Original Paper

Audiology & Neurotology

Audiol Neurotol 2006;11:38–52
DOI: 10.1159/000088853

Received: October 14, 2004
Accepted after revision: June 29, 2005
Published online: October 10, 2005

Improved Music Perception with Explicit Pitch Coding in Cochlear Implants

Johan Laneau[a] Jan Wouters[a] Marc Moonen[b]

[a]Lab. Exp. ORL, and [b]ESAT-SCD, Katholieke Universiteit Leuven, Leuven, Belgium

Key Words
Cochlear implant · Music perception · Pitch · Sound processing

Abstract
Music perception and appraisal is very poor in cochlear implant (CI) subjects partly because (musical) pitch is inadequately transmitted by the current clinically used sound processors. A new sound processing scheme (F0mod) was designed to optimize pitch perception, and its performance for music and pitch perception was compared in four different experiments to that of the current clinically used sound processing scheme (ACE) in six Nucleus CI24 subjects. In the F0mod scheme, slowly varying channel envelopes are explicitly modulated sinusoidally at the fundamental frequency (F0) of the input signal, with 100% modulation depth and in phase across channels to maximize temporal envelope pitch cues. The results of the four experiments show that: (1) F0 discrimination of single-formant stimuli was not significantly different for the two schemes, (2) F0 discrimination of musical notes of five instruments was three times better with the F0mod scheme for F0 up to 250 Hz, (3) melody recognition of familiar Flemish songs (with all rhythm cues removed) was improved with the F0mod scheme, and (4) estimates of musical pitch intervals, obtained in a musically trained CI subject, matched more closely the presented intervals with the F0mod scheme. These results indicate that explicit F0 modulation of the channel envelopes improves music perception in CI subjects.

Copyright © 2006 S. Karger AG, Basel

Introduction

Cochlear implants (CI) are effective in restoring hearing in the profoundly deaf and allow sentence recognition in quiet for at least some subjects. However, music perception is poor in CI subjects compared to normal-hearing (NH) subjects [Gfeller and Lansing, 1991; Kong et al., 2004; Leal et al., 2003; Schulz and Kerber, 1993] and only few subjects enjoy listening to music with their implant [Gfeller et al., 2000]. In this study, a new sound processing scheme is presented to improve pitch and music perception in CI subjects and tested in six Cochlear Nucleus CI24 subjects.

Studies assessing music perception in CI subjects listening through their sound processors have shown that rhythm perception is close to the performance of NH subjects but that the CI subjects have more problems with the pitch patterns constituting the musical melodies [Gfeller and Lansing, 1991; Kong et al., 2004; Schulz and Kerber, 1993]. The pitch or fundamental frequency (F0) discrimination performance of CI subjects is very poor compared to NH subjects and covers a wide range from 1 semitone up to 2 octaves depending on the subject and the condition [Geurts and Wouters, 2001; Gfeller et al., 2002]. Furthermore, apart from being crucial for melody perception, pitch is important in speech for prosodic cues in most languages and semantic or grammatical cues in tonal languages. Pitch may also be used in separating multiple concurrent sources [Darwin and Carlyon, 1995].

Pijl [1997] showed that at least part of the limited pitch perception performance of the CI subjects was due to ineffective sound processing because CI subjects were able

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2006 S. Karger AG, Basel
1420–3030/06/0111–0038$23.50/0

Accessible online at:
www.karger.com/aud

Johan Laneau
Lab. Exp. ORL, Katholieke Universiteit Leuven
Kapucijnenvoer 33
BE–3000 Leuven (Belgium)
Tel. +32 16 332416, Fax +32 16 332335, E-Mail Johan.Laneau@philips.com

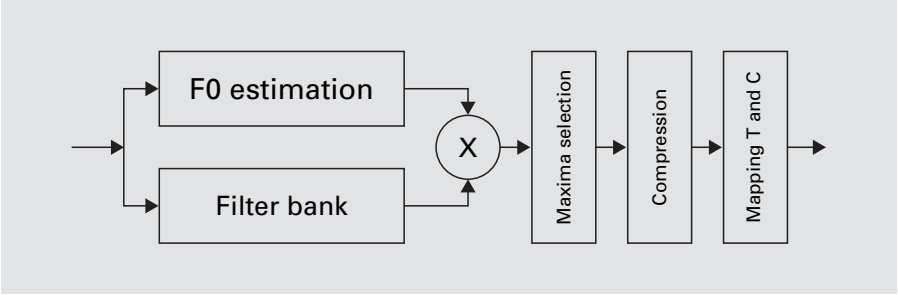


Fig. 1. Overview of the processing blocks of the newly proposed sound processing scheme (F0mod). See text for details.

to estimate musical intervals correctly for synthetic stimuli but the subjects were unable to do so when stimuli were presented through their speech processors. Therefore, a number of recent studies have developed new sound processing schemes to improve pitch perception in CI subjects [Geurts and Wouters, 2001, 2004; Green et al., 2004; Lan et al., 2004]. However, most of these studies have led to no or little improvement [Geurts and Wouters, 2001, 2004; Green et al., 2004] or were only evaluated with NH subjects using vocoders [Lan et al., 2004].

In the present study, a new sound processing scheme (F0mod) is proposed that is designed to enhance (musical) pitch perception. The scheme optimizes pitch perception related to both periodicity pitch and spectral pitch, being two dimensions of pitch [Licklider, 1954]. The periodicity pitch is presented to CI subjects by means of temporal pitch cues, while the spectral pitch is presented by means of place pitch cues.

Although the mechanism for periodicity pitch in NH subjects is incompletely understood [Oxenham et al., 2004], models based upon the autocorrelation function (all-order statistic) of the peripheral neural patterns can accurately predict the pitch perceived by NH subjects for most natural sounds occurring in the natural environment [Cariani and Delgutte, 1996; Licklider, 1951; Meddis and Hewitt, 1991]. In contrast, the purely temporal pitch percept in CI subjects is poorly modeled by an all-order statistic but is more closely related to a weighted sum of the first-order spike intervals [Carlyon et al., 2002]. To account for this discrepancy between pitch perception mechanisms, the sound processing scheme presented here extracts the periodicity pitch of the acoustic stimulus using an autocorrelation-based method and presents this extracted pitch to the CI subject through the first-order intervals of the modulation of the electrical pattern.

For NH subjects, the spectral pitch, sometimes also denoted sharpness or brightness of timbre, corresponds to the centroid of the spectrum [Anantharaman et al., 1993] or to the centroid of the excitation pattern [Zwicker and Fastl, 1999]. Similarly, the place pitch elicited by multichannel stimuli in CI subjects corresponds to the centroid of the excitation pattern [Laneau et al., 2004]. Consequently, the place pitch cues perceived by a CI subject will be matched to the spectral pitch cues as perceived by an NH subject when the excitation patterns along the basilar membrane are the same for electrical and acoustical stimuli. In the proposed F0mod scheme, it has been tried to minimize the distortion between the electrical and acoustical excitation pattern because the filter bank of the sound processing scheme is designed such that the electrically evoked excitation pattern is a linearly compressed version of the acoustically evoked excitation pattern.

The implementation of the new scheme is given in the section on sound processing. In the following sections, the new sound processing scheme (F0mod) is tested perceptually and the performance of CI subjects is compared to their performance with the current clinically used sound processing strategy for Nucleus (ACE). In the first two tests, F0 discrimination is measured for stylized vowels (see experiment 1) and for musical notes (see experiment 2). In experiment 3, melody recognition was measured and in experiment 4, a musically trained CI subject performed a musical interval labeling task.

Sound Processing

Implementation

An overview of the processing blocks of the new sound processing scheme (F0mod) is shown in figure 1. The input sound signal is presented to two parallel blocks. The first block, the filter bank, splits the signal into 22 channels and extracts the relatively slow varying envelopes of

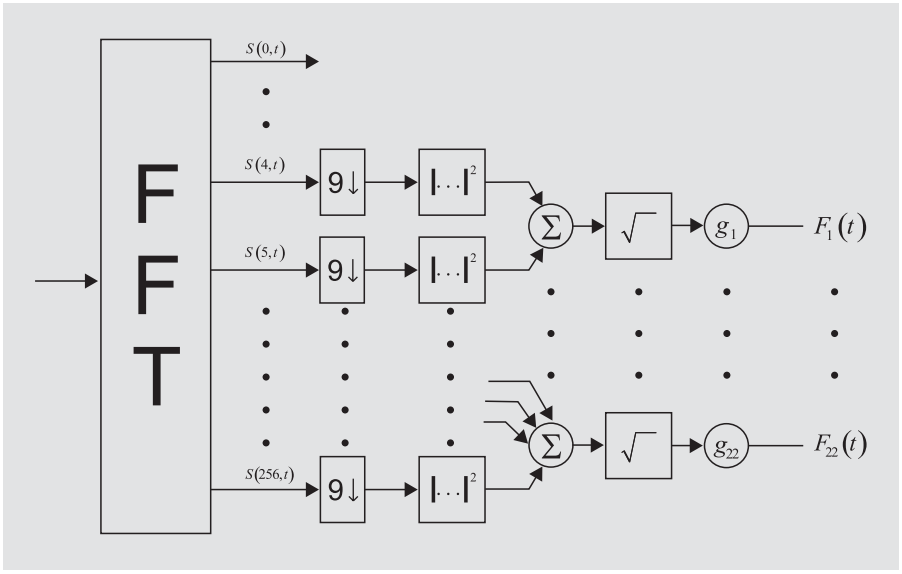


Fig. 2. Block diagram of the processing stages in the filter bank of the F0mod strategy. The input signal is analyzed using a 512-point FFT, then downsampled 9 times. The magnitude of the complex spectrum is calculated to obtain the Hilbert envelope. Subsequently, the different frequency bins are combined into 22 channels.

each channel. The second block estimates the F0 of the input signal based on the autocorrelation of the input signal. The 22 output channels of the filter bank are then modulated sinusoidally at the estimated F0. The next processing blocks perform maxima selection, compression of the waveform and map the waveform between the T and C levels of the subjects.

Filter Bank

The input signal is analyzed with a 512-point fast Fourier transform (FFT) and the Hilbert envelope is extracted for each frequency bin of the spectrum. The extracted envelopes of each bin are then summed into 22 separate channels. The block diagram of the processing stages in the filter bank is given in figure 2. The details of the implementation of the filter bank are discussed below.

The filter bank block reads in the input signal, sampled at 16 kHz, and stores it into buffers of 512 samples with 503 samples of overlap. This leads to an update rate (F_{update}) of approximately 1,778 buffers per second, equivalent to an effective 9-fold downsampling. The spectrum of each buffer is obtained through an FFT after applying a Hanning window. The envelope is estimated, based upon the definition of the envelope in the Hilbert transform, by calculating the magnitude of the complex spectrum. Because the bandwidth of each frequency bin is limited to 125 Hz, the extracted envelopes are limited to 62.5 Hz [Oppenheim and Schafer, 1999]. This cutoff frequency of the modulations is high enough to allow for good speech perception [Drullman et al., 1994; Shannon et al., 1995], but modulations at higher frequencies, leading to temporal pitch cues, are filtered out.

Because most speech signals do not contain energy below 100 Hz, the lowest 4 bins (including the DC bin) are discarded. Consequently, the filter bank spectrum ranges from 125 Hz up to 8000 Hz. The remaining bins are combined into 22 channels such that the bandwidths of these channels correspond to an equal number of equivalent rectangular bandwidths [Glasberg and Moore, 1990]. The assignment of the bins to each channel is shown in table 1.

The amplitude of each channel is defined by the total root mean square (RMS) of the assigned bins:

$$F_n(t) = g_n \sqrt{\sum_i |S(i,t)|^2}, \quad n = 1 \ldots 22 \tag{1}$$

where $F_n(t)$ is the amplitude of filter bank output channel n at time instant t, and $S(i,t)$ is the ith frequency bin of the FFT of the buffer at time instance t. The weights g_n are inserted to equalize the maximum output of all channels and to scale the output so that its dynamic range is exactly from 0 to 1, taking into account that the input signal is limited between –1 and 1 [Laneau, 2005].

Fundamental Frequency Estimation and Modulation

The second parallel block in figure 1 estimates the F0 of the input signal. A single F0 estimate ($F0_{est}$) is obtained for each buffer processed in the filter bank. The F0 is estimated by taking the inverse of the lag corresponding to the maximum in the circular autocorrelation. The circular autocorrelation is calculated through the inverse FFT

Table 1. Frequency allocation table for the 22 channels of the two filter banks that are compared in this study (F0mod and ACE)

Channel number																					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
F0mod																					
Number of bins																					
2	2	2	3	3	3	4	5	5	6	7	8	10	11	13	15	17	20	23	27	31	35
Center frequency low bin, Hz																					
125	188	250	313	406	500	594	719	875	1031	1219	1438	1688	2000	2344	2750	3219	3750	4375	5094	5938	6906
Center frequency high bin, Hz																					
156	219	281	375	469	563	688	844	1000	1188	1406	1656	1969	2313	2719	3188	3719	4344	5063	5906	6875	8000
ACE																					
Number of bins																					
1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	4	4	5	5	6	7	8
Center frequency low bin, Hz																					
250	375	500	625	750	875	1000	1125	1250	1375	1625	1875	2125	2375	2750	3125	3625	4125	4750	5375	6125	7000
Center frequency high bin, Hz																					
250	375	500	625	750	875	1000	1125	1250	1500	1750	2000	2250	2625	3000	3500	4000	4625	5250	6000	6875	8000

For the F0mod filter bank, the channels are combinations of 257 bins from the 512-point FFT. For the ACE filter bank, the channels are combinations of 65 bins from the 128-point FFT.

of the power spectrum. The estimated F0 is limited between 75 Hz and 593 Hz. The limits prevent the F0 detector to detect erroneous F0 values that are too high (above 593 Hz) or too low (below 75 Hz).

Subsequently, the channel amplitudes obtained from the filter bank are modulated with a sine at $F0_{est}$ and with 100% modulation depth. All channels are modulated with the same function and in phase:

$$A_n(t) = \left[0.5 + 0.5 \sin\left(\frac{2\pi}{F_{update}} \sum_{i=0}^{t} F0_{est}(i)\right)\right] F_n(t), \quad n = 1 \ldots 22 \quad (2)$$

where $A_n(t)$ and $F_n(t)$ are the modulated amplitude, and the output of the filter bank, respectively. F_{update} is the update rate of the filter bank (1778 Hz). More information about the signal processing aspects can be found in Laneau [2005].

The proposed sound processing strategy (F0mod) imposes the maximal modulation depth on all channels and makes this modulation in phase. Laneau et al. [2004] have shown that F0 discrimination improves when the modulation depth of the F0-related modulations in the channel amplitudes increases and that this increase is clearest when the modulation is in phase over the different channels. Consequently, the imposed modulations make the temporal pitch cues as clear as possible (maximal modulation depth) and as unambiguous as possible because only a single first-order interval is present in the modulation across channels.

Further Processing

The following processing blocks of the sound processing scheme are shown on the right hand side of figure 1. In the first processing block (maxima selection), only the amplitudes of the 8 channels with the largest amplitudes are retained per processed buffer and the other channels are set to zero. Subsequently, in the compression block, all channel amplitudes are compressed to accommodate for the reduced dynamic range of CI subjects and the steep loudness growth with increasing electrical current. The compression function is identical to the compression function used in the clinical Cochlear SPrint speech processor and is described in equation 1 of Laneau et al. [2004]. Finally, the resulting amplitudes are linearly mapped between the T and C levels for the appropriate channels of each subject.

The obtained current amplitudes were used to modulate pulse trains of 1800 pulses per second per channel. Consequently the total stimulation rate was 14400 pulses per second. The pulses on different channels were presented interleaved in time and ordered from base to apex. The pulses were biphasic pulses with a phase duration of 25 μs and an interphase gap of 8 μs. The stimulation mode was monopolar with both return electrodes active (MP1 + 2).

To accommodate for the small difference between the update rate of the filter bank (1778 Hz) and the stimulation rate per channel (1800 Hz), the channel amplitudes are resampled by duplicating every 81st sample. This

equalizes the sampling rate of the channel amplitudes with the stimulation rate.

Comparison Strategies

The sound processing strategy presented in the previous section (F0mod) is compared in the present study to two other strategies. The first strategy is the ACE strategy and it is based upon the ACE implementation in Cochlear SPrint processors. The second strategy is a variation of the F0mod strategy where the modulation of the envelopes is discarded so that temporal pitch cues are not present (denoted ACE512 in this study).

The ACE strategy is similar to the processing scheme presented in figure 1 but contains no F0 estimation block and the filter bank output channels are not explicitly modulated. The ACE filter bank reads in the input signal in buffers of 128 samples with 119 samples of overlap, leading to an update rate of 1778 Hz, as in the F0mod strategy. This contrasts with the implementation of the ACE strategy in the Cochlear SPrint processor where the update rate is limited to 760 Hz [Holden et al., 2002]. However, in the present implementation, the update rate was set to 1778 Hz to avoid aliasing effects in the perceived temporal pitch [McKay et al., 1994]. The buffers are analyzed using a 128-point FFT after application of a Hanning window. The frequency bin spacing is 125 Hz and the bandwidth of the main lobe of each frequency bin is 500 Hz. The envelope is extracted, as in the F0mod scheme, by calculating the magnitude of the spectrum. The lowest two bins are discarded and the remaining bins are combined into 22 channels using equation 1. The assignment of the bins to their respective channels is given in the bottom part of table 1. The relatively wide bandwidth of the bins in the ACE filter bank allows having frequencies in the extracted envelopes of up to 250 Hz. This includes modulation frequencies that elicit temporal pitch cues, without explicitly imposing them as in the F0mod strategy. The further processing of the ACE strategy is identical to the further processing of the F0mod strategy.

Because the ACE strategy and the F0mod strategy differ both in the filter bank used and in the presence/absence of the explicit modulation, a third strategy is added that only differs from the ACE strategy with regard to the filter bank and only differs from the F0mod strategy in the absence of explicit modulation. This ACE512 strategy has the same filter bank as the F0mod strategy, based upon a 512-point FFT. However, in contrast to the F0mod strategy but similar to the ACE strategy, the extracted envelopes are not explicitly modulated at F0. This means that only modulations up to 62 Hz are present in the output channels and consequently only place pitch cues can be elicited by stimuli processed with the ACE512 strategy.

Experiment 1: F0 Discrimination of Single-Formant Stimuli

Methods

Stimuli

Fundamental frequency discrimination was measured for stylized vowels consisting of single-formant stimuli. The stimuli were a subset of the stimuli used in Laneau et al. [2004]. The stimuli were generated by filtering a 500-ms pulse train sampled at 16 kHz with a cascade of two filters. The first filter was a second-order infinite impulse response low-pass filter with a cutoff frequency of 50 Hz so that the output of the first filter resembled the glottal volume velocity. The second filter was a second-order infinite impulse response resonator that created a formant in the spectrum. The details of the filters can be found in Laneau et al. [2004]. The formant frequency was 300 Hz, 350 Hz, 400 Hz, 450 Hz or 500 Hz and the bandwidth of the formant was kept fixed at 100 Hz.

The fundamental frequency of the reference signal was 133 Hz and the fundamental frequencies of the comparison signals were 140 Hz, 150 Hz, or 165 Hz (corresponding to relative F0 differences of 0.9, 2.1, or 3.7 semitones, respectively). All stimuli were set to equal RMS power and then maximally amplified with a common gain while the signal was kept between –1 and 1 so that most of the stimuli optimally used the dynamic range of the compression block in the sound processing. The resulting RMS of the signals before processing was approximately 40% of the maximum input amplitude. All stimuli were processed offline with the three sound processing strategies (F0mod, ACE and ACE512), and subsequently stored on disk.

Procedure

Fundamental frequency discrimination was measured in a 2-interval, 2-alternative forced choice paradigm similar to the procedure in Laneau et al. [2004] using the method of constant stimuli. The subjects were presented with the reference stimulus and the comparison stimulus in random order and were asked to indicate the highest

Table 2. Relevant information about each of the subjects who participated in the experiments

Subject	Duration of profound deafness, years	Etiology	Implant type	Speech processor	Clinical speech processing strategy	Age years	Implant experience at the start of experiments years; months
S1	3	Ménière's disease	Nucleus CI24R(CS)	SPrint	ACE	47	7; 5
S2	11	hereditary	Nucleus CI24R(ST)	ESPrit	Speak	20	3; 1
S3	30	Usher syndrome	Nucleus CI24R(CS)	ESPrit 3G	ACE	65	0; 4
S4	3	unknown	Nucleus CI24R(ST)	ESPrit	Speak	23	4; 5
S5	30	progressive	Nucleus CI24R(CS)	SPrint	ACE	56	2; 0
S6	32	progressive	Nucleus CI24R(CS)	ESPrit 3G	ACE	50	1; 8

one in pitch. The F0 of the reference stimulus was 133 Hz and the F0 of the comparison stimulus was set at one of the three higher F0 frequencies; but within one trial, both the reference and the comparison stimulus had the same formant frequency. No feedback was presented to enforce subjects to use their intuitive sense of pitch and to prevent the subjects from using cues other than pitch (for instance loudness). Reference and comparison stimuli were separated by a 500-ms silent gap.

The trials were presented in separate blocks for each of the three sound processing schemes. Each block contained trials with all three comparison frequencies for all 5 formant frequencies, and within each block each trial was repeated four times. Different formant frequencies were included per block in order to prevent subjects from identifying and learning the reference stimulus based upon other cues. In total, 3 different blocks (for each strategy) consisting of 60 trials were presented to every subject. The order of the blocks was randomized and each block was repeated twice.

To reduce the influence of loudness cues, amplitude was roved by multiplying the amplitude of all stimuli upon each presentation with a random factor between 0.77 and 1, with amplitude expressed in clinical current units and relative to the threshold level.

Subjects

Six experienced postlingually deafened CI recipients implanted with a Nucleus CI24 implant participated in the present study. The subjects were paid for their collaboration. Some relevant details about the subjects can be found in table 2. All subjects are relatively good performers with their implant in daily life. Subject S1 used the LAURA implant before being reimplanted with a Nucleus CI24 implant 11 months before the experiments begun. All subjects except subject S3 had previous experience in psychophysical experiments with pitch perception. Subject S3 was a professional musician before becoming deaf.

Results

The average proportion of correctly ranked trials is shown in figure 3 for the three different sound processing schemes and as a function of frequency difference. The proportions were averaged over the five formant frequencies and over the 6 subjects.

A repeated-measures analysis of variance (ANOVA) was performed on the obtained proportions with the F0 difference, sound processing scheme, formant frequency, and the measurement run (first or second presentation of the block) as within-subject variables, using the lower-bound statistics. The results showed a significant effect of both sound processing scheme [$F(1) = 75.9$; $p < 0.001$] and relative F0 difference [$F(1) = 18.3$; $p = 0.008$]. There was also a significant interaction effect between sound processing scheme and relative F0 difference [$F(1) = 10.2$, $p = 0.024$]. No other within-subject factor or interaction of within-subject factors had a significant effect. There were significant between-subject differences [$F(1) = 1386.8$; $p < 0.001$].

Pairwise comparisons using the least significant difference showed that the subjects performed significantly worse with the ACE512 scheme than with the ACE and F0mod schemes [$t(179) = 9.5$, $p < 0.001$ for the F0mod scheme; $t(179) = 10.4$, $p < 0.001$ for the ACE scheme]. However, there was no difference between the results obtained with the ACE and F0mod schemes [$t(179) = 0.78$, $p = 0.470$].

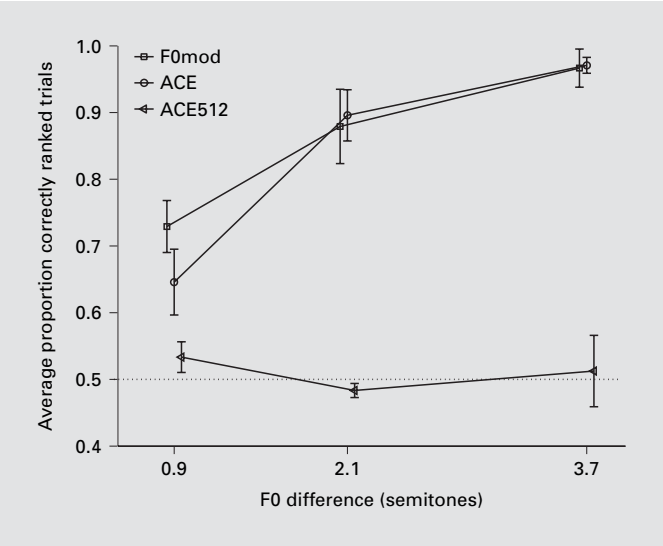


Fig. 3. Results for the F0 discrimination of single-formant stimuli with a reference frequency of 133 Hz. Each line shows the proportion of correctly ranked trials averaged over the six subjects as a function of relative F0 difference for the three different processing schemes. The dotted line indicates chance level. The error bars indicate ± one standard error of the mean over the six subjects.

Experiment 2: F0 Discrimination for Musical Tones

Methods

F0 discrimination was measured for notes of five different instruments and at three different reference fundamental frequencies. The musical notes were generated through a software MIDI generator based on the Creative Labs Live sound font wave table. The selected instruments were grand piano, clarinet, trumpet, guitar, and synthetic voice. The MIDI notes, dotted eighth notes, were played at 120 beats per minute. Consequently, the sustained portion of each note was approximately 400 ms. All recordings were truncated to 500 ms and set to equal RMS. All stimuli were processed with the three sound processing schemes offline and stored on disk.

F0 discrimination was measured for three reference F0 values: 130.8, 185.0, and 370.0 Hz (corresponding to the notes C3, F3# and F4#, respectively). The F0 values of the comparison signals were 1, 2 or 4 semitones higher (corresponding to approximately 6, 12 or 26% relative F0 difference).

The same subjects participated and the same procedures were used as in experiment 1. Upon each trial, the subjects were presented with a reference note (at one of the reference frequencies) and a comparison note (with higher F0) in random order and the subjects were asked to indicate the highest one in pitch. Both notes in a particular trial were played by the same instrument. The trials were presented in separate blocks for each combination of sound processing scheme and reference F0. Within each block, the three comparison F0 differences were presented for all five instruments and every trial was repeated four times. Consequently, this led to 9 different blocks consisting of 60 trials that were presented to each subject twice. Loudness roving was applied as in experiment 1.

Results

The average proportions of correctly ranked trials are shown in figure 4 for each sound processing scheme and as a function of relative F0 difference. Each panel shows the results for a different reference F0.

A repeated-measures ANOVA was performed on the proportions with five within-subject factors: sound processing scheme (F0mod, ACE, or ACE512), reference F0, musical instrument or timbre, relative F0 difference, and the measurement run (first or second presentation of the block) using the lower-bound adjustment for the degrees of freedom. The analysis indicated significant effects of the sound processing scheme [$F(1) = 9.87$; $p = 0.026$], and the relative F0 difference [$F(1) = 76.0$; $p < 0.001$]. There were also significant interaction effects between sound processing scheme and relative F0 difference [$F(1) = 8.12$; $p = 0.036$] and between the reference F0 and relative F0 difference [$F(1) = 7.05$; $p = 0.045$]. The interaction effect between sound processing scheme and reference F0 approached significance [$F(1) = 5.82$; $p = 0.061$]. The subjects differed significantly in their performances [$F(1) = 10{,}361$; $p < 0.001$].

The largest difference between the three sound processing schemes is found at the lowest reference frequency (130.8 Hz) shown in the left panel of figure 4. Paired t tests showed that for this reference F0 the ACE512 scheme produced lower scores than the ACE scheme [$t(179) = 3.84$, $p < 0.001$] and that the ACE scheme produced lower scores than the F0mod scheme [$t(179) = 5.92$, $p < 0.001$]. The ACE512 scheme, which only provides place pitch cues, had the poorest performance. However, the ACE and F0mod scheme, which provide both temporal and place pitch cues, enabled the subjects to discriminate F0 differences of 2–4 semitones.

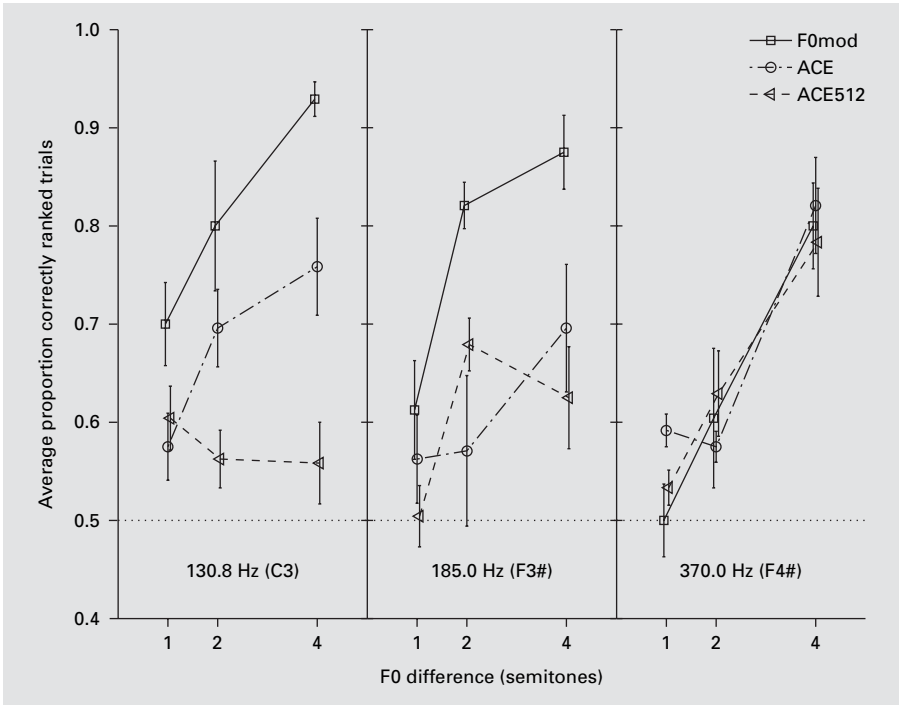


Fig. 4. F0 discrimination results for musical notes of five different instruments. The results are averaged over the instruments and the six subjects. Each line shows the proportion of correctly ranked trials as a function of relative F0 difference for the three processing schemes. The results for the reference frequency fixed at 130.8, 185.0, and 370.0 Hz are shown in the left, middle and right panel, respectively. The dotted line indicates chance level. The error bars indicate ± one standard error of the mean over the six subjects and five instruments.

At the middle reference F0 (185.0 Hz), a paired t test indicated no difference between the ACE and ACE512 schemes [$t(179) = 0.258$, $p = 0.797$]. However, the results with the F0mod scheme were significantly higher than with the other two schemes [$t(179) = 6.74$, $p < 0.001$; $t(179) = 7.48$, $p < 0.001$]. Only with the F0mod scheme was the average proportion of correctly ranked trials above 75%, even for relative F0 differences of 2 semitones.

At the highest reference F0 (370.0 Hz), paired t tests indicated no differences between any of the sound processing schemes. In this condition, all three sound processing schemes provided approximately the same level of performance to the subjects. On average, the proportion of correctly ranked trials increased monotonically with relative F0 difference and reached 75% at around 4 semitones.

Experiment 3: Melody Recognition

Methods

Although an improvement was found for F0 discrimination with the F0mod scheme in the first two experiments of the present study, it can be argued that it is questionable that this improvement will lead to an improvement in music perception in CI subjects. In the present experiment, we therefore compare melody recognition of familiar melodies with the ACE and the F0mod strategy in four CI subjects.

Stimuli

The most characteristic parts of 19 popular Flemish melodies (mostly nursery songs) were composed in MIDI format [Laneau, 2005]. The melodies were selected for their general familiarity and simplicity in rhythm. In order to remove rhythm cues for melody recognition, the rhythms of all melodies were adjusted so that all melodies contained 16 quarter notes. The notes of each melody were then transposed such that the median note of each melody was equal to F4# (370.0 Hz).

The MIDI melodies were rendered with the clarinet instrument at 120 beats per minute using the software synthesizer described in experiment 2. Each melody was rendered twice: in a high register (around F4#; 370.0 Hz) and a low register (around F3#; 185.0 Hz) for which all the notes were transposed one octave down. The recordings were truncated at a duration of 8.15 s. All stimuli were then processed with the ACE and F0mod sound processing schemes and stored on disk.

Procedure

Subjects S1, S2, S5, and S6 participated in the present experiment. They were presented with the names of the 19 melodies and were asked to indicate their familiarity

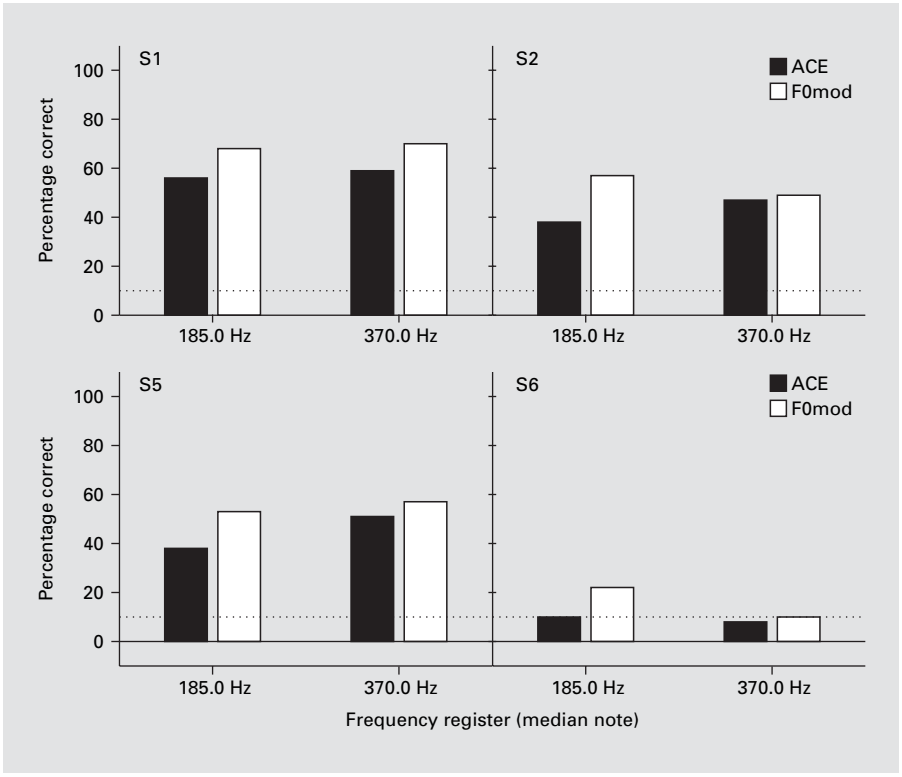


Fig. 5. Melody recognition of familiar songs without rhythmic cues. Each panel shows the results for a specific subject. The dotted line indicates chance level. Each bar is the result of 100 trials. The subjects scored significantly (at 0.05 level) above chance when the proportion was above 15%.

with each song. For each subject, the 10 most familiar songs were selected. The subjects were presented with blocks of melodies presented in random order and had to indicate the name of the melody from a closed set. Each melody was presented twice within each block, so that each block consisted of 20 melodies. Four different blocks were presented to the subjects for each combination of sound processing scheme (ACE or F0mod) and register (high or low) and the different blocks were presented in random order. Each block was presented five times to every subject spread over two or three test sessions. No feedback was given to the subjects.

At the beginning of each test session, the subjects were presented with each of their 10 most familiar melodies processed with the ACE scheme and with the original correct rhythm as an aid for the subjects to remember the melodies.

Results

Figure 5 shows the results for the melody recognition task averaged over the 5 runs and for each subject separately. The number of correctly identified melodies per block was analyzed using a repeated-measures ANOVA with three factors: sound processing scheme, register of the melodies (high or low), and measurement run number (1–5). There was a significant difference between the sound processing schemes [$F(1) = 100.1$; $p = 0.002$] but there was no difference between the registers [$F(1) = 0.124$; $p = 0.748$]. Moreover, a slightly significant effect of the measurement run [$F(1) = 10.343$; $p = 0.049$] was observed. There were no significant interaction terms. The subjects differed significantly in their performance [$F(1) = 15.98$; $p = 0.028$].

The subjects were able to recognize 9.9% more melodies with the F0mod scheme than with the ACE scheme in both the high- and the low-frequency register. Subjects S1, S2, and S5 scored relatively well on the melody recognition task with percentages of correct melody recognition ranging from 38% up to 70%. Subject S6 scored significantly worse and only scored significantly better than chance for the F0mod scheme in the lower register.

Subjects S1, S2, and S5 also exhibited a small training effect, as the number of correctly identified melodies increased monotonically for all conditions with the measurement run number. The average percentage of correctly identified melodies was approximately 25% higher in the fifth experimental run compared to the first.

Experiment 4: Musical Interval Identification

In the previous experiments, we showed that CI subjects could discriminate smaller F0 differences with the F0mod scheme and also that melodies could be recognized better. However, this does not mean that the pitches are correctly transmitted to the CI subjects. In this experiment, we test whether the distance between musical notes (musical interval) is correctly transmitted with the ACE and the F0mod strategy in a musically trained CI subject.

Methods

Subject S3, who was a professional organ and piano player before becoming deaf, participated in the present experiment. Interval identification was measured for frequency differences of –12, –10, –9, –7, –5, –4, –2, –1, 1, 2, 4, 5, 7, 9, 10, or 12 semitones (minor second, major second, third, fourth, fifth, sixth, seventh, and octave; both rising and falling in frequency). The subject was presented with a reference note (either C3, F3#, or F4#, corresponding to 130.8 Hz, 185.0 Hz, and 370.0 Hz) followed by a second note separated in frequency by one of the test intervals. The subject was then asked to indicate the interval separating both notes from the list of possible intervals. The intervals were presented in blocks and each block contained the 16 test intervals for the same reference note and processed with the same scheme. Each interval was repeated twice and the order of the intervals within one block was randomized. At the beginning of each block, five additional trials were presented (randomly selected from the trials in this block) to familiarize the subject with the presented processing condition prior to data collection. The results of these five trials were discarded for the analysis. Consequently, 6 different blocks (for both ACE and F0mod schemes with the three reference frequencies) consisting of 37 trials were presented to the subject. The blocks were presented four (for 130.8 Hz, and 370.0 Hz) or five (for 185.0 Hz) times to the subject. Blocks were always presented alternating between blocks processed with the ACE and F0mod scheme and all blocks with the same reference frequency were presented in sequence. No feedback was presented to the subject in order to force the subject to use his intuitive sense of musical intervals.

The intervals consisted of two dotted eighth notes rendered at 120 bpm with the software synthesizer as in experiment 2 and for the clarinet instrument. All notes were equalized in RMS and truncated to 500 ms. The intervals, consisting of a reference note and a comparison note with a 500-ms silent gap in between, were processed with the ACE and F0mod schemes. For the highest reference note, the upper limit of the F0 estimator of the F0mod scheme was set to 750 Hz to include the highest F0 frequency (740 Hz) of the comparison note.

Results

The average interval estimate is shown in figure 6 for each presented musical interval. In general, rising intervals were estimated as rising intervals, and falling intervals were generally estimated as falling intervals. Similarly, the estimated interval became larger as the presented interval was larger. However, the subject indicated a positive difference of approximately one octave in the lowest register for the F0mod scheme for the intervals of –12, –10 and –9 semitones. This large error between the presented interval and the estimated interval is due to the incorrect estimate of F0 in the F0mod scheme for the comparison note in these intervals. In these intervals, the F0 of the comparison notes were below the minimum F0 allowed in the F0 estimator. The results corresponding to these conditions were considered outliers and have been removed from further analysis.

The absolute errors made by the subject between the estimate of the musical interval and the presented interval were analyzed using Friedman's nonparametric two-way ANOVA. The absolute error is larger for the ACE scheme than for the F0mod scheme [$\chi^2(1) = 20.51$, $p < 0.001$]. Moreover, it is larger for the higher registers than for lower registers [$\chi^2(2) = 16.21$, $p < 0.001$].

The larger just noticeable differences in F0 with the ACE scheme are also reflected in a larger variance of the estimates with the ACE scheme. The average standard deviations of the estimates were 3.1 semitones and 2.1 semitones for the ACE and F0mod scheme, respectively.

Discussion

F0 Discrimination

With the F0mod scheme on the F0 discrimination task, the subjects reached a proportion of 75% correct answers between 1 and 2 semitones for stimuli with F0 below 250 Hz. For the ACE scheme, this level of perfor-

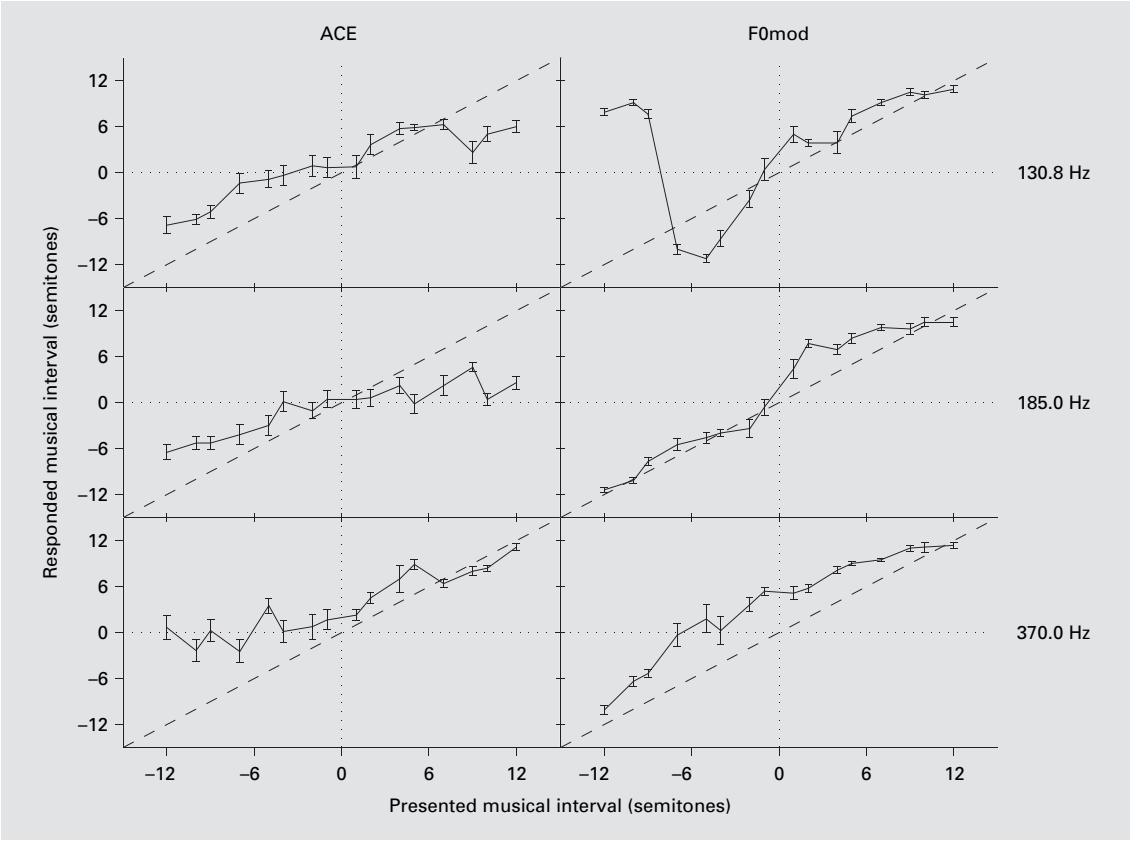


Fig. 6. Estimates of musical pitch intervals by subject S3 who was a professional musician before becoming deaf. The sounds processed with the ACE strategy and with the F0mod strategy are shown in the left and right columns, respectively. The reference note of the two-note interval for the estimates was 130.8, 185.0, and 370.0 Hz for the top, middle, and bottom row of panels, respectively. The dashed diagonal line in every panel indicates the relation expected for musically trained NH subjects. The error bars indicate ± one standard error of the mean.

mance was only reached for the single-formant stimuli. In fact, for the musical notes, the average slope of a normal cumulative distribution fit to the average proportions was 2.6 and 3.3 times higher for the F0mod strategy than for the ACE strategy for the reference F0 of 130.8 Hz and 185.0 Hz, respectively. This means that the just noticeable F0 differences are approximately 3 times smaller with the F0mod scheme than with the ACE scheme for musical notes below 250 Hz.

This large improvement in F0 discrimination performance with the new scheme for musical notes is in contrast with the results with single-formant stimuli (experiment 1) and with the results of Geurts and Wouters [2001] and Green et al. [2004]. In these experiments, no or only a minor improvement was obtained with the sound processing scheme providing enhanced temporal modulations. However, all these studies used synthetic vowels or synthetic diphthongs generated based upon the Klatt speech synthesizer [Klatt, 1980]. In this synthesizer, the stimuli are generated by filtering periodic pulse trains, a harmonic signal with all components of equal amplitude and in cosine phase. Qin and Oxenham [submitted] measured F0 discrimination with a noise band vocoder to simulate CI performance for various amounts of reverberation. They found the best performance for signals with all harmonics in cosine phase. However, when the harmonics were no longer in cosine phase, performance dropped significantly for vocoders with up to 8 channels. Consequently, the good performance of the ACE scheme in experiment 1 and the standard CIS schemes in Geurts and Wouters [2001] and Green et al. [2004] was probably due to the artificial nature of the stimuli in anechoic conditions. In this condition, the envelope modulations in the channels were already close to optimal to provide temporal pitch cues with the ACE scheme. Further enhancing the temporal pitch cues then did not result in large im-

provements. However, with more realistic signals, as with the stimuli in this experiment, the performance drops for the standard processing schemes as the modulation depth is reduced because of the temporal smearing. Consequently, the results of the subjects with the ACE scheme are better with the single-formant stimuli of experiment 1 than with the musical notes of experiment 2. This difference in performance is much smaller with the F0mod scheme because the modulation depth is preserved in all cases.

With the ACE512 scheme and at the highest reference F0 tested (370 Hz), the subjects did not use any temporal pitch cues because the stimuli did not contain temporal pitch cues or the modulations were above 300 Hz limiting their effectiveness. Without temporal pitch cues, F0 discrimination was poor. However, with increasing F0 the effectiveness of the place pitch cues for F0 discrimination improves because the performance of the subjects is better at 370.0 Hz for the ACE and ACE512 schemes than for the condition at 185.0 Hz. A similar effect was observed for the ACE scheme in Laneau et al. [2004].

Identical results were obtained with the 128-point FFT filter bank (ACE) and the 512-point FFT filter bank (ACE512 and F0mod) at the highest reference F0 tested, although the latter filter bank has more resolution in the lower frequencies. This is in contradiction with the results of Laneau et al. [2004] who found a benefit for the place pitch cues with filter banks having more resolution in the lower frequencies. However, in that study the difference in resolution in the F0 frequencies between the filter banks was larger than the difference in resolution between the filter banks of the present study. Moreover, in the study of Laneau et al. [2004], the fundamental frequency component had a large effect on the overall spectral shape of the stimuli because the formant frequencies were relatively low (between 300 Hz and 500 Hz). In the present experiment, the 'weight' of the fundamental on the overall spectral shape (or spectral centroid) was less prominent because the stimuli contained more harmonics.

Melody Recognition

All subjects scored better with the new F0mod scheme than with the standard ACE strategy, at least for melodies in the lower frequency register. Most probably, this is due to the improved F0 discrimination ability of the subjects when using the F0mod scheme. Gfeller et al. [2002] also found that subjects that were able to discriminate smaller F0 differences could recognize more melodies. For the high register, the improvement in melody recognition with the F0mod scheme probably depends on the improved F0 discrimination for the lower notes (down to 220 Hz) in the melodies.

The results of the present experiment show no difference between the melodies presented in the low register and the high register. This is in contrast with the results of Pijl and Schwarz [1995], who found a monotonic decay of melody recognition performance with increasing frequency register of the melodies. The melodies in that study were, however, solely presented by means of temporal pitch cues, while the note differences in the present experiment could be detected using both temporal and place pitch cues. The fact that subjects S1, S2 and S5 were able to recognize melodies in the high frequency register means that they were able to use place pitch cues for the task. This is definitely so for the ACE scheme where temporal pitch cues were not present because temporal modulations above 250 Hz were filtered out in the ACE scheme. Also, the effectiveness of temporal pitch cues is very limited above 300 Hz [Shannon, 1983; Zeng, 2002].

Musical Interval Estimation

Pijl [1997] reported that CI subjects were unable to label the intervals correctly for musical notes presented in sound field through the subjects' speech processor using the SPEAK speech processing strategy. Although the interval identification with the ACE scheme is relatively poor in the present experiment, it is better than the interval labeling performance in Pijl's report [1997]. This difference is probably due to the low pulse rate per channel in the SPEAK strategy (below 300 Hz), which is too low to correctly transmit the modulation frequency of the beatings in the channels [McKay et al., 1994].

It is also noteworthy that temporal pitch cues were only effective up to notes of approximately 185 Hz for the ACE scheme (because of the processing) and up to notes of approximately 370 Hz for the F0mod scheme (because the efficacy of temporal pitch cues decreases above 370 Hz; see experiment 2). Above these frequencies, only the place pitch cues in the stimuli were effective. For the ACE scheme, the place pitch cues were ineffective in eliciting the correct musical pitch difference between 185 Hz and 370 Hz as the two notes in the intervals were estimated to be similar in this frequency range. However, above 370 Hz, the place pitch cues were able to elicit mu-

sical pitch differences that moreover also resembled the presented musical pitch differences based upon F0. A similar result is observed for the F0mod scheme above 370 Hz where place pitch cues elicit a monotonically rising musical interval percept. However, in this case, the musical interval was consistently overestimated. McDermott and McKay [1997] have also reported that interval estimation is possible using solely place pitch cues and that the perceived interval rises monotonically with increasing spatial separation of the stimulation site.

Implications for CI Sound Processors

The present study indicates that providing explicit modulation of the channel envelopes at F0 with large modulation depth and in phase across channels, as in the F0mod scheme, improves (musical) pitch perception in CI subjects. This result is obtained from a comparison to conventional sound processing schemes where temporal pitch cues related to F0 are only present through beatings (of relatively shallow depth) originating from the interactions of the harmonics falling within the same analysis filter. Moreover, with the F0mod scheme the modulation depth at F0 is unaffected by reverberation or background noise (as long as the F0 estimation is performed correctly).

A number of issues still need to be resolved before an explicit F0 modulating scheme can be used clinically in the CI population. Firstly and most importantly, it must be assessed whether the proposed sound processing scheme does not have an adverse effect on speech intelligibility. Previous studies have shown that speech perception is improved in most subjects when the stimulation rate per channel is raised above 150 Hz [Fu and Shannon, 2000] or 250 Hz [Skinner et al., 2002a, b]. In the proposed scheme, the effective stimulation rate may be reduced to frequencies in the range of 100–300 Hz and consequently a negative effect on speech perception may be expected. A preliminary study of vowel identification with the scheme proposed by Green et al. [2004] showed that with the explicit F0-modulating scheme vowel recognition was worse than with standard CIS [Macherey, 2003]. But even with slightly reduced speech intelligibility with the F0mod scheme, it could still be very useful as a program in CI sound processors for listening to music.

A second issue that needs to be resolved before clinical application of the F0-modulating schemes has a more technical origin. A robust F0 estimation is essential for the correct functioning of the proposed scheme. Moreover, it has to be computationally inexpensive and has to have a low latency. Combining all these features in adverse circumstances has been proven difficult in the past.

Finally, although the enhanced envelope pitch cues in the F0mod scheme resulted in improved music perception, these cues are weak compared to pitch cues in NH subjects. Optimally, the salience of pitch cues related to resolved harmonics would be transmitted to CI subjects.

Mechanisms of Musical Pitch Perception

It has been shown that purely temporal pitch cues can convey musical pitch [Burns and Viemeister, 1976; Moore and Rosen, 1979; Pijl and Schwarz, 1995]. However, there is an ongoing debate whether purely place pitch meets the strict definition of musical pitch. Some researchers have argued that this is not the case [Moore and Carlyon, 2004]. Our results indicate that also purely place pitch cues can convey musical pitch. Firstly, three of the four subjects were able to recognize melodies without rhythm cues in the high register with the ACE strategy. In this condition, the lowest note was around 220 Hz, so for all notes temporal pitch cues were absent because of the envelope filtering in the ACE strategy. This means that most probably the subjects recognized the melodies purely based on place pitch cues. Secondly, in the interval estimation experiment, the musically trained subject S3 was able to approximately label the musical intervals for notes above 370 Hz. In this frequency region, temporal pitch cues are assumed to be completely absent in case of the ACE scheme, and at least to be less effective in case of the F0mod scheme.

Our results appear to indicate that purely place pitch cues can be used for musical pitch in some particular tasks, similar to purely temporal pitch cues. This does not mean that both cues elicit the same musical pitch percept. Most likely, this is not the case because of their psychophysical [McKay et al., 2000; Tong et al., 1983] and physiological [Warren et al., 2003] independence.

Conclusion

In this study, a new sound processing scheme for CI (F0mod) was proposed to enhance pitch and music perception by CI subjects. In this scheme, the envelopes of

all channels are modulated sinusoidally at F0 with 100% modulation depth and in phase across all channels.

With the new F0mod scheme, music perception was found to be improved significantly with respect to the current most often used sound processing strategy for Cochlear Nucleus recipients (ACE). This was evidenced by three benefits.

(1) F0 discrimination was three times better with the F0mod strategy compared to the standard strategy for musical notes of five different instruments when the F0 was below approximately 250 Hz.

(2) Subjects were able to recognize more melodies (without any rhythmic cues) with the F0mod sound processing scheme.

(3) The perceived pitch differences between musical notes were more in accordance with the musical scale as perceived by NH subjects.

There was, however, no clear benefit with the new sound processing scheme for F0 discrimination at 370 Hz and for interval labeling above 370 Hz where only place pitch cues were available. This means that there was no significant difference in perception of the place pitch cues between the new scheme and the standard scheme, at least for these conditions.

These findings indicate that the presented F0mod scheme is very likely to improve CI subjects' music perception or speech understanding in tonal languages.

Acknowledgements

We thank the subjects for their time and enthusiastic cooperation. 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.

References

Anantharaman JN, Krishnamurthy AK, Feth LL: Intensity-weighted average of instantaneous frequency as a model for frequency discrimination. J Acoust Soc Am 1993;94:723–729.

Burns EM, Viemeister NF: Non-spectral pitch. J Acoust Soc Am 1976;60:863–869.

Cariani PA, Delgutte B: Neural correlates of the pitch of complex tones. 1. Pitch and pitch salience. J Neurophysiol 1996;76:1698–1716.

Carlyon RP, van Wieringen A, Long CJ, Deeks JM, Wouters J: Temporal pitch mechanisms in acoustic and electric hearing. J Acoust Soc Am 2002;112:621–633.

Darwin CJ, Carlyon RP: Auditory grouping; in Moore BCJ (eds): Hearing. Handbook of Perception and Cognition. Orlando, Academic Press, 1995, vol 6, pp 387–424.

Drullman R, Festen JM, Plomp R: Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 1994;95:1053–1064.

Fu QJ, Shannon RV: Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners. J Acoust Soc Am 2000;107:589–597.

Geurts L, Wouters J: Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J Acoust Soc Am 2001;109:713–726.

Geurts L, Wouters J: Better place-coding of the fundamental frequency in cochlear implants. J Acoust Soc Am 2004;115:844–852.

Gfeller K, Christ A, Knutson JF, Witt S, Murray KT, Tyler RS: Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. J Am Acad Audiol 2000;11:390–406.

Gfeller K, Lansing CR: Melodic, rhythmic, and timbral perception of adult cochlear implant users. J Speech Hear Res 1991;34:916–920.

Gfeller K, Turner C, Mehr M, Woodworth G, Fearn R, Knutson J, Witt S, Stordahl J: Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. Cochlear Implants Int 2002;3:29–53.

Glasberg BR, Moore BC: Derivation of auditory filter shapes from notched-noise data. Hear Res 1990;47:103–138.

Green T, Faulkner A, Rosen S: Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants. J Acoust Soc Am 2004;116:2289–2297.

Holden LK, Skinner MW, Holden TA, Demorest ME: Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy. Ear Hear 2002;23:463–476.

Klatt DH: Software for a cascade-parallel formant synthesizer. J Acoust Soc Am 1980;67:971–995.

Kong YY, Cruz R, Jones JA, Zeng FG: Music perception with temporal cues in acoustic and electric hearing. Ear Hear 2004;25:173–185.

Lan N, Nie KB, Gao SK, Zeng FG: A novel speech-processing strategy incorporating tonal information for cochlear implants. IEEE Trans Biomed Eng 2004;51:752–760.

Laneau J: When the Deaf Listen to Music – Pitch Perception with Cochlear Implants; PhD thesis, Leuven, 2005 (http://hdl.handle.net/1979/57).

Laneau J, Moonen M, Wouters J: Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. J Acoust Soc Am 2004;116:3606–3619.

Leal MC, Shin YJ, Laborde ML, Calmels MN, Verges S, Lugardon S, Andrieu S, Deguine O, Fraysse B: Music perception in adult cochlear implant recipients. Acta Otolaryngol 2003;123:826–835.

Licklider JCR: A duplex theory of pitch perception. Experientia 1951;7:128–134.

Licklider JCR: Periodicity pitch and place pitch. J Acoust Soc Am 1954;26:945.

Macherey O: Evaluation of a Method to Improve Perception of Voice Pitch in Users of CIS Cochlear Implants; Master thesis, Paris, 2003.

McDermott HJ, McKay CM: Musical pitch perception with electrical stimulation of the cochlea. J Acoust Soc Am 1997;101:1622–1631.

McKay CM, McDermott HJ, Carlyon RP: Place and temporal cues in pitch perception: are they truly independent? Acoust Res Lett Online 2000;1:25–30.

McKay CM, McDermott HJ, Clark GM: Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. J Acoust Soc Am 1994;96:2664–2673.

Meddis R, Hewitt MJ: Virtual pitch and phase sensitivity of a computer-model of the auditory periphery. 1. Pitch identification. J Acoust Soc Am 1991;89:2866–2882.

Moore BC, Carlyon RP: Perception of pitch by people with cochlear hearing loss and by cochlear implant users; in Plack C, Oxenham AJ (eds): Pitch. Springer Handbook of Auditory Research. New York, Springer, 2004.

Moore BCJ, Rosen SM: Tune recognition with reduced pitch and interval information. Q J Exp Psychol 1979;31:229–240.

Oppenheim AV, Schafer RW: Discrete-Time Signal Processing, ed 2, Revised. Upper Saddle River, Prentice Hall, 1999.

Oxenham AJ, Bernstein JGW, Penagos H: Correct tonotopic representation is necessary for complex pitch perception. Proc Natl Acad Sci USA 2004;101:1421–1425.

Pijl S: Labeling of musical interval size by cochlear implant patients and normally hearing subjects. Ear Hear 1997;18:364–372.

Pijl S, Schwarz DW: Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. J Acoust Soc Am 1995;98:886–895.

Qin MK, Oxenham AJ: F0 discriminability and utility with acoustic simulation of cochlear implant signal processing. Ear Hear, submitted.

Schulz E, Kerber M: Music perception with the MED-EL implants; in Hochmair-Desoyer IJ, Hochmair ES (eds): Advances in Cochlear Implants. Vienna, Manz, 1993, pp 326–332.

Shannon RV: Multichannel electrical stimulation of the auditory nerve in man. 1. Basic psychophysics. Hear Res 1983;11:157–189.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M: Speech recognition with primarily temporal cues. Science 1995;270:303–304.

Skinner MW, Arndt PL, Staller SJ: Nucleus 24 advanced encoder conversion study: performance versus preference. Ear Hear 2002a;23:2S–17S.

Skinner MW, Holden LK, Whitford LA, Plant KL, Psarros C, Holden TA: Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. Ear Hear 2002b;23:207–223.

Tong YC, Blamey PJ, Dowell RC, Clark GM: Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant. J Acoust Soc Am 1983;74:73–80.

Warren JD, Uppenkamp S, Patterson RD, Griffiths TD: Separating pitch chroma and pitch height in the human brain. Proc Natl Acad Sci USA 2003;100:10038–10042.

Zeng FG: Temporal pitch in electric hearing. Hear Res 2002;174:101–106.

Zwicker E, Fastl H: Psychoacoustics: Facts and Models, ed 2, Revised. Berlin, Springer, 1999.