Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants

Johan Laneau[a)]
Laboratory for Experimental ORL, K.U.Leuven, Kapucijnenvoer 33, B 3000 Leuven, Belgium

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

Jan Wouters
Laboratory for Experimental ORL, K.U.Leuven, Kapucijnenvoer 33, B 3000 Leuven, Belgium

(Received 30 June 2004; revised 10 October 2005; accepted 11 October 2005)

Although in a number of experiments noise-band vocoders have been shown to provide acoustic models for speech perception in cochlear implants (CI), the present study assesses in four experiments whether and under what limitations noise-band vocoders can be used as an acoustic model for pitch perception in CI. The first two experiments examine the effect of spectral smearing on simulated electrode discrimination and fundamental frequency (F0) discrimination. The third experiment assesses the effect of spectral mismatch in an F0-discrimination task with two different vocoders. The fourth experiment investigates the effect of amplitude compression on modulation rate discrimination. For each experiment, the results obtained from normal-hearing subjects presented with vocoded stimuli are compared to results obtained directly from CI recipients. The results show that place pitch sensitivity drops with increased spectral smearing and that place pitch cues for multi-channel stimuli can adequately be mimicked when the discriminability of adjacent channels is adjusted by varying the spectral slopes to match that of CI subjects. The results also indicate that temporal pitch sensitivity is limited for noise-band carriers with low center frequencies and that the absence of a compression function in the vocoder might alter the saliency of the temporal pitch cues. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2133391]

PACS number(s): 43.66.Hg, 43.66.Ts, 43.66.Fe [RAL] Pages: 491–506

I. INTRODUCTION

Previous studies have shown that noise-band vocoders can be used as an acoustic model for speech perception in cochlear implants (Blamey *et al.*, 1984b; Dorman and Loizou, 1997, 1998; Friesen *et al.*, 2001; Nelson *et al.*, 2003). These studies have indicated that at least good qualitative agreement exists between the model results and the results obtained from cochlear implant (CI) subjects as long as the number of channels is below eight. More recently, several studies have used noise-band vocoders to study pitch perception in CI (Green *et al.*, 2002) or pitch-related tasks (Fu *et al.*, 1998; Faulkner *et al.*, 2000; Xu *et al.*, 2002; Kong *et al.*, 2004). However, it is generally unknown whether noise-band vocoders provide a valid acoustic model to study pitch perception in CI. It has been shown that speech perception relies on different mechanisms than pitch perception (Smith *et al.*, 2002). Smith *et al.* showed that while speech perception is mostly based upon envelope information in broad frequency bands, pitch perception is based upon the fine structure within every frequency band. The principal goal of the present study was to assess the validity of noise-band vocoders as acoustic models for pitch perception in CI subjects. Valid models are defined here as models for which the average pitch discrimination performance does not differ significantly from the average performance obtained from CI subjects on any of a number of different tasks. In order to find such a model, pitch discrimination results from post-lingual CI subjects are compared to results from normal-hearing (NH) subjects listening to noise-band vocoder processed signals for different parameters of the noise-band vocoder.

There are several benefits of using acoustic models in parallel to performing experiments directly with cochlear implant recipients. First, the results obtained from cochlear implant patients are generally more prone to inter- and intra-subject variability (Friesen *et al.*, 2001). Possible reasons for this higher variability might be the different etiologies of the patients, differences in duration of auditory deprivation leading to various amounts of degeneration of the auditory functions, and the high variability in peripheral processes, such as neural excitation profiles, between subjects and even within subjects. A second benefit of using acoustic models is the fact that in acoustic models most of the parameters can be changed independently, whereas in cochlear implant patients some these factors are fixed, such as width of excitation pattern or implant insertion depth. Third, comparing the results of NH subjects using acoustic models, on the one hand, and cochlear implant patients, on the other hand, reveals some details of the basic functioning of the auditory system. Basic mechanisms of the auditory system can be more solidly founded when the results in electrical and acoustical hearing reveal the same trends (Carlyon *et al.*, 2002). Fourth, com-

[a)]Electronic mail: johan.laneau@philips.com

paring results obtained through acoustic models and results obtained directly from CI recipients also allows indicating the cause of limitations in CI recipients' performance. Finally, there is a practical benefit to acoustic models because the pool of NH subjects is vastly more extended than the pool of CI subjects, and research interfaces to stimulate acoustically are more easily available than research interfaces to stimulate implants directly.

Previously several acoustic models have been developed and used to study pitch perception in cochlear implants through NH subjects. First, noise-band vocoders are most often used (Fu *et al.*, 1998; Faulkner *et al.*, 2000; Green *et al.*, 2002; Xu *et al.*, 2002; Kong *et al.*, 2004) because temporal and place pitch cues appear to be accurately modeled. Modulation of noise carriers elicits weak purely temporal pitch percepts that correspond to the modulation frequency (Burns and Viemeister, 1976). Moreover, the difference limens in rate discrimination for CI resemble the modulation rate discrimination of noise carriers as a function of frequency (Shannon, 1983a; Blamey *et al.*, 1984a). Furthermore, the place pitch cues can be modeled by changing the center frequency of the noise band because the pitch of a steady-state noise band corresponds approximately to its center frequency (Stover and Feth, 1983; Dai *et al.*, 1996; Zwicker and Fastl, 1999).

Whereas noise-band vocoder signals sound very stochastic, CI subjects never define the perceived sound as noisy but rather qualify the perceived sound of single channels as beeps. To accommodate for this fact a second acoustic model was developed that uses pure tones as carriers for the different channels in the resynthesis of the vocoder (Dorman *et al.*, 1997). More recently, such a vocoder was used to study the pitch perception through discrimination of lexical tones for two different sound processing schemes (Lan *et al.*, 2004). The sinusoidal vocoder might be less accurate in modeling pitch perception in CI because adjacent channels are extremely well discriminable in the model while in general CI subjects can have problems discriminating adjacent channels (Nelson *et al.*, 1995). Furthermore, amplitude modulation of a pure-tone carrier generates sidebands which can be detected when the sidebands are spaced wide enough to be resolved by the peripheral filters of the normal ear (Kohlrausch *et al.*, 2000), while this is impossible in CI subjects.

A third acoustic model was proposed to study the purely temporal pitch percepts and uses harmonic complexes that are band-pass filtered in relatively high-frequency regions so that all harmonics are unresolved (McKay and Carlyon, 1999; Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003; Deeks and Carlyon, 2004). Using this scheme place and temporal pitch cues can be adjusted independently by varying the center frequency of the band-pass filter and the fundamental frequency of the harmonic complex, respectively. However, the model is restricted to lower pulse rates because for higher pulse rates the harmonics become resolved.

The present study examines the factors that affect the use of a noise-band vocoder as an acoustic model of pitch perception in CI. Four experiments are described. The first two experiments examine the effect of spectral smearing on pitch perception of noise-band vocoded signals in NH subjects. These experiments focus on spectral or mimicked place pitch cues. The amount of spectral smearing is changed by varying the spectral slope of the noise bands in the vocoder. The first experiment (experiment 1) assesses center frequency discrimination for noise bands, simulating electrode discrimination. The second experiment (experiment 2) measures fundamental frequency (F0) discrimination. In the third experiment (experiment 3) F0-discrimination results for postlingually deafened CI subjects from Laneau *et al.* (2004) are compared to F0-discrimination results for NH subjects for two different vocoders to assess the effect of spectral mismatch. Both temporal and place pitch cues are evaluated in this experiment. In the last experiment (experiment 4) the effect of the absence of amplitude compression in the vocoder is assessed on modulation rate discrimination, focusing on temporal pitch discrimination.

II. EXPERIMENT 1: THE EFFECT OF SPECTRAL SMEARING ON SIMULATED ELECTRODE DISCRIMINATION

A. Subjects

Five NH subjects (S1–S5) aged between 24 and 39 participated in this and all further experiments of the presented study. All subjects were members of the departmental staff and one subject was the first author of this paper. All subjects had experience in psychophysical studies with similar stimuli.

The results of the NH subjects will be compared to the results of four adult postlingually deafened CI subjects that performed equivalent tasks. The results of the CI subjects have been reported in Laneau and Wouters (2004). All CI subjects were implanted with the Nucleus CI24 implant and were relatively good performers. Some of the subjects' relevant details can be found in Table I. Although some subjects had relatively short implant experience, their pitch discrimination performance was expected to asymptote because they already participated in previous psychophysical studies assessing pitch perception including training sessions.

B. Presentation

In this and all further experiments of the present study, the acoustical stimuli were routed via a desktop PC and a 24-bit PCI sound card (Lynx ONE, Lynx Studio Technology) to a mixer-amplifier (Eurorack MX1604A) and played to the right earpiece of a TDH39 headphone. The APEX program controlled stimulus playback and response collection (Geurts and Wouters, 2000). The attenuated intensity of the stimuli was adjusted to be at a comfortable level, which in most cases was approximately 70 dB SPL. The subjects were seated in a double-walled soundproof booth.

C. Methods

Simulated place pitch discrimination was measured using the acoustical analogy of electrode discrimination based on a noise-band vocoder: center frequency discrimination for

TABLE I. Relevant information about each of the CI subjects who participated in the experiments.

Subject	Duration of profound deafness (years)	Etiology	Speech processor	Clinical speech processing strategy	Age (years)	Implant experience (year;month)
CI1	11	Hereditary	ESPrit	Speak	19	1;11
CI2	3	Unknown	ESPrit	Speak	21	3;2
CI3	30	Progressive	SPrint	ACE	54	0;8
CI4	32	Progressive	ESPrit 3G	ACE	49	0;6

noise bands. The center frequency discrimination was assessed as a function of the slope of the noise bands and at two different reference center frequencies.

Stimuli consisted of white noise filtered with a custom designed filter. This filter (CISIM filter) is loosely based upon the physiology of electrical stimulation through a cochlear implant. The center frequency of the filter is set to the frequency corresponding to the cochlear location of a simulated electrode. The shape of the frequency response of the filter is designed to mimic the exponential decay of current density along the basilar membrane (Black and Clark, 1980). The magnitude of the desired frequency response of the CISIM filter was thus set to decay exponentially with increasing distance between the characteristic point corresponding to the frequency and the position of the stimulating electrode. The space constant of the exponential decay (λ) and the position of the simulated electrode ($x_{\text{electrode}}$) are two free parameters. Consequently the desired frequency response of the CISIM filter is given by

$$F_{\text{CISIM}}(x(f)) = \exp\left(\frac{-|x_{\text{electrode}} - x(f)|}{\lambda}\right). \quad (1)$$

The conversion of distance along the cochlea to frequency $x(f)$ was done using Greenwood's formula (1990). The desired frequency response was calculated for 400 frequencies spaced equidistantly from 1 to 34 mm, assuming a 35-mm cochlea. The obtained desired frequency response was approximated using a 150-tap linear phase FIR filter using the FIR2 command in Matlab.

Center frequency discrimination was measured at two reference center frequencies 565 and 2643 Hz (corresponding to cochlear locations of 10.75 and 20.5 mm; cochlear length is expressed from apex to base and assuming a 35-mm cochlea), and for ten different space constants: $\lambda = 0.5$, 0.67, 1, 1.33, 2, 2.5, 3.33, 5, 10, and 20 mm. Stimuli were 500 ms long including 10-ms linear on- and off-ramps and were sampled at 20 kHz.

Discrimination was measured using a two-interval two-alternative forced choice (2I-2AFC) constant stimuli paradigm. The subjects were presented with a reference and a comparison stimulus in random order, and they were asked to indicate the highest in pitch. The center frequency of the reference stimulus corresponded to one of the two reference center frequencies tested. The center frequency of the comparison stimulus was higher than that of the reference stimulus, corresponding to a shift towards the base of the cochlea. For space constants (λ) of 2 mm up to 20 mm three comparison stimuli were used with cochlear locations (corresponding to their center frequencies) shifted 0.75, 1.5, or 2.25 mm towards the base from the cochlear location of the reference stimulus. This corresponds to shifts of one, two, or three electrodes in the Nucleus CI24 device. For space constants (λ) of 0.5 up to 1.33 mm, five comparison stimuli were used and the shifts were 0.1875, 0.375, 0.5625, 0.75, or 1.5 mm towards the base. This corresponds to shifts of $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1, or 2 electrodes in the Nucleus CI24 device. Note that a shift of 0.75 mm corresponds to a relative change of center frequency of approximately 12% and that the relative center frequency is roughly proportional to the shift. The intensity of the stimuli was randomly attenuated over a 20-dB range to minimize loudness cues. No feedback was presented to the subjects.

The trials were presented in blocks. Each block contained trials for a given space constant and for both reference center frequencies. Every trial was repeated ten times per block, and this led to 60 or 100 trials per block for space constants greater or smaller than 2 mm, respectively. Every block was presented twice to the subjects; hence every subject was presented 20 blocks.

Just noticeable differences (jnd's) of simulated electrode position were estimated for each combination of reference center frequency and space constant. Using a nonlinear least squares fitting routine, a normal cumulative distribution function was fitted to the average proportion of trials where the comparison stimulus was indicated as higher than the reference stimulus as a function of distance between cochlear locations of comparison and reference stimulus. The jnd was set to the distance that resulted in a proportion of 75% correct according to the fitted curve. If the listener could not perform the task, the discrimination threshold was arbitrarily set to 75 mm for the statistical analysis. This procedure is identical to that used in Laneau and Wouters (2004).

D. Results and discussion

Figure 1 shows the average jnd's (expressed in mm) of the subjects as a function of the space constant for both reference center frequencies. The logarithms of the jnd's were used for averaging and to obtain a normal distribution for the statistical analysis.

A repeated measures analysis of variance (ANOVA) was performed on the logarithms of the estimated jnd data with

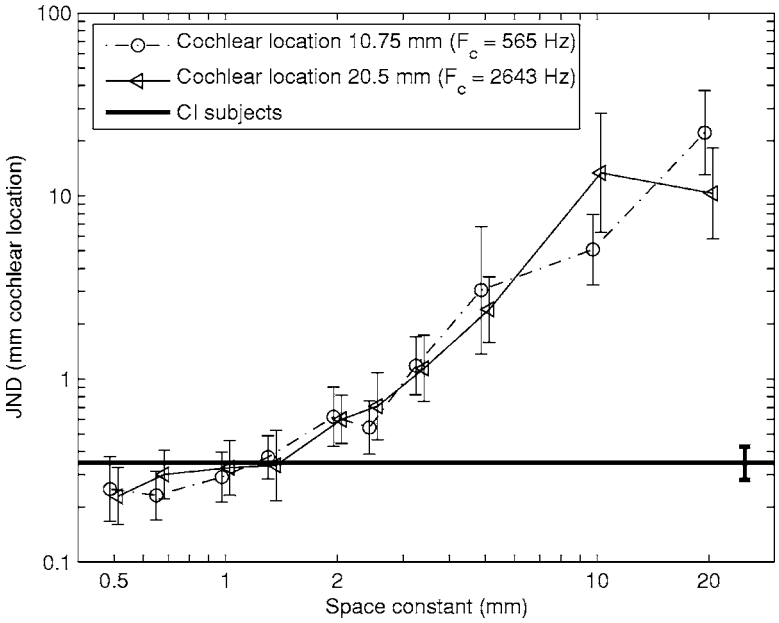

FIG. 1. Results for simulated electrode discrimination using the CISIM filter. Electrode discrimination was simulated by discrimination of center frequency of noise bands. Just noticeable differences were estimated as a function of the spectral overlap between the noise bands and expressed as distance shift in cochlear position. As the space constant decreases, the noise bands become narrower with less overlap, and the discrimination improves. The error bars indicate ±1 standard error of the mean of the intersubject variability. There was no significant difference between the two simulated insertion depths. The average result of electrode discrimination of four Nucleus CI24 subjects as measured in a separate study (Laneau and Wouters, 2004) is added with the horizontal solid line. The standard error of the mean for the CI subjects is approximately 0.07 electrodes.

the reference center frequency and the space constant as factors. There was a significant effect of space constant ($p = 0.002$). No other effects were significant.

The variation of jnd as a function of space constant is qualitatively equivalent to the resonance frequency discrimination of filtered noise (Gagne and Zurek, 1988). Gagne and Zurek found that when the Q factor of the noise-shaping filter was decreased, the center frequency discrimination was impaired. Similarly in the present experiment, increasing the space constant resulted in an increase of place pitch jnd's. Gagne and Zurek also found that the difference limen expressed relative to the center frequency is roughly constant when the Q factor or spectral slope was held constant. This is also found in the present study in that there is no difference between the results obtained at both reference center frequencies and noting that a shift in cochlear position corresponds to a constant relative frequency difference.

The jnd's measured by Gagne and Zurek are, however, significantly smaller than the jnd's obtained in the present study, especially for large space constants. (In comparing the jnd's, the d' values derived from the jnd's of both studies were assumed to vary linearly with distance along the basilar membrane.) For the larger space constants it is, however, hard to compare the two filter shapes, since for low values of Q the resonance filters of Gagne and Zurek become low-pass filters with a steeper slope compared to the slopes of the filters of the present study. Part of the lower sensitivity in the present study is also most probably due to the relatively large range of intensity roving in the present study while Gagne and Zurek did not rove the intensity of their stimuli. Also differences between subjects of the two studies might have contributed to the lower sensitivity reported in the present study. In the present study a large variability in subjects' performance is observed (see Fig. 1 and below) which is mainly caused by the performance of two NH subjects performing relatively poorly on pitch discrimination (see experiment 4).

Using the present custom-made CISIM filter there was no difference in place pitch sensitivity across reference center frequency or equivalent simulated electrode position. Similarly, in cochlear implants place pitch sensitivity is not significantly different at different electrode locations (Nelson *et al.*, 1995).

The place pitch discrimination of the CI subjects was measured in Laneau and Wouters (2004). In this study, electrode discrimination was measured for single- and multichannel stimuli. The experimental procedure was very similar to the procedure used for the present simulated electrode discrimination experiment. In the cochlear implant experiment it was, however, impossible to change the space constant (λ) of the decay of the excitation pattern. The average jnd for electrode discrimination of the four CI subjects is added as a solid horizontal line to Fig. 1 for comparison. The average performance of the NH subjects matches the average CI performance for a space constant of approximately 1 mm. Note that the filter designed based on a space constant of 1 mm has relatively steep slopes of approximately 40 dB/oct.

III. EXPERIMENT 2: THE EFFECT OF SPECTRAL SMEARING ON F0 DISCRIMINATION

In experiment 2, F0 discrimination of stylized synthetic vowels was measured in NH subjects for different amounts of spectral smearing in the noise-band vocoder. The vocoder processing was designed such that only spectral pitch (or mimicked place pitch) cues were useful for the subject. The subjects and equipment for stimulus presentation were the same as in experiment 1.

A. Methods

1. Stimuli

The stimuli consisted of stylized vowels processed with a noise-band vocoder. The unprocessed stylized vowels were a subset of the original stimuli used in the comparison study (Laneau *et al.*, 2004). Stylized vowels were generated by

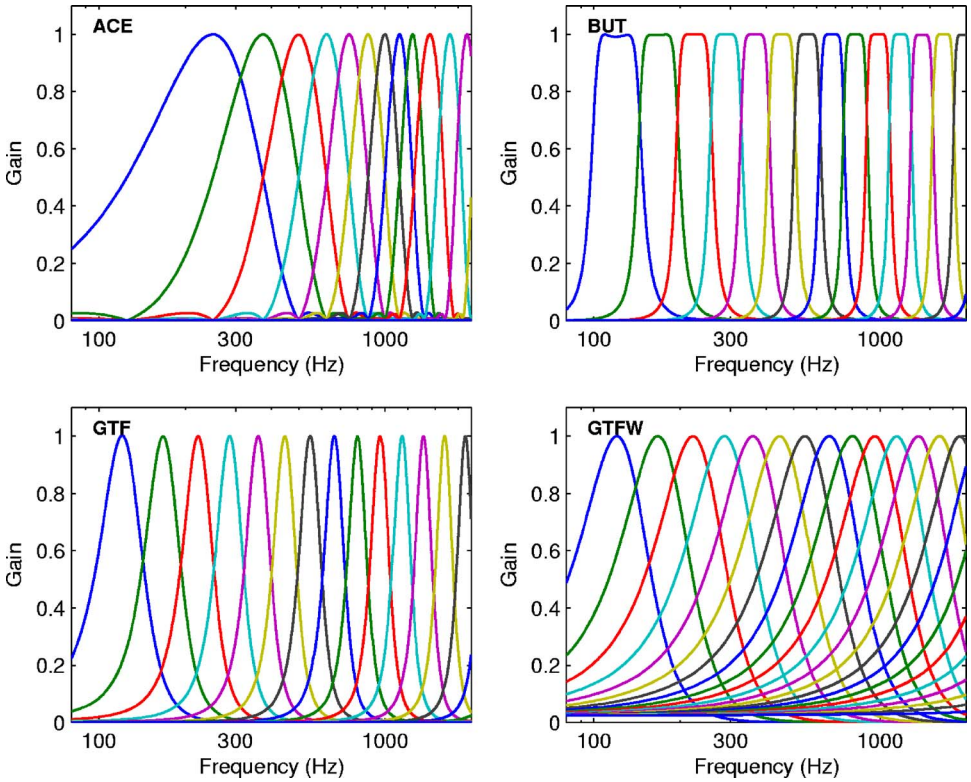

FIG. 2. (Color online) Details of the frequency response of the four analysis filter banks used in the present study.

passing a pulse train through a low-pass filter and a single resonator. The low-pass filter was a second-order IIR filter with a cutoff of 50 Hz and the output of the low-pass filter resembled the glottal volume velocity. The resonator was a second-order IIR resonating filter that created a formant in the spectrum. The details of both filters can be found in Laneau *et al.* (2004). The formant frequency was set at 300, 350, 400, 450, or 500 Hz while the bandwidth of the resonator was fixed at 100 Hz.

These single formant stimuli (SFS) were created with fundamental frequencies 133, 140, 150, 165, 180, 210, 250, 325, and 450 Hz. The sampling frequency was 16 kHz. All stimuli were 500 ms long and had equal rms power.

2. Processing—analysis

The analysis part of the vocoder is identical to the analysis stage of the processing in the study of Laneau *et al.* (2004) with the Gammatone filter bank. The Gammatone filterbank (GTF) has 22 bands and its frequency response is shown in Fig. 2. The GTF filter bank is based upon a model for the filtering of the normal ear (Patterson *et al.*, 1995). The GTF filter bank consists of 22 fourth order IIR gammatone band-pass filters with center frequencies spaced to simulate auditory filters distributed equidistantly along the length of the basilar membrane. The bandwidth of the filters was set to the ERB_N of the respective auditory filter (Glasberg and Moore, 1990).

In order to obtain a condition mainly focusing on spectral pitch (simulating place pitch), the envelopes of the filter outputs were low-pass filtered at 10 Hz so that all fluctuations causing temporal pitch cues were removed. More details of the implementation of the GTF filter bank and the envelope extraction are given in Laneau *et al.* (2004).

Finally, maxima selection was performed in two steps. First, the obtained envelopes were resampled from the audio sampling rate to 900 Hz (the stimulation rate per channel in the CI). Then, for each time slot (sampled at 900 Hz) only the 8 most intense channels of the 22 channels were retained, while the other channels were set to zero. In contrast to CI processing, no compression of the envelopes was included in the noise-band vocoder schemes.

3. Processing—resynthesis

In the resynthesis stage, the output of the analysis stage is upsampled to 30 kHz and then modulated with noise bands. These noise bands are generated by filtering white Gaussian noise through a bank of 22 CISIM filters, to simulate an electrode array consisting of 22 electrodes spaced 0.75 mm apart, as the Cochlear Nucleus CI24(M) electrode array. The space constant λ of the 22 CISIM filters was set to 2, 1.33, 1, 0.8, 0.67, 0.5, 0.33, or 0.25 mm, in order to assess the effect of changing the slope of the resynthesis filters (space constants or equivalently changing the channel discrimination) on place pitch cues for F0 discrimination. The center frequency of the lowest channel was set at 1148 Hz, corresponding to the cochlear location of a simulated electrode array inserted relatively shallow to 20 mm from the round window. This was done to minimize the negative effect of the peripheral auditory filters on temporal pitch sensitivity (Hanna 1992). Consequently, the center frequencies of the 22 noise-band carriers in the CISIM vocoder ranged from 1148 Hz up to 11 410 Hz, and the center frequencies of adjacent carriers were separated by approximately 12%.

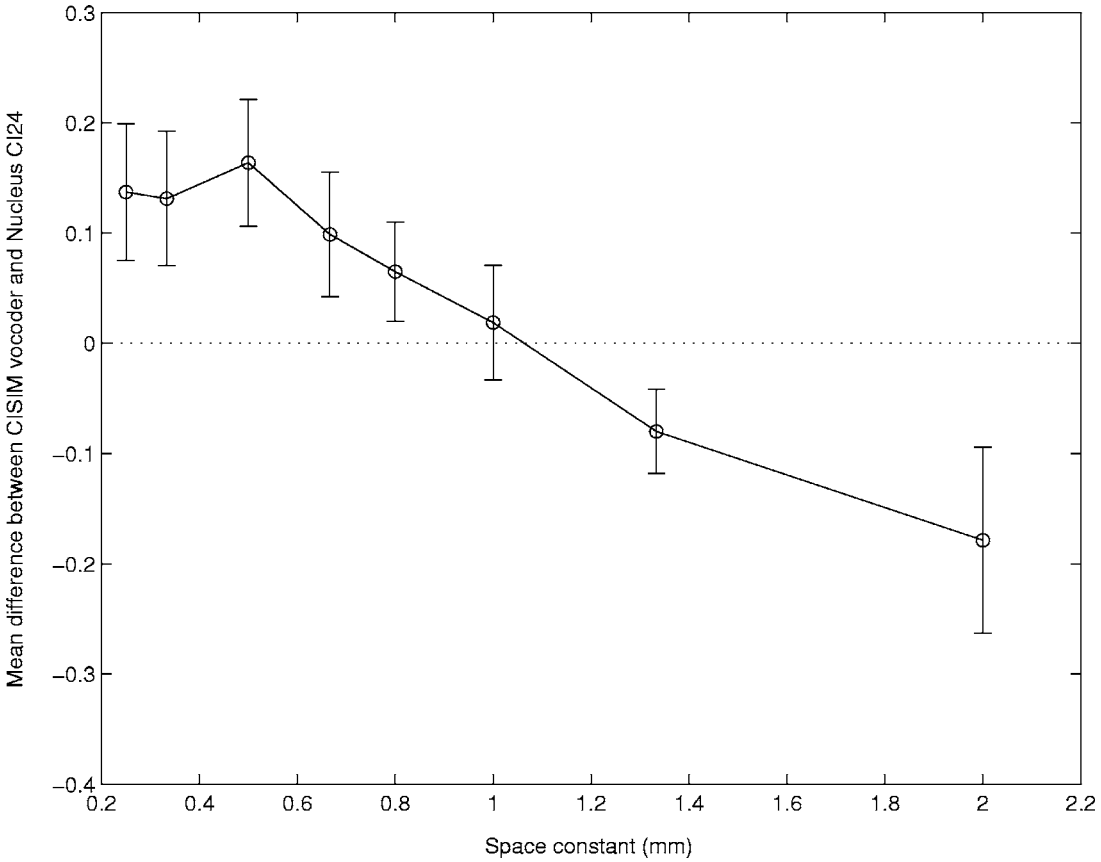

FIG. 3. The mean difference between the average psychometric curve of the CI subjects and the average psychometric curve of the NH subjects with the CISIM vocoder as a function of the space constant of the CISIM filters. Positive values indicate that the psychometric curves for the NH subjects were on average higher than for the CI subjects and consequently had better F0 discrimination and vice versa. All stimuli were processed with the GTF filter bank and with the envelopes low-pass filtered at 10 Hz. The best correspondence between the vocoder and CI results is a space constant of 1 mm. The error bars indicate ±1 standard error of the mean of the five NH subjects.

4. Procedure

F0 discrimination was measured in a 2I-2AFC procedure using the method of constant stimuli. The subjects were presented with two stimuli separated by a 500-ms silent gap and were asked to indicate which of the two intervals contained the stimulus with higher pitch, while ignoring the loudness cues. No feedback was presented to enforce subjects to use their intuitive sense of pitch and to prevent the subjects from using other cues than pitch. Both stimuli in a trial had the same formant frequency but one stimulus' F0 frequency was set at the reference frequency (133 Hz) and the F0 of the other stimulus was set at a higher frequency.

The trials were presented in blocks in random order, where each block contained only stimuli processed with the same space constant (or amount of spectral smearing). Each block contained trials with every comparison frequency for all five formant frequencies, and within each block each trial was repeated twice. Different formant frequencies were included per block in order to prevent the subject from identifying and learning the reference stimulus based upon other cues. To minimize loudness cues, the intensity of the stimuli was randomly attenuated over a 20-dB range.

B. Results and discussion

The psychometric curves were averaged across the 5 NH subjects and the difference was calculated between the average psychometric curve of the NH subjects and the average psychometric curve obtained for CI subjects in the same condition (Laneau *et al.*, 2004). For each space constant, the mean difference is calculated and this mean error is shown in Fig. 3 as a function of space constant. Positive values indicate that the F0 discrimination was better for the NH subjects using the CISIM vocoder than for the CI subjects. Negative values indicate that the CI subjects were able to discriminate better F0 differences than the NH subjects using solely place pitch cues.

NH listeners were better able to make use of place pitch cues using the vocoder with small space constants, or easily discriminable adjacent channels, than the CI subjects. However, by increasing the channel overlap the effect of the place pitch cues decreased and ultimately became less salient in the vocoder than in the CI subjects. The closest correspondence between the results using the CISIM vocoder and the CI results is obtained for a space constant of 1 mm. This is the value for the space constant that was also found to optimally mimic electrode discrimination (see experiment 1).

IV. EXPERIMENT 3: THE EFFECT OF SPECTRAL MISMATCH ON F0 DSICRIMINATION

A. Methods

In order to asses the effect of spectral mismatch on F0 discrimination of noise-band vocoded signals, F0 discrimina-

tion of stylized synthetic vowels was measured in NH subjects for two different noise-band vocoders: the standard vocoder and the CISIM vocoder. The subjects and equipment for stimulus presentation were the same as in experiment 1. The results obtained with each vocoder are compared to the results for CI subjects from the comparison study (Laneau *et al.*, 2004). The stimuli are the same as in experiment 2.

1. Processing—analysis

The analysis part of the vocoder is identical for both noise-band vocoders tested in the present experiment and is identical to the analysis stage of the processing in the comparison study (Laneau *et al.*, 2004). Four different filter banks were used to process the stimuli in the analysis stage: ACE, GTF, BUT, and GTFW. 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 and their frequency response is shown in Fig. 2. One filter bank (ACE) is identical to the filter bank currently implemented in the commercial speech processor of Nucleus devices (Cochlear Ltd., 2002). It consists of a 128-point fast Fourier transform (FFT) with a Hamming window. The center frequencies of the resulting frequency bins are spaced by 125 Hz, and their bandwidth is approximately 250 Hz. The 64 frequency bins are summed together to form 22 channels with an approximately logarithmic frequency resolution. The three other filter banks resemble more closely the frequency analysis of the normal human ear and have more resolution in the low frequencies compared to the ACE filter bank. The GTF filter bank is described in experiment 2. The BUT and GTFW are derived from this filter bank by changing the filters' shape and bandwidth, respectively. The GTF, BUT, and GTFW filter banks all consist of 22 fourth-order IIR band-pass filters with center frequencies spaced approximately equidistantly along the length of the basilar membrane. The filters of the GTFW filter bank were fourth-order gammatone filters, while Butterworth filters were used for the BUT filter bank. The bandwidth of the GTFW filter bank was set to half the filter's center frequency. For the BUT filter bank the bandwidth was set to have crossover frequencies at 3 dB down from the pass band.

In order to separate the effects of temporal and place pitch cues, the envelopes of the filter outputs were obtained under two different conditions. 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 was low-pass filtered at 10 Hz so that all fluctuations causing temporal pitch cues were removed. More details of the implementation of the four filter banks and the envelope extraction are given in Laneau *et al.* (2004). Further processing in the analysis stage is the same as in experiment 2.

2. Processing—resynthesis

The specification of the noise bands, serving as carriers in the resynthesis stage of the noise-band vocoders, differs between the two vocoders: standard vocoder and CISIM vocoder.

a. Standard vocoder. In the standard vocoder the output of the analysis stage is up-sampled to 16 kHz using sample-and-hold resampling and then modulated with noise bands. These noise bands are generated by filtering white Gaussian noise through fourth-order Butterworth band-pass filters. The cutoff frequencies of the resynthesis filters are set equal to the crossover frequencies of the filters of the analysis filter bank. Consequently, the resynthesis filters differ depending on the filter bank used in the analysis.

For the BUT and ACE filter bank the crossover frequencies of the analysis filter bank were determined analytically. For the GTF filter bank the cutoff frequencies of the resynthesis filter bank were set to 23 frequencies spaced evenly along the basilar membrane between 100 and 8000 Hz, where the conversion between place along the basilar membrane and frequency was calculated using Greenwood's formula (Greenwood, 1990). The standard vocoder is not used with the GTFW filter bank.

b. CI-simulation vocoder (CISIM). The main difference of the CISIM vocoder with the standard vocoder was the frequency response of the resynthesis filters. The CISIM vocoder used 22 CISIM filters, as described in experiment 2. The space constant of the 22 CISIM filters was set to 1 mm, because for this value the performance of the simulated place pitch discrimination in NH subjects was closest to the performance of the CI subjects (see experiments 1 and 2).

In contrast to the standard vocoder, the resynthesis filters were now kept fixed independent of the analysis filter bank. This resembles more closely CI operation because when the sound processing scheme is altered, the channels perceived by the subject do not immediately change accordingly because the electrodes remain at the same location and the current spread does not alter. It is noted, however, that after some time CI subjects do tend to adapt to new schemes (Fu *et al.*, 2002). Such adaptation is not active in Laneau *et al.* (2004), which reports on an acute experiment.

3. Procedure

The test procedure was identical to the procedure of experiment 2. The trials were presented in blocks. The conditions resulting from the combinations of type of envelope extraction (with or without temporal pitch cues) and filter bank (ACE, GTF, BUT, or GTFW) were presented into separate blocks. The block contents, number of trial repetitions, and the loudness roving were as in experiment 2.

For the standard vocoder the F0 of the comparison stimuli only ranged up to 250 Hz and the GTFW filter bank was not included, while for the CISIM vocoder the fundamental frequency of the comparison stimuli ranged up to 450 Hz, as in the comparison study (Laneau *et al.*, 2004). In summary, 12 blocks of 60 trials containing stimuli processed with the standard vocoder and 16 blocks of 80 trials containing stimuli processed with the CISIM vocoder were presented to every subject.

4. Data analysis

A measure of the effect of temporal and place pitch cues was derived from the proportions of correctly ranked trials. Because no significant effect of the formant frequencies was found (Laneau *et al.*, 2004), the proportions were averaged over the different formant frequencies. The resulting average

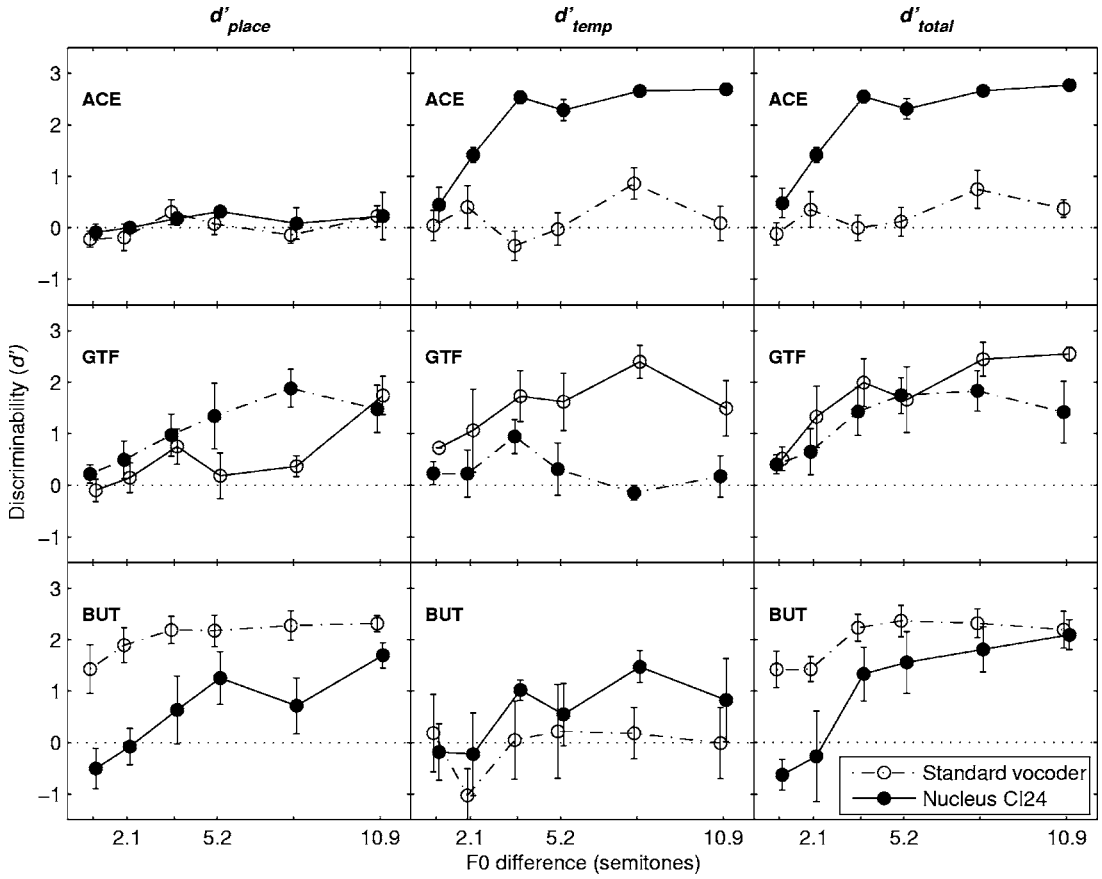

FIG. 4. The average effect of the place pitch cues, the temporal pitch cues, and the combined temporal and place pitch cues for F0 discrimination present in the stimuli processed with the standard vocoder for the ACE, GTF, and BUT filter banks. The results are averaged over the five NH subjects. The average result for four CI subjects from Laneau *et al.* (2004) is included for comparison. The error bars indicate ±1 standard error of the mean of the intersubject variability.

proportions were transformed into d' values, taking into account an adjustment for perfect performance. For perfect performance (for all proportions equal to one), half a trial error was introduced as suggested by MacMillan and Creelman (1991).

The d'_{place} values were estimated from the condition where the temporal fluctuations in the envelopes above 10 Hz were filtered out and represent an estimate of the effect of place pitch cues. The d'_{total} values were estimated from the condition where temporal fluctuations in the envelopes were present up to 200 Hz and represent an estimate of the combined effect of temporal and place pitch cues. The effect of the pure temporal pitch cues (d'_{temp}) was calculated by taking the difference between the effect of the combined pitch cues and the effect of the place pitch cues using the following formula:

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

In this formula, temporal and place pitch cues are assumed independent and negative d' values are possible (Laneau, 2005).

In order to compare the results obtained with the vocoder and the results obtained for CI subjects, the d' values for place pitch cues, temporal pitch cues, and combined temporal and place pitch cues were calculated for the results of the NH subjects using the vocoders and for the results of four CI subjects (Laneau *et al.*, 2004). For both vocoders, a repeated measures ANOVA was performed on the d' values with two within-subject factors (analysis filter bank and F0-difference) and one intersubject factor (separating NH subjects from CI subjects). This analysis was performed for d'_{place}, d'_{temp}, and d'_{total} separately.

B. Results

1. Standard vocoder

Figure 4 displays the effects of place pitch cues, temporal pitch cues, and combined pitch cues in the left, middle, and right columns, respectively, obtained with the standard vocoder. Each row of panels shows results of stimuli processed with a different analysis filter bank. The average results for four CI subjects (Laneau *et al.*, 2004) are included in the figures for comparison.

(i) The repeated measures ANOVA performed on the d'_{place} values (shown in the left column of Fig. 4) comparing the vocoder results with the CI results indicate that there was a significant difference between the results obtained with the vocoder and the results obtained with the CI subjects ($p=0.048$). Moreover, this difference depended significantly on the analysis filter bank ($p=0.042$). The place pitch cues had a signifi-

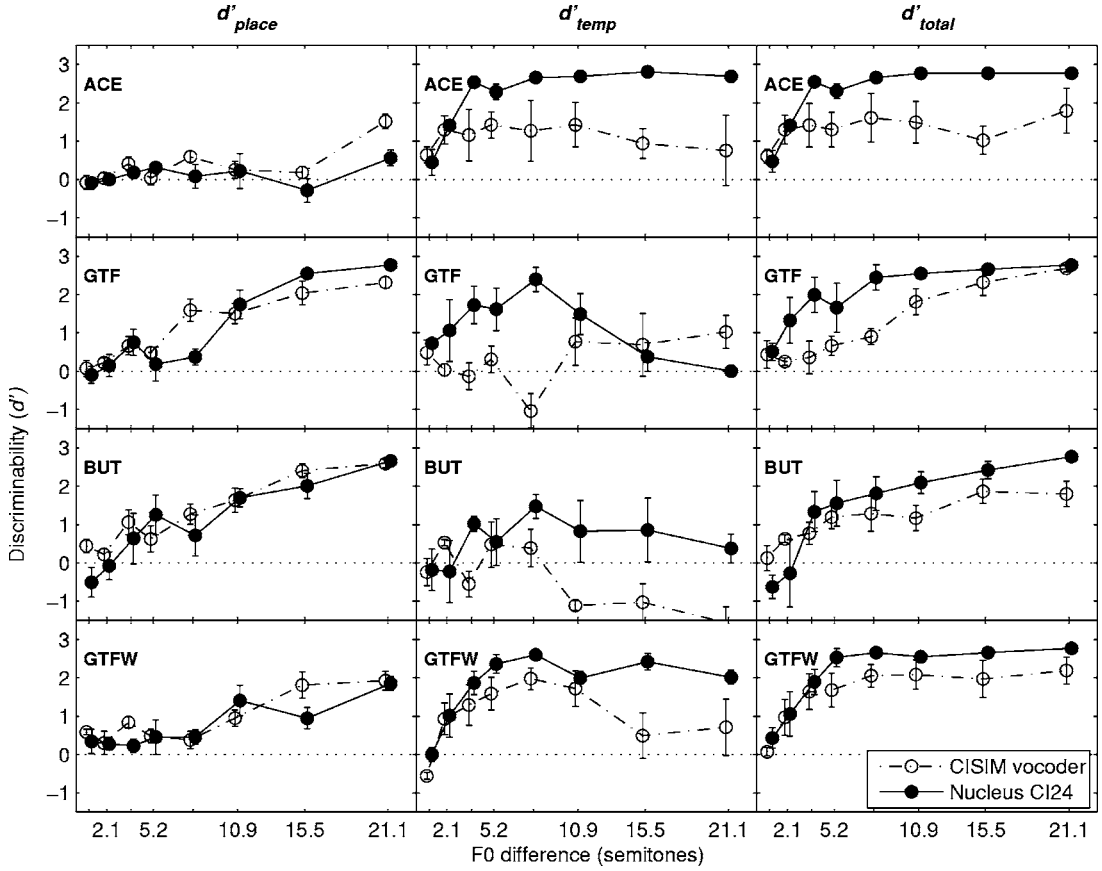

FIG. 5. The average effect of the place pitch cues, the temporal pitch cues, and the combined temporal and place pitch cues for F0 discrimination present in the stimuli processed with the CISIM vocoder for different filter banks and as a function of relative F0 difference. The results are averaged over the five NH subjects. The error bars indicate ±1 standard error of the mean of the intersubject variability. The average result for four CI subjects from Laneau *et al.* (2004) is included for comparison.

cantly larger effect in the NH subjects than in the CI subjects. The largest difference between the standard vocoder results and CI results occurred for the BUT filter bank, where the standard vocoder allowed the subjects to discriminate frequency differences smaller than a semitone (approximately 5%) solely based upon place pitch cues. For the ACE filter bank the place pitch cues did not allow F0 discrimination with either the standard vocoder or with the CI.

(ii) The analysis of the d'_{temp} values (shown in the middle column of Fig. 4) indicated that the temporal pitch cues had significantly different effects for the NH subjects with the standard vocoder than for the CI subjects ($p=0.002$). The perceived pitch effect of adding temporal modulations in the envelope up to 200 Hz was smaller for the NH subjects with the standard vocoder than for the CI subjects.

(iii) The combined effects of both temporal and place pitch cues (d'_{total} shown in the right column of Fig. 4) were also compared between the standard vocoder and the CI. Although there is no significant difference for F0 discrimination with combined place and temporal pitch cues between the CI subjects and NH subjects with the standard vocoder over all the filter banks ($p=0.295$), there is a significant difference depending on the processing filter bank used ($p=0.001$). For example, the ACE filter bank led to the best performance in the CI subjects but to the worst performance in the NH subjects using the standard vocoder. Similarly, while the BUT filter bank led to the best performance in NH subjects, it had the worst performance of all filter banks in the CI subjects.

2. CISIM vocoder

Figure 5 shows the effect the place pitch cues (d'_{place}), the effect of the temporal pitch cues (d'_{temp}), and the effect of the combined place and temporal pitch cues (d'_{total}) averaged over the five NH subjects for the CISIM vocoder, in the right, middle and left columns, respectively. Each row of panels shows the results for a different analysis filter bank. The average results for four CI subjects (Laneau *et al.*, 2004) are included in each figure for comparison.

(i) The statistical analysis of the perceptual effect of the place pitch cues (d'_{place}), showed no significant difference between the CISIM vocoder and the CI results ($p=0.412$). There were significant effects of the filter bank ($p=0.002$), the relative F0 difference ($p<0.001$), and a significant interaction effect between the filter bank and the relative F0 difference ($p=0.023$).

(ii) The analysis of the amount of effect of temporal pitch cues (d'_{temp}) showed that both the filter bank and the relative F0 difference had a significant effect on performance ($p=0.002$, and $p=0.032$, respectively). There was a significant difference between the results obtained with the CISIM vocoder and with the CI subjects ($p=0.028$). However the procedure for obtaining these results (with the CISIM vocoder or with CI subjects) did not interact significantly with either F0 difference or analysis filter bank.

(iii) The analysis of the effects of the combined place and temporal pitch cues (d'_{total}) showed only a significant effect of the relative F0 difference ($p<0.001$). There was no significant overall difference between the CISIM data and the results obtained from the CI subjects ($p=0.140$).

The CISIM vocoder performed well in modeling the effect of the place pitch cues present in the CI results. There is no significant difference between the results obtained from NH subjects with the CISIM vocoder and from CI subjects for the condition where only place pitch cues were present. Moreover, the comparison in performance over the four different analysis filter banks leads to similar results when using the CISIM vocoder results or the CI results: The effect of the place pitch cues was largest for the BUT and GTF filter banks, while for the ACE filter bank there was almost no effect of place pitch cues.

However, the effects of the temporal pitch cues were smaller with the CISIM vocoder than those present in the CI results. For both the ACE and GTFW filter bank F0-discrimination performance improved by adding temporal modulations in the envelopes up to 200 Hz. Temporal pitch cues thus had an effect for these filter banks, although the effect is smaller than the effect found in the CI subjects. It should also be noted that the intersubject variability is greater for the CISIM vocoder than for the CI results. For the GTF and BUT filter bank no clear benefit in F0 discrimination is obtained by providing temporal modulations in the envelopes up to 200 Hz for the CISIM vocoder. Even more, for the BUT filter bank for large F0 differences the F0-discrimination performance dropped after adding the temporal fluctuations.

In summary, the CISIM vocoder and the CI provide the approximately equivalent place pitch cues for F0 discrimination, but the CISIM vocoder provides less effective temporal pitch cues than the CI. Consequently, the CISIM vocoder provides less effective pitch cues in total (combined place and temporal pitch cues) for F0 discrimination than the CI, although not significantly.

C. Discussion

In most noise-band vocoders used for CI research white noise is modulated before the resynthesis filters are applied (e.g., Shannon *et al.*, 1995). In the present study the modulation of the noise is performed after the filtering for two reasons. First, the band-pass filtering may reduce the depth of the modulation of the envelope for higher modulation frequencies. And, second, the different group delays of the resynthesis filters alter the phase relations between the modulations on the different channels.

Neither the standard nor the CISIM vocoder was completely suitable as acoustic models for CI. For the standard vocoder neither the effects of the place pitch cues nor the effects of the temporal pitch were correctly modeled. For the CISIM vocoder the effect of the temporal pitch cues was somewhat underestimated, but the effects of the place pitch cues were accurately modeled for all analysis filter banks. In the next sections we discuss the possible factors affecting the results for place pitch cues and temporal pitch cues for both vocoders.

1. Place pitch cues

With the standard vocoder the NH subjects obtained better F0 discrimination than the CI subjects when the discrimination was solely based on place pitch cues. This higher pitch sensitivity using the standard vocoder was most likely caused by the relatively high-frequency resolution of the vocoder processed signals and the fact that there is no spectral mismatch in the standard vocoder. The resynthesis filters of the standard vocoder were relatively steep and did not overlap. Consequently, the spectrum of the signals was relatively little smeared by the standard vocoder. In contrast, excitation patterns of adjacent channels in CI probably had shallow slopes and overlapped greatly because of the current spread along the cochlea (Shannon, 1983b). Such overlap would smear the resulting excitation pattern and consequently weakened the place pitch differences between the compared signals.

The high performance of the NH subjects with the BUT filter bank using purely place pitch cues was likely due to the fact that for the BUT filter bank the analysis and resynthesis filters were identical. In this way the spectrum was least distorted for the BUT filter bank and the spectrum was minimally warped along the cochlea. Presenting the correct frequency to the correct place of stimulation is crucial for NH subjects to obtain good frequency discrimination (Oxenham *et al.*, 2004).

For the CISIM vocoder the slopes of the resynthesis filters were set to obtain equal channel discrimination using the CISIM vocoder as found in CI subjects (see experiment 1 and 2). With this adjusted channel discrimination, the results of the NH subjects with the CISIM vocoder successfully model the results of CI subjects in an F0-discrimination task based solely on place pitch cues independent of the analysis filter bank used.

2. Temporal pitch cues

Although amplitude-modulated noise is known to elicit pitch percepts (Burns and Viemeister, 1976) and the sensitivity to temporal pitch in CI is similar to the sensitivity of rate discrimination of amplitude-modulated noise (Blamey *et al.*, 1984a), the effect of the temporal pitch cues was smaller for the NH subjects using the vocoders than for CI subjects. This limited temporal pitch sensitivity using the vocoder was most probably caused by a combination of factors. The most

important factors are the absence of envelope compression/expansion in the vocoder, the poor pitch sensitivity of some of the NH subjects in the present study, the ringing of the peripheral auditory filters of the normal ear, and the possible interference between the channel envelope and the noisy envelope inherent in the narrow noise band.

The first factor that most probably contributed to the limited effect of the temporal pitch cues using the noise-band vocoders is the absence of any compression/expansion of the envelopes in the vocoders. In the CI system the envelopes are compressed to accommodate for the reduced dynamic range and the steep loudness growth of CI subjects. However, even with compression, the resulting modulation depth in the neural excitation pattern may be larger for CI subjects compared to the modulation depth in the excitation pattern of NH subjects listening to the vocoders. This is because excitation is an expansive function of the input for electrical stimulation, while it is a compressive function for acoustic stimulation (Zeng and Shannon, 1994). Consequently, the relative reduced modulation depth may have impaired the effectiveness of the temporal pitch cues for the NH subjects because lower modulation depths lead to poorer modulation rate discrimination (Patterson *et al.*, 1978; Grant *et al.*, 1998). This hypothesis is tested and confirmed in experiment 4.

In experiment 4 it is shown that NH listeners with a vocoder require more modulation depth to achieve the same performance of modulation rate discrimination compared to CI subjects (see Fig. 7). This difference in perception has different effects for the four analysis filter banks. For the ACE and GTFW filter banks the modulation depth present in the stimuli ranged from 59% up to 65% for the ACE filter bank and from 35% up to 60% for the GTFW filter bank. The modulation depth was calculated using the "in-phase" method from Laneau *et al.* (2004). For these analysis filter banks the modulation depth was sufficient to elicit temporal pitch cues in the three NH subjects (S2, S4, and S5) and in all four CI subjects. For the GTF and BUT filter banks the modulation depth of the reference stimuli ranged from approximately 10% to 33%, and from 11% to 36%, respectively. This modulation depth was insufficient to generate discriminable temporal pitch percepts in any of the NH subjects while in contrast it was sufficient for at least some stimuli for the three better CI subjects. Although the modulations for the GTF and BUT filter bank were undiscriminable in rate for the NH subjects, they were still detectable. The required modulation depth for modulation detection at 133 is in the order of 10% for NH subjects (Bacon and Viemeister, 1985). The presence of these fluctuations may have elicited a roughness sensation (Zwicker and Fastl, 1999) that may have interfered with the place pitch cues and resulted in negative temporal pitch cues for the BUT filter bank and for some subjects with the GTF filter bank (see the middle column of Fig. 5).

This difference for rate discrimination as a function of modulation depth between the vocoder and the CI subjects may be overcome by the insertion of an additional compression/expansion stage into the vocoder. An expansion stage would increase the modulation depth and thus provide more effective temporal pitch cues. This suggested compression/expansion stage is not equivalent to the compression stage found in CI systems. The suggested compression/expansion stage is intended to overcome any residual differences in loudness growth (or more specifically perceptual modulation depth) between the CI with compression and the vocoder.

The second factor contributing to the difference in effectiveness of the temporal pitch cues between NH listeners and CI subjects is the relatively poor performance in pitch related tasks of two of the NH subjects. In experiment 4, two NH listeners, S1 and S3, were unable to perform the pitch discrimination task while this task is within the limits of normal performance for most NH subjects (Patterson *et al.*, 1978; Grant *et al.*, 1998). Futhermore, these same subjects also performed below average performance on another modulation rate discrimination task reported in Laneau (2005). There is no clear reason why these two subjects performed so poorly at modulation rate discrimination tasks.

A third reason why the effect of temporal pitch cues was lower in the NH subjects using the vocoder compared to the effect in CI is the peripheral filtering of the normal ear. Hanna (1992) showed that at low center frequencies the smaller bandwidth of the peripheral filters limits the effect of the temporal pitch cues. This effect is largest for the standard vocoder because for the standard vocoder the spectral region of the output signal is matched to the spectral region of the input signal and the original stimuli in the present study only contained energy in the lower frequencies because of the relatively low formant frequencies and the maxima selection in the processing. Consequently, the output signals of the standard vocoder in the present study only contained lower frequencies where the limiting effect of the peripheral filters exists. The F0-related modulations in the envelope of the different channels of the standard vocoder can be obscured by the modulations already present on the basilar membrane. Due to the limited bandwidth of the basilar membrane at lower frequencies the effective modulation depth of the F0-related modulations can be reduced.

This effect is absent for the CISIM vocoder because the center frequencies of the resynthesis filters are shifted up in frequency with respect to the center frequencies of the analysis filters. The stimuli processed with the CISIM vocoder contained energy at higher frequencies where the effect of the temporal pitch cues is not limited by the peripheral filters of the normal ear. There was a significant increase in the effect of temporal pitch cues between the standard and the CISIM vocoder ($p=0.031$) for the filter bank with sufficient modulation depth (ACE) and for the three better performing NH subjects.

Summarizing the three previous factors, a noise band vocoder can be used as an acoustic model for temporal pitch research in cochlear implants for NH subjects with relatively good pitch sensitivity, for stimuli with enough modulation depth, and for a vocoder with noise bands at higher frequencies (i.e., the CISIM vocoder). This is shown in Fig. 6 where the average effect of the temporal pitch cues (d'_{temp}) for subjects S2, S4, and S5 are depicted for the ACE and GTFW filter banks obtained with the CISIM vocoder. There exists

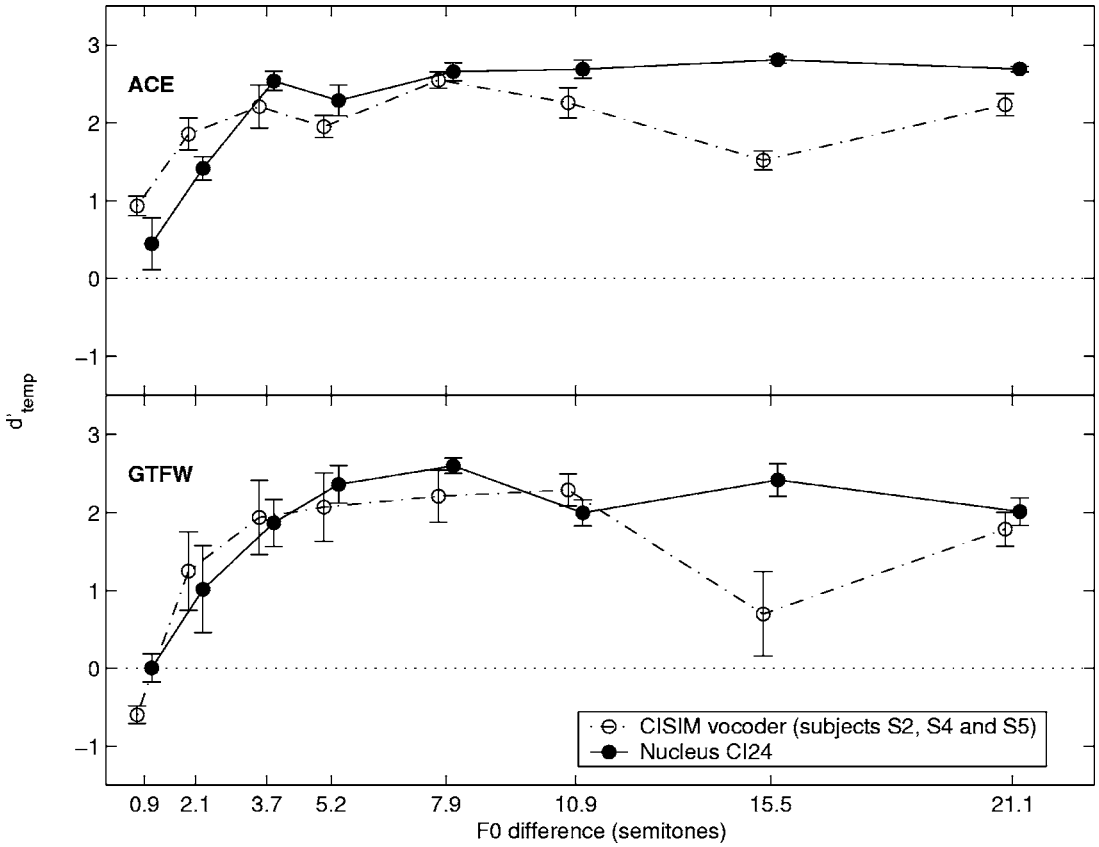

FIG. 6. The amount of effect of the temporal pitch cues for F0 discrimination for stimuli processed with the CISIM vocoder for filter banks ACE and GTFW and as a function of relative F0 difference. The results are averaged over NH subjects S2, S4, and S5 who have relatively good temporal pitch sensitivity. The results for four CI subjects from Laneau *et al.* (2004) are included for comparison. For this reduced set of conditions and for these subjects, the CISIM vocoder succeeds in modeling the CI data. The error bars indicate ±1 standard error of the mean of the intersubject variability.

relatively good correspondence between the results obtained with the CISIM vocoder for these subjects and for the results obtained with the four CI subjects from Laneau *et al.* (2004) added to the figure for comparison.

Finally, a fourth factor may have contributed to the poor effectiveness of the temporal pitch cues for the NH listeners using the vocoder compared to the CI subjects. The noise-band carriers of the vocoder have inherent random envelope fluctuations, creating an "external" variability. These random modulations may interfere with the envelope modulations in the analysis-channel envelopes related to the F0 and eliciting the temporal pitch cues (Formby and Muir, 1988; Hanna, 1992). Therefore, a deterministic carrier with limited envelope modulations in the temporal pitch frequency range might be more suitable as a carrier for a vocoder intended as an acoustic model for pitch sensation in CI (Carlyon *et al.*, 2002; van Wieringen *et al.*, 2003; Deeks and Carlyon, 2004).

V. EXPERIMENT 4: MODULATION RATE DISCRIMINATION

A. Methods

To assess the effect of the absence of a compression/expansion stage in the vocoder on temporal pitch cues, the minimal modulation depth required to discriminate a 20% change in modulation frequency (approximately 3.2 semitones) was measured in the same five NH subjects using the CISIM vocoder and in the four CI subjects of Laneau *et al.* (2004). For the CI subjects, the stimuli consisted of amplitude-modulated pulse trains presented interleaved on the three most apical channels with 900 pulses per second per channel. A dc-shifted sinusoid was compressed using the standard compression function to accommodate for loudness growth in Nucleus CI24 subjects (Laneau *et al.*, 2004) and used to modulate the amplitude of the pulses of the three channels. The modulation was in phase across the three channels and the modulation depth was varied adaptively. The subjects were presented two signals in random order on each trial: one was modulated at 133 Hz and the other at approximately 160 Hz. Subjects were asked to indicate the higher in pitch. After two consecutive correct answers the modulation depth was decreased by 1 dB and after each incorrect answer the modulation depth was increased again by 1 dB, leading to an asymptotic average of 71% correct responses (Levitt, 1971). The procedure was continued until eight reversals were obtained and the mean of the last four reversals was taken as the result for that particular run. The three best runs out of five runs were retained for each subject. Intensity was roved identically as in Laneau *et al.* (2004) by randomly varying the electrical output gain from 85% up to 110% of the dynamic range, to minimize loudness cues.

For the NH subjects, stimuli consisted of the sum of three modulated noise bands filtered with CISIM filters with

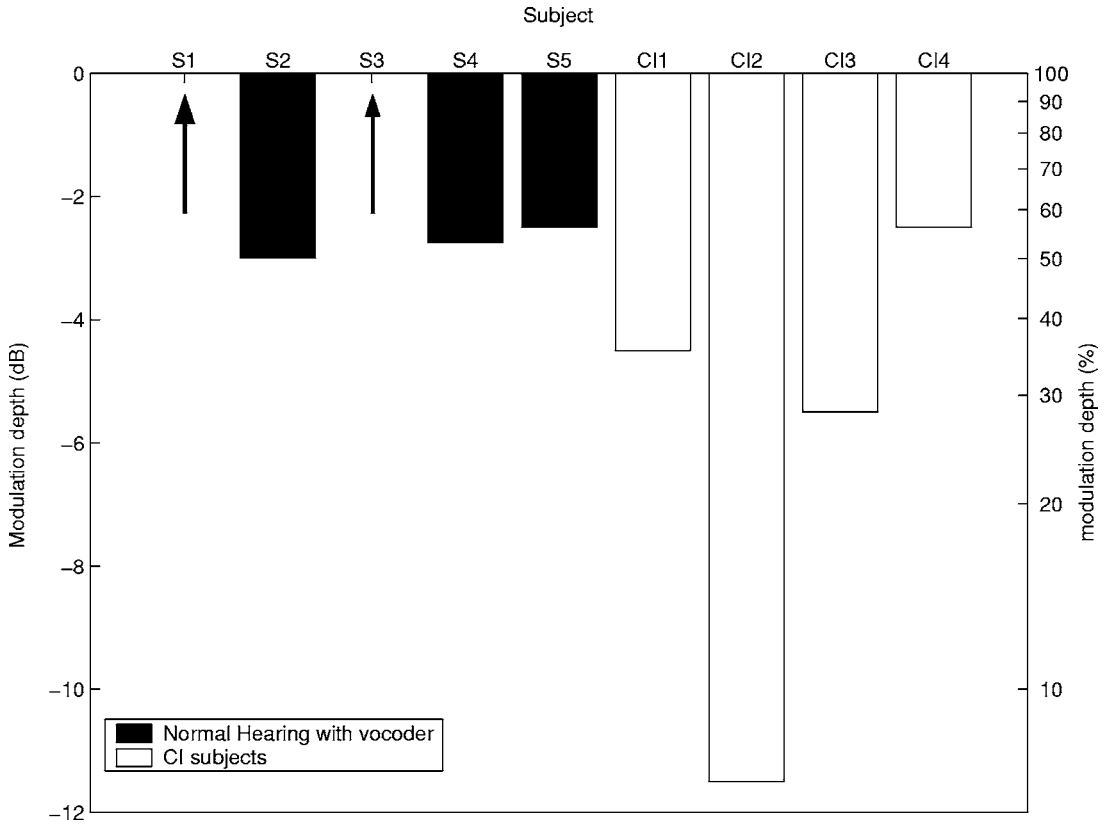

FIG. 7. Minimal required modulation depth to discriminate a 20% modulation rate difference on three simultaneously stimulated channels. The filled bars show the results of the NH subjects (S1–S5) where the modulation depth was measured using modulated noise bands (CISIM filters with space constant 1 mm). The open bars show the results of the CI subjects (CI1–CI4) where pulse trains were modulated. The arrows for subjects S1 and S3 indicate that they were unable to correctly rank the 20% modulation rate difference above chance level even for 100% modulation depth. The results show that the required modulation depth for rate discrimination is generally lower for the CI subjects compared to the NH subjects.

space constant of 1 mm and center frequencies of 1148, 1291, and 1451 Hz, corresponding to the cochlear locations of the three most apical electrodes of the simulated electrode array of the CISIM vocoder (15, 15.75, and 16.5 mm, respectively). The modulation was sinusoidal and in-phase over the three channels. The procedure was identical to the procedure for the CI subjects. Intensity was roved over a 20-dB range. No compression or expansion of the envelope was included for the acoustical stimuli.

B. Results and discussion

Figure 7 shows the results for both the NH subjects and the CI subjects. Two of the NH subjects (S1 and S3) were unable to discriminate the rate difference even at 100% modulation depth. CI subjects CI1, CI2, and CI3 require the smallest modulation depth to discriminate the 20% modulation rate difference. CI subject CI4 is comparable in performance to the better NH subjects. In general, CI subjects thus require less modulation depth compared to NH subjects with the CISIM vocoder to discriminate the 20% difference in modulation rate.

VI. GENERAL DISCUSSION

Our results indicate that the relative contributions of spatial and temporal pitch cues for F0 discrimination can be altered by varying the width of the single-channel excitation patterns. Narrower excitation patterns with steeper slopes elicit highly salient place pitch cues and with very narrow excitation patterns these place pitch cues may become more salient than the temporal pitch cues.

In currently used CI systems F0 discrimination is mediated by temporal pitch cues more than by place pitch cues (Geurts and Wouters, 2001; Green *et al.*, 2002; Laneau *et al.*, 2004). However, the present results suggest that narrower excitation patterns may provide more salient place pitch cues. Narrower excitation patterns were reported using bipolar or tripolar stimulation compared to the monopolar mode used in most current CI systems (Hartmann and Kral, 2004). Ultimately, these narrow excitation patterns may elicit place pitch cues which are more salient than the temporal pitch cues. This might enhance F0 discrimination in CI subjects whereas the limit for the temporal pitch cues appears to be reached (Green *et al.*, 2004).

In most studies using vocoders to investigate the effects of processing for pitch perception (or pitch-related tasks) in CI subjects the carriers of the vocoders are spectrally matched, as in the standard vocoder in the present study, i.e., the analysis filters and the resynthesis filters are identical for noise-band vocoders (Fu *et al.*, 1998; Faulkner *et al.*, 2000; Green *et al.*, 2002; Xu *et al.*, 2002; Green *et al.*, 2004; Qin and Oxenham, 2005; Kong *et al.*, 2004; Fu *et al.*, 2004) or pulse-train vocoders (Deeks and Carlyon, 2004), or the frequency of the sinusoidal carrier is set at the center frequency

of the analysis filters for sinusoidal vocoders (Lan *et al.*, 2004; Fu *et al.*, 2004). First, our results indicate that the place (or spectral) pitch cues with the spectrally matched vocoders are more salient than those found in the CI subjects, especially for the case where analysis and resynthesis filters were completely identical as in the BUT condition for the standard vocoder. This suggests that the spectral pitch cues found in the mentioned vocoder studies may have been stronger than what may be obtained in CI subjects, especially when many channels are used in the vocoder. Second, our results also indicate that the temporal pitch cues obtained with the standard vocoder are less salient than the cues CI subjects can use because of the peripheral filtering in the normal ear. This suggests that with the spectrally matched vocoders of most other studies the obtained temporal pitch cues may be smaller than what may be found in CI subjects because the channels with center frequencies below 1 kHz are less effective in providing salient temporal pitch cues. Taken together the last two findings, it is possible that the spectral pitch cues were relatively more important than the temporal pitch cues with the vocoders than what may be obtained with CI subjects, especially as the number of channels would be high and the spectral resolution would be good. For sinusoidal vocoders the relative contribution of the spectral pitch cues may be even more exaggerated as the excitation profile of sinusoids is very narrow and this leads to very discriminable spectral pitch cues.

Pitch or F0 discrimination solely based upon place pitch cues was strongly affected in the present study by spectral smearing (see Fig. 3). This is in contrast with the smaller effect of spectral smearing on speech perception. Spectral smearing only affects speech understanding when the slopes of the noise-band carriers of the vocoder are more shallow than 18 dB/oct (Shannon *et al.*, 1998). This observation is consistent with the higher spectral resolution needed for melody recognition than for speech understanding (Smith *et al.*, 2002).

There exists, however, a difference between the amount of spectral smearing necessary to replicate CI subjects' speech perception performance and the amount of spectral smearing necessary to replicate CI subjects' pitch discrimination performance. The mean speech-in-noise recognition thresholds of implant users are close to those of NH subjects listening to four-channel spectrally smeared (with 6 dB/oct resynthesis filters) noise-band vocoded speech (Fu and Nogaki, 2005). In the present study the place pitch performance was matched for a 22-channel noise-band vocoder with slopes of 40 dB/oct (the CISIM vocoder). The cause of this difference in smearing necessary for matching CI performance is unknown, but two factors may be important. First, in the CISIM vocoder the spectrum is spectrally shifted and compressed along the cochlear axis. This probably affected performance because for a vocoder with a matched spectrum and shallower slopes (the standard vocoder in this study) the performance of place pitch discrimination was better even though the slopes of the carriers were shallower (24 dB/oct). This is similar to the reduced speech understanding for spectrally shifted and compressed speech (Fu and Shannon, 1999; Baskent and Shannon, 2003). In the studies assessing the effect of spectral smearing on speech perception the spectrum was not warped along the cochlear axis. Second, channel interactions in CI subjects may be greater during speech than during the stationary signals used in the present study because dynamic stimuli may cause stronger channel interactions (Chatterjee, 2003).

Only postlingually deafened CI subjects participated in the pitch discrimination experiments that were used to verify the acoustic model. However, both place pitch and temporal pitch mechanisms can be impaired in prelingually deafened subjects that were implanted at a relatively late age (approximately after 12 years of age) (Busby *et al.*, 1993; Busby and Clark, 2000a, b). In those subjects often no pitch ordering across the electrode array is observed (Busby and Clark, 2000b). Therefore, the acoustic model presented in this study is not likely to be generalized to this particular group of CI subjects. For prelingually deafened CI subjects that were implanted early in life, normal pitch perception can be observed (Busby and Clark, 2000a, b). In contrast to the previous group of subjects, these subjects experience auditory sensations in the "critical period" of their brain development and this allows for at least partial maturation of their auditory system (Hartmann and Kral, 2004). It may thus be possible to extend the present acoustic model to also include early-implanted prelingually deafened CI subjects, although this remains to be verified in future experiments. The model thus appears applicable for postlingually deafened CI subjects and early-implanted prelingually deafened subjects. These two groups constitute the major portion of all CI subjects.

The application of the presented acoustic model (the CISIM vocoder) to CI subjects in general may be complicated by the fact that the CI subjects participating in this study were all relatively good performers with very good electrode discrimination compared to other postlingually deafened CI subjects (Nelson *et al.*, 1995). Second, as already mentioned, the performance on the pitch discrimination task of some of the NH subjects in this study was poorer than that of other NH subjects on similar tasks reported in other studies. Because of these two factors the pitch discrimination obtained with the acoustic model may be somewhat too optimistic to replicate CI performance in general. Especially the length constant λ could be considered somewhat longer, or equivalently more spectral smearing, in more general applications of the model or a compressive function, might be used to limit the effectiveness of temporal pitch cues.

VII. CONCLUSIONS

Although noise-band vocoders have been proven to provide a successful acoustic model to study speech perception in cochlear implants, it may not be straightforward to extend the model for pitch perception research in CI. The results of the present study indicated that both temporal and place pitch sensitivity can be affected by parameters of the noise-band vocoder.

1. The degree of spectral overlap between adjacent resynthesis filters of the vocoder is inversely proportional to the place pitch sensitivity subjects obtain with the vocoder.

2. The discrimination of noise bands based on place pitch cues and as a function of simulated mm shifts is independent of center frequency for noise bands filtered with the CISIM filter.
3. If the discrimination of adjacent channels with the vocoder, or equivalently single-channel place pitch sensitivity, is adjusted (by varying the spectral overlap of the channels) to match the electrode discrimination performance in CI subjects, then the place pitch sensitivity with the vocoder matches the place pitch sensitivity of CI subjects also for complex multi-channel stimuli.
4. Temporal pitch sensitivity is limited for lower frequency channels of the vocoder. For noise bands having lower center frequencies the peripheral auditory filters limit the modulation depth and consequently the temporal pitch sensitivity. A noise band vocoder where the resynthesis filters all have high center frequencies (e.g., mimicking a shallow insertion depth) may overcome this effect of the peripheral auditory filters.
5. There is large variation in subjects' performance on pitch discrimination tasks, even for NH subjects.
6. A compression/expansion stage in the vocoder has to be adjusted to the effect of the compression in the CI system. The resulting modulation depths in the neural excitation pattern with electrical stimulation and with acoustical stimulation may differ if the CI compression is insufficient or too high. An extra compression or expansion function in the vocoder may compensate for these differences.

Taking these factors into account, a noise-band vocoder may be used as an acoustic model for pitch perception research in CI. Nevertheless, one has to take great care with the interpretation of the results because noise bands differ fundamentally from the deterministic pulse trains used in cochlear implants. Even for the CISIM vocoder, which was specifically designed to mimic pitch discrimination performance of CI subjects, the obtained results differed from the CI results for at least some conditions or subjects. This outcome limits the noise-band vocoder as a general model for electric hearing with a CI.

ACKNOWLEDGMENTS

We thank the subjects for their enthusiastic cooperation. We also thank Fred Wightman and Bob Shannon for valuable 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 No. IWT 020540), by the Fund for Scientific Research—Flanders/Belgium (Project No. G.0233.01), and by Cochlear Ltd.

Bacon, S. P., and Viemeister, N. F. (**1985**). "Temporal-modulation transfer-functions in normal-hearing and hearing-impaired listeners," Audiology **24**, 117–134.

Baskent, D., and Shannon, R. V. (**2003**). "Speech recognition under conditions of frequency-place compression and expansion," J. Acoust. Soc. Am. **113**, 2064–2076.

Black, R. C., and Clark, G. M. (**1980**). "Differential electrical excitation of the auditory nerve," J. Acoust. Soc. Am. **67**, 868–874.

Blamey, P. J., Dowell, R. C., Tong, Y. C., and Clark, G. M. (**1984a**). "An acoustic model of a multiple-channel cochlear implant," J. Acoust. Soc. Am. **76**, 97–103.

Blamey, P. J., Dowell, R. C., Tong, Y. C., Brown, A. M., Luscombe, S. M., and Clark, G. M. (**1984b**). "Speech processing studies using an acoustic model of a multiple-channel cochlear implant," J. Acoust. Soc. Am. **76**, 104–110.

Burns, E. M., and Viemeister, N. F. (**1976**). "Non-spectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Busby, P. A., and Clark, G. M. (**2000a**). "Electrode discrimination by early-deafened subjects using the cochlear limited multiple-electrode cochlear implant," Ear Hear. **21**, 291–304.

Busby, P. A., and Clark, G. M. (**2000b**). "Pitch estimation by early-deafened subjects using a multiple-electrode cochlear implant," J. Acoust. Soc. Am. **107**, 547–558.

Busby, P. A., Tong, Y. C., and Clark, G. M. (**1993**). "Electrode position, repetition rate, and speech perception by early- and late-deafened cochlear implant patients," J. Acoust. Soc. Am. **93**, 1058–1067.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

Chatterjee, M. (**2003**). "Modulation masking in cochlear implant listeners: envelope versus tonotopic components," J. Acoust. Soc. Am. **113**, 2042–2053.

Cochlear Ltd (**2002**). "Nucleus implant communicator (NIC) system overview," N95291 Iss.1, Cochlear Ltd.

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.

Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.

Dorman, M. F., and Loizou, P. C. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," Am. J. Otol. **18**, S113-S114.

Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

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.

Formby, C., and Muir, K. (**1988**). "Modulation and gap detection for broadband and filtered noise signals," J. Acoust. Soc. Am. **84**, 545–550.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q. J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**(1), 19–27.

Fu, Q. J., and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q. J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.

Fu, Q. J., Shannon, R. V., and Galvin, J. J., III (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Gagne, J. P., and Zurek, P. M. (**1988**). "Resonance-frequency discrimination," J. Acoust. Soc. Am. **83**, 2293–2299.

Geurts, L., and Wouters, J. (**2000**). "A concept for a research tool for experiments with cochlear implant users," J. Acoust. Soc. Am. **108**, 2949–2956.

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.
Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Grant, K. W., Summers, V., and Leek, M. R. (**1998**). "Modulation rate detection and discrimination by normal hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 1051–1060.
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.
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**, 2289–2297.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Hanna, T. E. (**1992**). "Discrimination and identification of modulation rate using a noise carrier," J. Acoust. Soc. Am. **91**, 2122–2128.
Hartmann, R., and Kral, A. (**2004**), *Cochlear Implants: Auditory Prostheses and Electric Hearing*, edited by F. G. Zeng, A. N. Popper, and R. R. Fay (Springer-Verlag, New York, 2004), Chap. 6, pp. 213–285.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
Lan, N., Nie, K. B., Gao, S. K., and Zeng, F. G. (**2004**). "A novel speech-processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.
Laneau, J. (**2005**). "When the deaf listen to music—pitch perception with cochlear implants," Ph.D. dissertation, Katholieke Universiteit Leuven, Belgium (http://hdl.handle.net/1979/57).
Laneau, J., and Wouters, J. (**2004**). "Multi-channel place pitch sensitivity in cochlear implant recipients," J. Assoc. Res. Otolaryngol. **5**, 285–294.
Laneau, J., Moonen, M., and Wouters, J. (**2004**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **116**, 3606–3619.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U. P., Cambridge).
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.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.
Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
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.
Patterson, R. D., Johnson-Davies, D., and Milroy, R. (**1978**). "Amplitude-modulated noise—detection of modulation versus detection of modulation rate," J. Acoust. Soc. Am. **63**, 1904–1911.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**(5), 451–460.
Shannon, R. V. (**1983a**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.
Shannon, R. V. (**1983b**). "Multichannel electrical-stimulation of the auditory-nerve in man. 2. Channel interaction," Hear. Res. **12**, 1–16.
Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
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.
van Wieringen, A., Carlyon, R. P., Long, C. J., and Wouters, J. (**2003**). "Pitch of amplitude-modulated irregular-rate stimuli in acoustic and electric hearing," J. Acoust. Soc. Am. **114**, 1516–1528.
Xu, L., Tsai, Y. J., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.
Zeng, F. G., and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric-stimulation of the human auditory-system," Science **264**, 564–566.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models, 2* (Information Sciences, Springer-Verlag, Berlin).