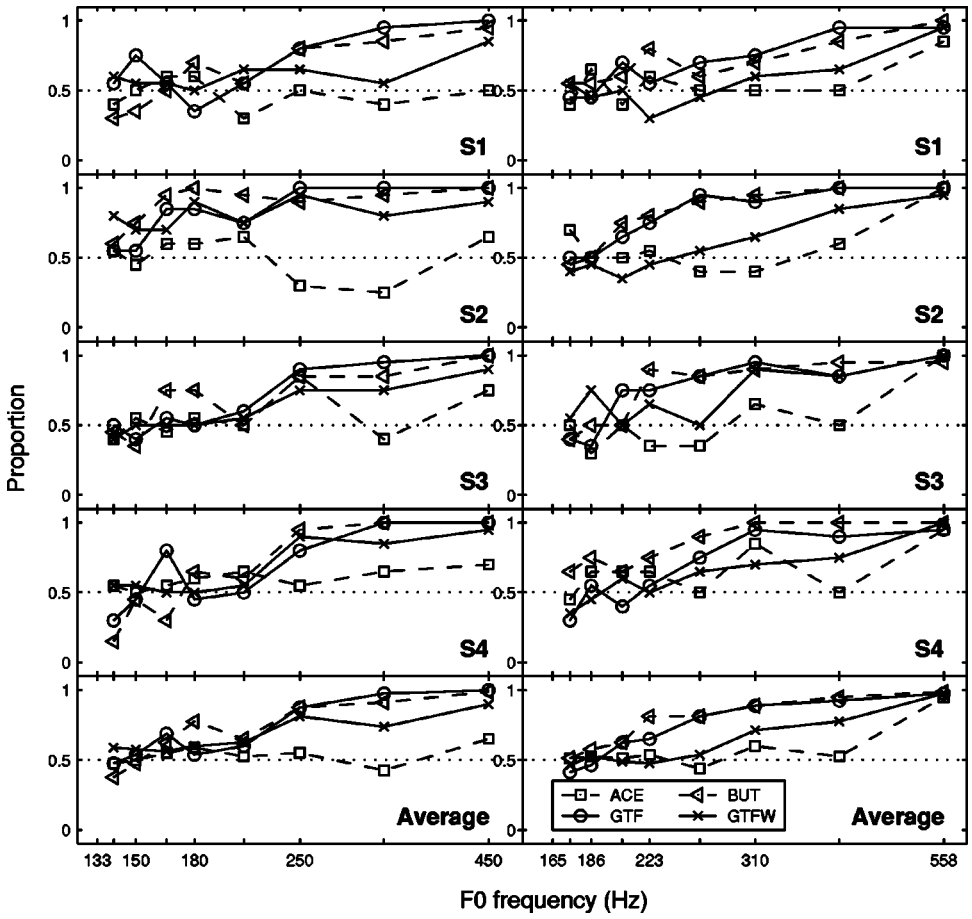

FIG. 3. The average proportion of correctly pitch-ranked stimuli pooled over the various formant frequencies. The different lines in the figures represent the results for the various filter banks examined in the present study, with all temporal fluctuations above 10 Hz filtered out. The fundamental frequencies of the signals are plotted on the horizontal axis on a log scale. The left and right columns of panels depict the results for the condition where the reference $F0$ was 133 and 165 Hz, respectively. Rows 1 to 4 represent a different subject. The bottom panels represent the average of the four subjects. The dotted line indicates the chance level.

Figures 3 and 4 indicate that subjects were able to rank the stimuli based upon their fundamental frequency. They also show that as the relative $F0$ difference increases the difference is perceived more clearly. Also, comparing Figs. 3 and 4, the results indicate that providing temporal fluctuations in the envelope of the channels up to 200 Hz improves the ability of the subjects to discriminate $F0$ differences. This improvement is, however, dependent on the filter bank used, and the largest improvement is found for the ACE filter bank.

Because of the significant interaction between presence/absence of temporal pitch cues and the filter banks, the results are presented in two parts: with or without temporal pitch cues.

A. Without temporal pitch cues

A four-factorial repeated-measures analysis of variance was performed on the proportions of correctly pitch-ranked trials that were collected with stimuli without temporal pitch cues. The main effects of filter banks ($p=0.030$) and relative $F0$ difference ($p=0.010$) were significant. The effects of formant frequency and reference $F0$ were not significant, as well as all interaction terms. The between-subject differences were significant ($p<0.001$). Pairwise comparisons were performed using the least-significant difference (LSD) on the filter bank design. The GTF and BUT filter bank produced significantly better performance than the ACE filter bank ($p=0.035$ and $p=0.018$, respectively for GTF and BUT) and the GTFW filter bank ($p=0.049$ and $p=0.015$, respectively for GTF and BUT), but there was no significant difference between BUT and GTF. The GTFW filter bank allowed for significantly better performance than the ACE filter bank ($p=0.046$).

The results indicate that even without temporal pitch cues it was still possible for the subjects to rank the stimuli based on $F0$. However, it must be emphasized that this was only possible for relatively large differences in $F0$ of about one octave for the BUT and GTF filter bank and more than one octave for the GTFW filter bank. For the ACE filter bank, the subjects found it almost impossible to use place pitch cues to rank $F0$, even when stimuli differed more than 20 semitones.

B. Including temporal pitch cues

A four-factorial repeated measures ANOVA was performed on the proportions of correctly ranked trials where the stimuli now contained temporal pitch cues. The main factors reference $F0$ ($p=0.006$) and relative $F0$ difference ($p=0.013$) had a significant effect. No other main effects were significant. The effect of the filter bank was not significant ($p=0.126$).

The discrimination was significantly poorer at the higher reference $F0$ tested. This result agrees with previous studies showing that the effectiveness of temporal pitch cues decrease with increasing overall rate (Shannon, 1983).

Three of the four subjects (S1, S3, and S4) did not reach perfect performance for the ACE filter bank and reference

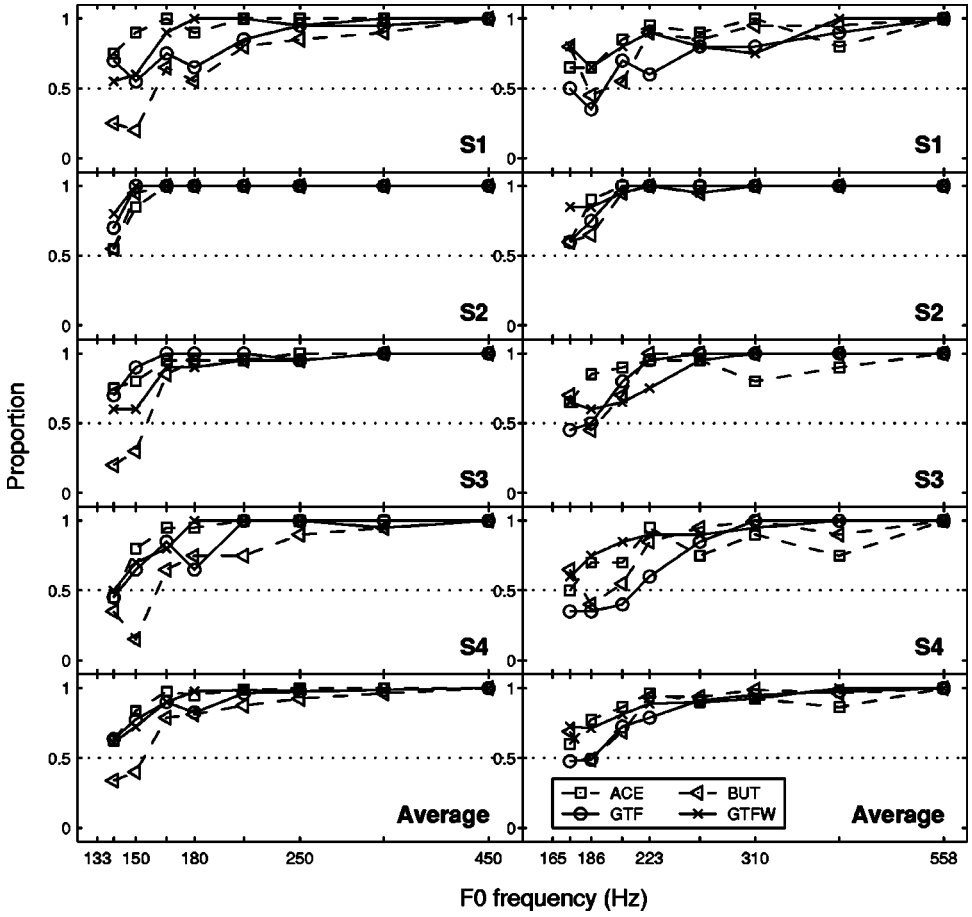

FIG. 4. The average proportion of correctly pitch-ranked stimuli pooled over the various formant frequencies. The different lines in the figures represent the results for the various filter banks examined in the present study, with all temporal fluctuations included up to 200 Hz. The fundamental frequencies of the signals are plotted on the horizontal axis on a log scale. The left and right columns of panels depict the results for the condition where the reference $F0$ was 133 Hz and 165 Hz, respectively. Rows 1 to 4 represent a different subject. The bottom panels represent the average of the four subjects. The dotted line indicates the chance level.

$F0$ of 165 Hz until the relative $F0$ difference was more than one octave. Nevertheless, the performance was above chance level for relatively small $F0$ differences. This indicated that further increasing the $F0$ difference did not make the two processed stimuli perceptually more dissimilar. This was probably due to the poor resolution of the ACE filter bank in the $F0$ range causing $F0$ to be encoded solely by temporal pitch cues, whose effectiveness was limited at high $F0$'s. This limitation might have been caused by the limited modulation depth of the envelopes at higher frequencies or by the reduced perceptual differences for temporal pitch cues at higher rates. The three IIR filter banks (GTF, BUT, and GTFW) did not suffer from this shortcoming because, for larger $F0$ differences, the pitch discrimination was most probably also mediated by place pitch. The difference in results between the condition without temporal pitch cues and the condition with both place and temporal pitch cues quantifies the influence of temporal pitch cues present in the presented stimuli. To assess this, the proportions of correctly ranked trials were averaged over the four subjects and converted into d' values for all conditions and all stimuli. To prevent infinite d' values for perfect responses, the perfect responses are adjusted by introducing an error of half a trial as suggested by MacMillan and Creelman (1991). The amount of perceptual difference, introduced by the temporal pitch cues, is the difference of the d' values between the condition with temporal pitch cues ($d'_{\text{place+temp}}$) and the condition without temporal pitch cues (d'_{place}). Assuming that both temporal and place pitch cues are optimally used and independent (McKay *et al.*, 2000), and taking into account the possibility of pitch reversals, the amount of temporal pitch cues is calculated by

$$d'_{\text{temp}} = \text{sign}(X) * \sqrt{|X|} \quad \text{with} \quad X = \text{sign}(d'_{\text{place+temp}}) \times (d'_{\text{place+temp}})^2 - \text{sign}(d'_{\text{place}})(d'_{\text{place}})^2. \tag{3}$$

This formula simplifies to $d'_{\text{temp}} = \sqrt{d'^{\,2}_{\text{place+temp}} - d'^{\,2}_{\text{place}}}$ for discrimination experiments as stated in McKay *et al.* (2000) because in discrimination experiments only positive d' values are possible. However, in a pitch-ranking experiment pitch reversals can occur and these lead to negative d' values. Moreover, in a pitch-ranking test the subjects are forced to project their percepts onto a single axis (''the combined pitch axis''). When temporal and place pitch cues of two compared stimuli are conflicting the two stimuli are discriminable (McKay *et al.*, 2000), but ranking them in pitch is at least very hard and might possibly lead to a d' value close to zero.

All data points where both conditions, with and without temporal pitch cues, resulted in perfect discrimination were discarded because these points did not provide any information on the effect of adding temporal pitch cues. The additional perceptual effect based upon the temporal pitch cues is shown in Fig. 5. Data points represent averages over formant frequency and reference $F0$. The temporal pitch cues become clearer as the $F0$ difference increases for all filter banks, but the clarity of the temporal pitch cues differs between filter banks. The ACE and GTFW filter bank, having broad filters,

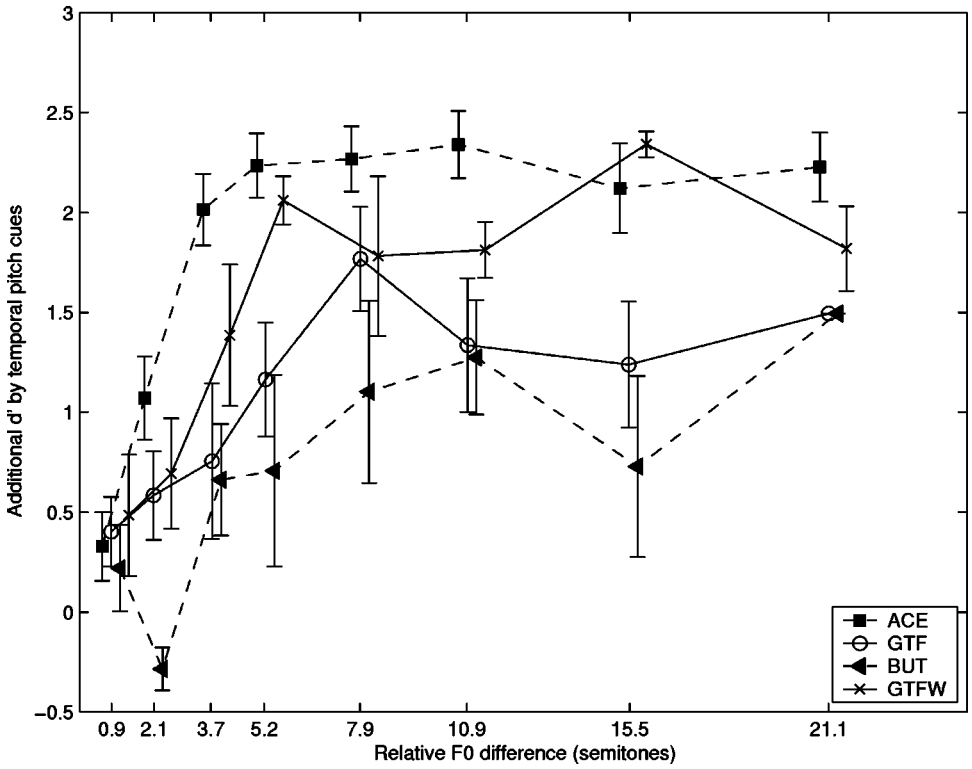

FIG. 5. The additional d' caused by the inclusion of temporal pitch cues in the processed stimuli, for the four filter banks tested in this study. The additional d' values are calculated based upon the average response of the four CI subjects, and are derived from the difference in d' between the corresponding data points with and without temporal pitch cues (see the text for details). The additional d' is averaged over the five different formant frequencies and over both reference $F0$ conditions. The error bars indicate the standard error of the mean. The data points are slightly shifted along the x axis for clarity.

allow for the largest increase in performance based upon temporal pitch cues. The GTF and BUT filter bank, which have narrower filters, exhibit less increase of the discrimination performance possibly because the baseline performance was lower.

IV. DISCUSSION

A. Without temporal pitch cues

Previous research has shown that $F0$ discrimination can take place in the absence of temporal pitch cues in CI subjects (Geurts and Wouters, 2001; Geurts and Wouters, 2004). This is confirmed by our results. Additionally, it is shown that this ability to discriminate $F0$ based on place pitch cues is highly dependent on the filter bank used. The ACE filter bank did not allow the $F0$ information to be coded by place pitch cues. However, the BUT, GTF, and GTFW filter banks, which have more resolution in the lower frequencies, did enable the subjects to perceive $F0$ differences based upon place pitch. The GTFW has slightly less resolution compared to GTF and BUT due to the large filter overlap, and consequently the $F0$ discrimination using this filter bank is somewhat lower compared to the latter two filter banks. The actual shape of the filter bank, being more rectangular with steep slopes and less overlap (BUT), or more triangular with shallower slopes and more overlap (GTF), did not significantly change the performance. This contrasts with the results of Geurts and Wouters (2004), who found an improved performance for a triangular-like filter bank. However, given the present results, the improved performance of the triangular filter bank in that study was probably facilitated by the different cutoff frequencies used in the triangular filter bank, leading to more resolution in the lower frequency register. For subjects who have poor temporal pitch sensitivity but good place pitch sensitivity, the filter banks with high resolution in the $F0$ register may enable them to better discriminate $F0$.

Some of the psychometric curves of Fig. 3 are nonmonotonic and contain pitch reversals, some of which are consistent over the various subjects. In addition, when pooling the data for the different subjects and plotting the separate psychometric curves for the different $F1$ values, some significant pitch reversals can be observed, e.g., the psychometric curve for $F1=500$ Hz in Fig. 6(A). An analysis of all conditions that resulted in significant pitch reversals after pooling the data over the different subjects reveals a consistent pattern of the average channel amplitude before compression. This pattern is most easily interpreted by means of an example. Figure 6(B) shows the average amplitude of the 22 channels for a stimulus with $F1=500$ Hz, processed with the GTF filter bank. The thick line represents the normalized average amplitude for the reference stimulus having an $F0$ of 133 Hz. The two thin lines represent the normalized average amplitude for two comparison stimuli, $F0=150$ Hz (solid line) and $F0=250$ Hz (dashed line). The former stimulus resulted in a pitch reversal and the latter stimulus resulted in a correct pitch order [the arrows in Fig. 6(A)]. When comparing $F0$'s 133- and 150 Hz, the position of the fundamental frequency moves from the first to the second channel (towards the base), but the peak caused by the harmonics under the formant frequency shifts downward from channel 7 to channel 6, towards the apex. So, the pitch percept was not determined by the place of the fundamental frequency, but was largely determined by the place of the largest excitation along the electrode array. In the case where the channels with the largest amplitude coincide, as in Fig. 6(B) for $F0$ 133 Hz (thick solid line) and $F0$ 250 Hz (thin

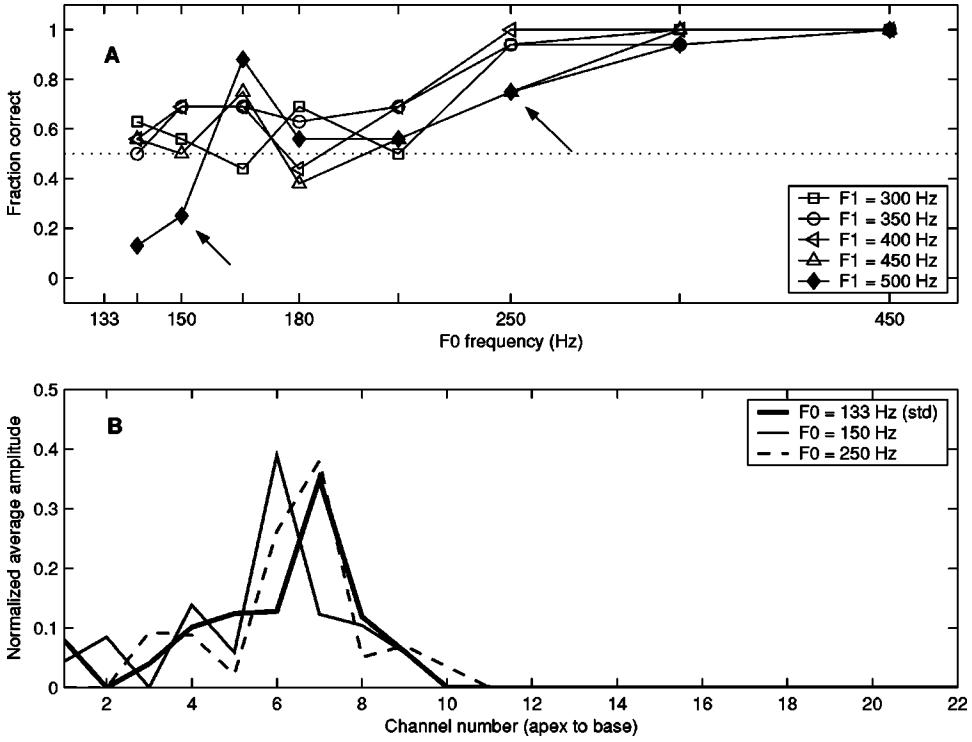

FIG. 6. Top panel: The proportion of correctly pitch-ranked stimuli pooled over all subjects for the condition with reference $F0$ 133 Hz and stimuli processed with the GTF filter bank and all temporal modulations above 10 Hz removed. The different lines in the figure depict the results for the various formant frequencies used. The fundamental frequencies of the signals are plotted on the horizontal axis on a log scale. A significant pitch reversal is noticed for the case where the formant frequency is 500 Hz (filled symbols) for the two lowest tested $F0$ frequencies (140 and 150 Hz). Bottom panel: The average channel amplitudes for the single-formant stimuli with formant frequency 500 Hz and fundamental frequencies 133, 150, and 250 Hz when processed with the GTF filter bank and with all temporal modulations above 10 Hz removed. The thick line represents the average amplitude of the reference stimulus with $F0$ 133Hz, $F1$ 500 Hz, and processed with the GTF filter bank (i.e., the reference stimulus of the line with the filled symbols in the upper panel). The thin solid line represents the average channel amplitude of the comparison stimulus mentioned in the upper panel and causing a pitch reversal ($F0=150$ Hz). The peak of the average amplitude has shifted towards the apex with respect to the average amplitude of the reference stimulus (thick line), causing the comparison stimulus to be perceived at a lower pitch. The thin dotted line represents the average amplitude of a comparison stimulus correctly ranked in pitch ($F0=250$ Hz). Although the peak of the average amplitudes of the reference and comparison stimulus are at the same location, most probably the correct ranking is due to the shift of the fainter channels towards the base.

dashed line), the pitch difference was most probably mediated by the excitation shift along the electrode array of the smaller peaks. Summarizing the two previous effects, we hypothesize that the pitch of the different sounds was determined by a weighted position of the peak of the excitation pattern. The simplest form of a weighted peak position was the center of gravity or the centroid, and can be expressed as

$$C=\frac{\Sigma_{e=1}^{22} e.m(e)}{\Sigma_{e=1}^{22} m(e)} \tag{4}$$

In Eq. (4) C is the centroid, e indicates the channel number (ranked from apex to base), and $m(e)$ is the average amplitude on channel e. The centroid is computed based upon the average amplitude of the envelopes (the compression accommodating for loudness growth is not applied). According to the proposed model, the discriminability of two sounds (based upon place pitch cues) was determined by the distance between the centroids of the excitation patterns of the two sounds. The centroid has the additional benefit of being independent of overall intensity. Figure 7 shows a scatter plot of the model prediction of all stimuli and the respective average proportion of correctly ranked trials, for $F0$'s of 133 Hz (top panel) and 165 Hz (bottom panel). The solid line in Fig. 7 represents the normal cumulative distribution fit to the pooled data over all conditions and subjects. This function was fitted to the average proportion using a Gauss–Newton nonlinear least-squares fitting method.

Similarly, a normal cumulative distribution function was fitted to the scatter plots for every subject separately. The slope of this fitted normal cumulative distribution function was taken as a performance measure of place pitch sensitivity. This fitting procedure is similar to a linear regression with the data points converted into d' values. Just-noticeable differences (jnd) were derived from the fitted curves as the minimal distance expressed in number of electrodes to obtain 75%-correct pitch rankings.

Figure 8 depicts the jnd estimates that were obtained from these fitted curves for every subject. The place pitch sensitivity of the subjects was also measured using an electrode ranking experiment in a separate study (Laneau and Wouters, 2004). These jnd estimates are also depicted in Fig. 8. In the latter study the subjects were presented two stimuli that differed in site of stimulation along the electrode array, and had to indicate the highest one. The jnd's shown in Fig. 8 are the mean of the jnd's for conditions with different numbers of active channels. No significant effect was found

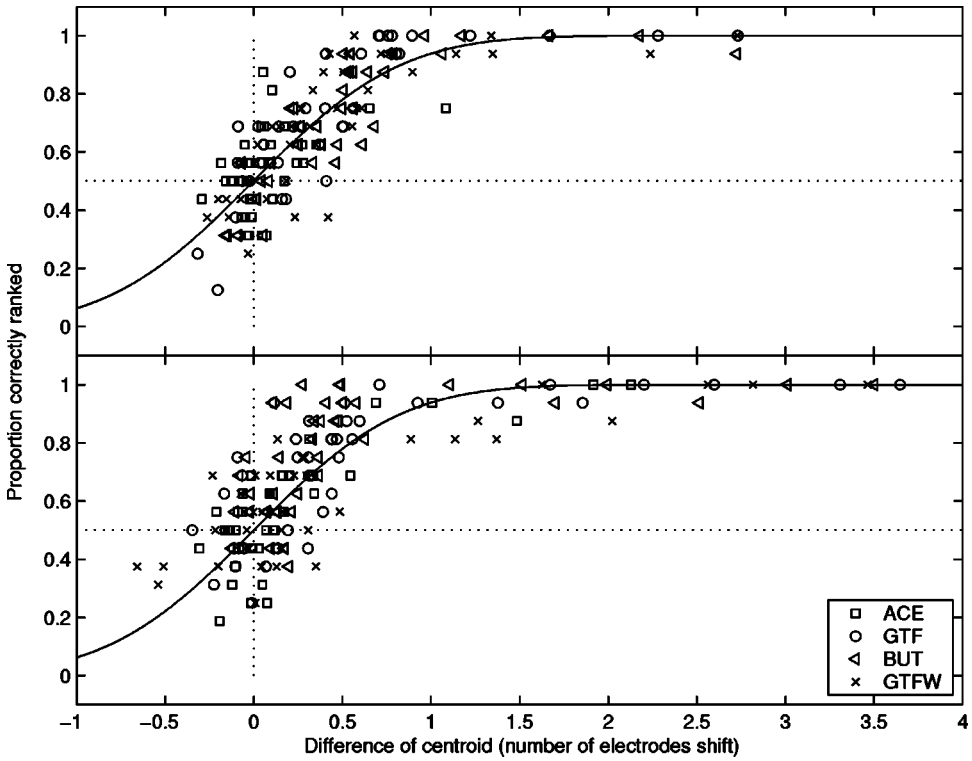

FIG. 7. The proportion of correctly ranked stimuli pooled over all subjects plotted as a function of the difference in centroid. Only data of stimuli that allowed no temporal pitch cues were included in this picture. For the sake of clarity the data for the two reference $F0$'s are depicted in separate panels ($F0$ 133 Hz in the upper panel and $F0$ 165 Hz in the lower panel). A normal cumulative distribution function was fitted to the pooled data and is represented by the solid line in both panels.

of the number of active channels of the stimuli on place pitch sensitivity in the latter study.

The jnd's estimated from the $F0$ discrimination data and the jnd's estimated from the place pitch measurements are nearly equal for all subjects (within error bars). However, for subject S3, the place pitch sensitivity is somewhat worse when compared to the centroid sensitivity. This can be due to the fact that we arbitrarily choose two electrode locations to estimate the place pitch sensitivity. At other locations the place pitch sensitivity for subject S3 might be better.

Previous support for the centroid model has been presented by McDermott and McKay (1994) for the case of two stimulation channels in cochlear implants. They investigated the pitch perceived when two bipolar channels were stimulated simultaneously and interleaved using constant current levels. The pitch of the two-channel stimuli varied monotonically between the pitches of the stimuli on the two single electrodes, as the relative current amplitudes were changed in orderly fashion. They interpreted their results by stating that the centroid of the excitation pattern shifted from one electrode location to the other as relative currents varied, and that this centroid shift caused the varying pitch. Their results also indicated that the pitch of two interleaved stimulated channels was perceived as one-dimensional, at least up to separations of 3 mm between stimulating electrodes. McKay *et al.* (1996) later confirmed and extended this result using dissimilarity matrices and multidimensional scaling analysis. They concluded that there exist two basic perceptual dimen-

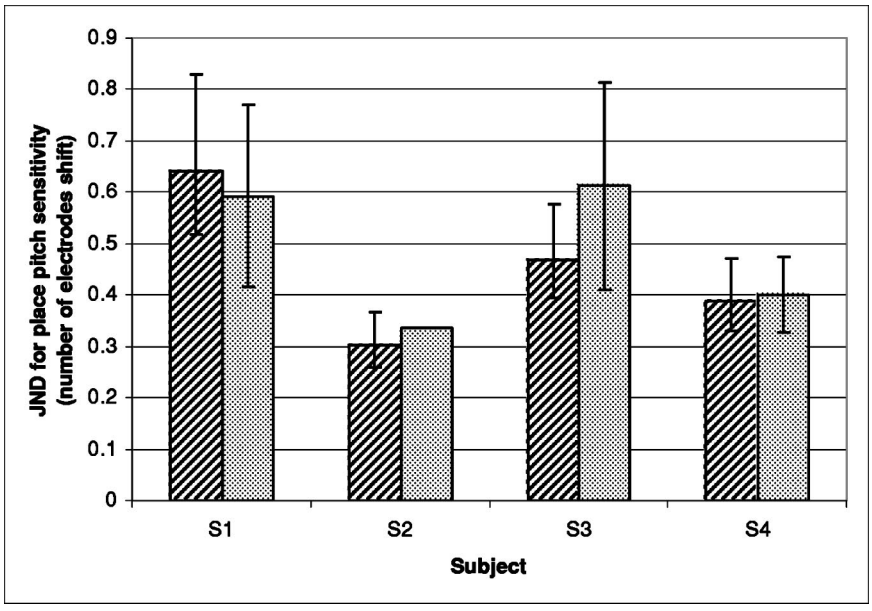

FIG. 8. Estimated just-noticeable differences (jnd's) for the four subjects. The dashed bars present the jnd's from the centroid differences calculated from the $F0$-discrimination data. The jnd's estimated from site of stimulation discrimination of the multichannel stimuli in the same patients (Laneau and Wouters, 2004) are depicted by means of the dotted bars. The error bars indicate the 95% confidence interval on the jnd estimated given by the fitting routine for the centroid jnd's. The error bars on the place pitch sensitivity of the multielectrode stimuli indicate ± 1 standard deviation.

sions with two-channel stimulation. The first percept was associated with the middle of the two active channels, corresponding to the centroid of the excitation pattern. The second percept was associated with the separation of the two active channels. However, McKay *et al.* also argued that their experiment did not rule out that the dimensions could also be interpreted as simply being associated with the location of the two channels. The results of McKay *et al.* (1996) were consistent for separations up to 12 mm, the largest separation tested, between active channels. Due to the nature of the acoustic stimuli used in the present study and the maxima selection in the sound-processing stage, the pulse sequences only stimulated channels that were at most ten electrodes apart, resulting in a maximal separation of 7.5 mm between active channels.

Cohen *et al.* (1996) examined the direct relationship between the forward-masking pattern and the numerical pitch estimates that single-channel stimuli exert. They found a clear qualitative correspondence between the centroid of the forward-masking pattern and the numerical pitch estimates, for all stimulation modes. They were also able to predict the reversed (compared to normal tonotopy) pitch ranking of some pairs of electrodes based upon the centroid of the forward-masking pattern for several subjects.

In normal-hearing subjects, models based upon the centroid of the spectrum (Stover and Feth, 1983; Anantharaman *et al.*, 1993) have been used to explain the results of pitch ranking and discrimination experiments with noise bands (Dai *et al.*, 1996; Versfeld, 1997). These models accurately described the discriminability of narrow-band sounds, but the models had to be extended for the discrimination of sounds with larger bandwidths.

It is, however, questionable whether the place pitch is truly a pitch percept or an aspect of the sounds' timbre. According to the ANSI definition of pitch (American National Standards Institute, 1994), place pitch is pitch as subjects are able to rank stimuli from high to low. Nevertheless, there is a striking resemblance between the proposed model for place pitch and the model for sharpness in normal-hearing subjects (Zwicker and Fastl, 1999). Both models calculate the centroid of the excitation pattern.

The results of the present study, and the studies of Geurts and Wouters (2001; 2004) have shown that $F0$ discrimination in cochlear implant recipients is possible solely based upon place pitch cues. These studies used stationary synthetic signals while running speech is nonstationary and formant frequencies change over time. Green *et al.* (2002) measured the discrimination of direction of sweeps of fundamental frequency using noise-band vocoders in normal-hearing subjects. They showed that for stationary signals the discrimination was possible based solely on place pitch cues, but the discrimination became almost impossible when the signals had nonstationary formant frequencies. Moreover, the model presented above to account for the $F0$-discrimination results based upon place pitch cues indicates a strong dependence of the place pitch cues on the overall spectral shape of the stimuli. This might limit the usefulness of place pitch cues to encode $F0$ in cochlear implants during running speech.

On the other hand, although formant frequencies and $F0$ frequencies originate from physically independent processes, formant frequencies and $F0$ frequencies are often correlated. For example, both the formant frequencies and the $F0$ frequencies of a female speaker are on average higher than those of a male speaker. The combined effect of higher formant frequencies and higher fundamental frequencies will shift the centroid of the excitation pattern toward the base. This shift might be detectable for cochlear implant subjects even during running speech. This might help the subject listening to running speech to discriminate pitch differences.

Moreover, the possible problem for $F0$ discrimination based upon place pitch cues coming from the interaction between overall spectral shape and $F0$ cues is less likely to exist for some musical sounds, such as woodwinds, because these musical notes have stationary spectral shapes throughout their duration (Gray, 1977). Second, the dimensions and form of the instrument, defining the resonance frequencies, are fixed so the spectral shape is fixed across different notes. Consequently, the differences in place pitch between two different notes are related to $F0$ differences.

B. Including temporal pitch cues

Comparing Figs. 3 and 4, the improvement in performance from adding the temporal pitch cues was largest for the ACE and GTFW filter banks. However, the range for improvement was much larger for the ACE filter bank than for the other filter banks because the latter already enabled some $F0$ discrimination without temporal pitch cues.

The limited improvement obtained by adding temporal pitch for the GTF and BUT filter bank might have been due to the fact that higher harmonics of the test stimuli is also resolved in those filter banks. Resolved harmonics prevented the output of filters from beating at the fundamental frequency, and the only fluctuation seen at the output was at the frequency of the respective harmonic. The filter outputs with the largest amplitude were filters having center frequencies close to the formant frequencies (between 300 and 500 Hz). In this range the BUT and GTF filter banks resolved the individual harmonics, so the largest output of the filters fluctuated at multiples of the fundamental frequency. As an example, the modulations of channels 1 to 7 of one of the sounds processed with the BUT filter bank and after envelope detection are depicted in Fig. 9. The most intense channels in this example do not beat at $F0$ frequency but at some higher harmonic frequency. Although this higher frequency fluctuation was partly attenuated by the envelope detection low-pass filter, it is still clearly present in the channel output. Only the output of the lower channel, containing the fundamental frequency, and possibly higher channels, containing unresolved harmonics, provide the temporal pitch cues related to $F0$. The channels containing resolved higher harmonics are modulated at multiples of $F0$ and consequently elicit higher pitches because higher modulation frequencies lead to higher pitches (McKay *et al.*, 1994).

The ACE filter bank did not suffer from this shortcoming because all the bins in the FFT filter bank are approximately 250 Hz wide, so all harmonics was unresolved (at least for the reference stimuli), causing all filter outputs to

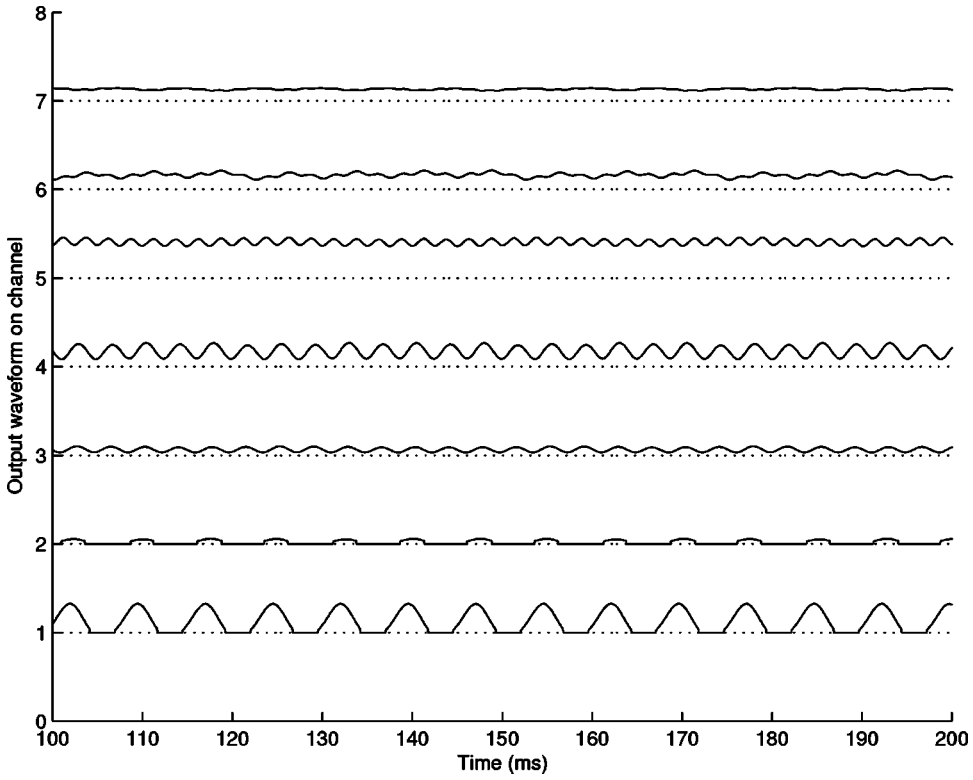

FIG. 9. Time fragment of the amplitudes of channels 1 to 7 across time for the SFS stimulus with $F0=133$ Hz and $F1=400$ Hz, processed with the BUT filter bank. Channels 1 and 2 only contain the first harmonic, and consequently the outputs beat at the fundamental frequency. Channels 3 and 4 only contain the second harmonic, leading to a modulation at the second harmonic frequency. Channel 5 contains only the third harmonic, and consequently the output beats at the third harmonic frequency. The beating at higher harmonics makes the temporal pitch cues less clear.

beat at the fundamental frequency. Similarly, the GTFW filter bank was designed to only resolve the fundamental frequency and with all overtones unresolved. This also made the output of all channels beat at $F0$ for the GTFW filter bank.

In order to assess the effect of the modulation depth of the beatings at $F0$ in the different channels, the average modulation depths of all stimuli were computed. We also assess the effect of the phase differences of the beating across channels. These interchannel phase differences of the beating at $F0$ are caused by the phase differences between the different harmonics of the original sound and by the phase response across frequency of the filters in the filter bank. The average modulation depth over all channels is computed in two different ways to assess the effect of the phase differences. In the first way, all phase information of the channels was discarded (''without phase''). In the second way the phase information was included and contributed to the average as out-of-phase channels could interfere destructively (''with phase'').

As a first step, the average spectrum of the modulations in all channels was calculated. For the ''without phase'' approach, the spectrum of all channels was calculated separately and the average of the magnitude of the spectrum over the channels was taken, discarding the phase components of the modulations. In the ''with phase'' approach, the time signal was averaged over all channels and then the magnitude of the spectrum of this average time signal was calculated. In the second step, the modulation depth was calculated for both approaches from the respective average spectrum, using Eq. (5)

$$\text{modulation depth} = \frac{2*\|S(F0)\|}{\|S(F0)\|+\|S(0)\|}, \tag{5}$$

where $\|S(F0)\|$ was the magnitude of the spectrum at the fundamental frequency and $\|S(0)\|$ was the magnitude of the dc component of the spectrum.

As a third step, the obtained modulation depths were correlated with the amount of temporal pitch cues (d'_{temp}) for all processed reference stimuli and all test conditions. The correlation was done only with the modulation depth of the reference stimuli and not with the modulation depth of the higher $F0$ comparison stimuli, because of the test procedure used in the present experiment. The subjects were asked to indicate the highest-sounding stimulus and this could be the one with a higher modulation rate or with no modulation at all. This is because for modulated stimuli the perceived pitch is related to the modulation frequency but steady-state (or poorly modulated) stimuli elicit high pitches related to the high overall pulse rate (McKay and Carlyon, 1999). So, clear pitch differences may have been perceived when the comparison contained no modulations at all.

The analysis revealed that the d'_{temp} values are significantly ($p<0.0001$ for both) correlated with the modulation depth of the reference stimulus of the trials for both computational methods ($r=0.4216$; and $r=0.3090$ for with phase and without phase, respectively). Moreover, although both methods of calculating the modulation depth are significantly correlated ($r=0.5910$) with each other, the correlation with benefit from inclusion of temporal pitch cues was 36% higher when phase information was included. Using a test comparing the two sets of predictors (Tabachnick and Fidell, 1996), this lead to a significant higher correlation for the ''with phase'' approach ($Z=2.42$; $p=0.0077$). However, both correlations are rather weak and the results of this analysis should be interpreted with some caution.

The poorest-performing filter bank for temporal pitch

cues (BUT) produced the lowest modulation depths averaged over all reference stimuli. The best-performing filter bank for temporal pitch cues (ACE) produced the highest average modulation depth for the reference stimuli, when modulation depths are calculated taking into account the relative phase differences between the channels. When the phase information is discarded, GTF produced the lowest modulation depths and GTFW with the highest modulation depths. So, the ''with phase'' modulation depth present in the processed stimuli can be a predictor of the amount of temporal pitch cues that can be used by the cochlear implantee.

This may mean that out-of-phase modulations on different channels probably allow for less clear pitch percepts than equally large in-phase modulations, at least when the channels are within a relatively narrow region, as is the case in this study. It has been shown before that modulation patterns on two nearby electrodes are integrated, while more widely separated modulation patterns are perceived separately (McKay and McDermott, 1996).

V. CONCLUSIONS

In the present study, properties of filter banks which affect $F0$ discrimination in cochlear implants were examined. Temporal and place pitch cues to $F0$ discrimination were studied separately and for both cues the mechanisms relating the pitch percepts to physical parameters were investigated.

The results of our experiments demonstrate that both place and temporal pitch cues allow the discrimination of the fundamental frequency, as also evidenced by previous research. However, $F0$ discrimination based upon purely place pitch cues is weak and only possible for differences exceeding one octave for the best-performing filter banks. Moreover, the ACE filter bank, as used by most of the Nucleus cochlear implant recipients in their daily life, did not allow for $F0$ discrimination based upon place pitch cues. Our experiments indicate that our new filter banks having the highest frequency resolution in the $F0$ register gave the best results using place pitch cues. The newly proposed filter banks are likely to improve $F0$ discrimination, relative to the currently clinically available filter bank, in subjects having good place sensitivity and poor temporal sensitivity, at least for spectrally stationary signals. A previous study has shown that for some cochlear implantees the place pitch cues are very important for $F0$ discrimination (Geurts and Wouters, 2004).

When temporal pitch cues were added to the stimuli, performance increased significantly and jnd's for $F0$ discrimination ranged from 6% up to 60% depending on subject, processing condition, and reference $F0$. There were no overall significant differences between the performance of the filter banks when both temporal and place pitch cues were presented to the subjects. The performance of cochlear implant subjects is still very poor compared to the performance of normal-hearing subjects where jnd's for $F0$ are typically less than 1%. The largest temporal pitch cues for $F0$ discrimination were possible with the filter banks that did not resolve the harmonics. In these filter banks all output channels beat at the fundamental frequency. When the beatings of the different channels were in phase, the temporal pitch cues seemed to be clearer compared to the case where beatings of neighboring channels are out of phase.

A modeling study of the results with purely place pitch cues indicated that place pitch is related to the centroid of the excitation pattern along the cochlea. As place pitch cues enable subjects to rank stimuli from low to high, it can be considered a pitch percept. However, taking into account the present model, place pitch is more related to the sharpness or brightness of timbre than to the repetition rate of the processed sound in cochlear implants.

The proposed model indicates that the place-pitch-mediated $F0$ cues and the overall spectral shape are closely related. In other words, there might be interference in the perception between $F0$ and formant frequencies when both are coded onto place of stimulation. This might possibly lead to a reduction of the usefulness of the place-pitch-mediated $F0$ cues when formants vary as in running speech, and might possibly also lead to reduced speech perception when $F0$ strongly varies. Although the spectral shape or brightness of timbre also interferes with pitch perception in normal-hearing subjects (Ohgushi, 1978), the effect in cochlear implants appears to be larger, at least when only place pitch cues are present.

With the filter banks studied, optimal transmission of both temporal and place pitch cues was difficult for one and the same filter bank.

ACKNOWLEDGMENTS

We thank the subjects for their enthusiastic cooperation. We also thank Luc Geurts and Astrid van Wieringen for helpful comments. We thank Robert Carlyon, Andrew Faulkner, and Robert Shannon for comments on earlier versions of this manuscript. This study was partly supported by the Flemish Institute for the Promotion of Scientific-Technological Research in Industry (Project IWT 020540), by the Fund for Scientific Research-Flanders/Belgium (Project G.0233.01), and by Cochlear Ltd.

American National Standards Institute (**1994**). ''American National Standard Acoustical Terminology for physiological and psychological acoustics.'' ANSI S1.1-1994 (American Standards Association, New York).

Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (**1993**). ''Intensity-weighted average of instantaneous frequency as a model for frequency discrimination,'' J. Acoust. Soc. Am. **94**, 723–729.

Barry, J. G., Blamey, P. J., Martin, L. F., Lee, K. Y., Tang, T., Ming, Y. Y., and Van Hasselt, C. A. (**2002**). ''Tone discrimination in Cantonese-speaking children using a cochlear implant,'' Clin. Linguist Phon. **16**, 79–99.

Cochlear Ltd. (**2002**), ''Nucleus Implant Communicator (NIC) System Overview,'' **N95291 Iss.1**

Cohen, L. T., Busby, P. A., and Clark, G. M. (**1996**). ''Cochlear implant place psychophysics. II. Comparison of forward masking and pitch estimation data,'' Audiol. Neuro-Otol. **1**, 278–292.

Dai, H. P., Nguyen, Q., Kidd, G., Feth, L. L., and Green, D. M. (**1996**). ''Phase independence of pitch produced by narrow-band sounds,'' J. Acoust. Soc. Am. **100**, 2349–2351.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.

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.

Geurts, L., and Wouters, J. (**2004**). ''Better place-coding of the fundamental frequency in cochlear implants,'' J. Acoust. Soc. Am. **115**, 844–852.

Gfeller, K., and Lansing, C. (**1992**). ''Musical perception of cochlear implant users as measured by the primary measures of music audiation—An item analysis,'' J. Music. Ther. **29**, 18–39.
Glasberg, B. R., and Moore, B. C. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Green, T., Faulkner, A., and Rosen, S. (**2002**). ''Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,'' J. Acoust. Soc. Am. **112**, 2155–2164.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Grey, J. M. (**1977**). ''Multidimensional perceptual scaling of musical timbres,'' J. Acoust. Soc. Am. **61**, 1270–1277.
Jones, P. A., McDermott, H. J., Seligman, P. M., and Millar, J. B. (**1995**). ''Coding of voice source information in the Nucleus cochlear implant system,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 363–365.
Klatt, D. H. (**1980**). ''Software for a cascade-parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.
Laneau, J., and Wouters, J. (**2004**). ''Multi-channel place pitch sensitivity in cochlear implant recipients,'' J. Assoc. Res. Otolaryngol. **5**, 285–294.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).
McDermott, H. J., and McKay, C. M. (**1994**). ''Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea,'' J. Acoust. Soc. Am. **96**, 155–162.
McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.
McKay, C. M., and McDermott, H. J. (**1996**). ''The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites,'' J. Acoust. Soc. Am. **100**, 1081–1092.
McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). ''Place and temporal cues in pitch perception: Are they truly independent?'' ARLO **1**, 25–30.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). ''Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees,'' J. Acoust. Soc. Am. **96**, 2664–2673.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1996**). ''The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees,'' J. Acoust. Soc. Am. **99**, 1079–1090.
Moore, B. C. J., and Glasberg, B. R. (**1990**). ''Frequency discrimination of complex tones with overlapping and nonoverlapping harmonics,'' J. Acoust. Soc. Am. **87**, 2163–2177.
Ohgushi, K. (**1978**). ''On the role of spatial and temporal cues in the perception of the pitch of complex tones,'' J. Acoust. Soc. Am. **64**, 764–771.
Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Pijl, S. (**1997**). ''Labeling of musical interval size by cochlear implant patients and normally hearing subjects,'' Ear Hear. **18**, 364–372.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). ''Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Slaney, M. (**1993**), ''An efficient implementation of the Patterson–Holdsworth auditory filter bank,'' Tech. Rep. **35**, Apple Computer.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Stover, L. J., and Feth, L. L. (**1983**). ''Pitch of narrow-band signals,'' J. Acoust. Soc. Am. **73**, 1701–1707.
Tabachnick, B. G., and Fidell, L. S. (**1996**). *Using Multivariate Statistics*, 3rd ed. (Harper Collins College Publishers, Northridge).
Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). ''Psychophysical studies for 2 multiple-channel cochlear implant patients,'' J. Acoust. Soc. Am. **71**, 153–160.
Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.
Versfeld, N. J. (**1997**). ''Discrimination of changes in the spectral shape of noise bands,'' J. Acoust. Soc. Am. **102**, 2264–2275.
Wilson, B. S. (**1997**). ''The future of cochlear implants,'' Br. J. Audiol. **31**, 205–225.
Wouters, J., Damman, W., and Bosman, A. J. (**1994**). ''Vlaamse opname van woordenlijsten voor spraakaudiometrie,'' ''Flemish recording of the wordlists for speech audiometry.'' Logopedie **7**, 28–33.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Information Sciences, Springer, Berlin).