Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid

L. V. Straatman\textsuperscript{a)}
Department of Otorhinolaryngology, Head and Neck Surgery, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

A. C. M. Rietveld
Department of Linguistics, Radboud University Nijmegen, Erasmusplein 1, 6525 HT Nijmegen, The Netherlands

J. Beijen, E. A. M. Mylanus, and L. H. M. Mens
Department of Otorhinolaryngology, Head and Neck Surgery, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

(Received 13 June 2008; revised 27 June 2010; accepted 12 July 2010)

Cochlear implants are largely unable to encode voice pitch information, which hampers the perception of some prosodic cues, such as intonation. This study investigated whether children with a cochlear implant in one ear were better able to detect differences in intonation when a hearing aid was added in the other ear (“bimodal fitting”). Fourteen children with normal hearing and 19 children with bimodal fitting participated in two experiments. The first experiment assessed the just noticeable difference in F0, by presenting listeners with a naturally produced bisyllabic utterance with an artificially manipulated pitch accent. The second experiment assessed the ability to distinguish between questions and affirmations in Dutch words, again by using artificial manipulation of F0. For the implanted group, performance significantly improved in each experiment when the hearing aid was added. However, even with a hearing aid, the implanted group required exaggerated F0 excursions to perceive a pitch accent and to identify a question. These exaggerated excursions are close to the maximum excursions typically used by Dutch speakers. Nevertheless, the results of this study showed that compared to the implant only condition, bimodal fitting improved the perception of intonation.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3474236]

PACS number(s): 43.64.Me [RYL] Pages: 1884–1895

I. INTRODUCTION

Advances in implant technology and speech processing have significantly improved the performance of cochlear implants (CIs). Nevertheless, current CI systems are still largely unable to encode voice pitch information. Following the Continuous Interleaved Sampling (CIS) design principles (Wilson et al., 1991), these systems provide temporal envelope information but discard much of the temporal fine structure required to perceive fundamental frequency (F0) (Zeng, 2004). In the CIS sound processing strategy, the pulse rate of the stimulus is fixed and therefore unrelated to F0. Earlier speech processing strategies conveyed voice pitch by the rate of pulsatile stimulation, however, the superior speech perception obtained with CIS made these strategies obsolete. Fixed-rate stimulation pulses are amplitude-modulated by low-pass filtered envelope signals extracted from a filter bank. In CIS processors, the lower cut-off frequency for the bandpass filter with the lowest center frequency just reaches the higher F0 frequencies (Wilson et al., 2005), so most of the temporal fine structure related to F0 is discarded (Green et al., 2005). Furthermore, for most subjects pitch perception with electrical stimulation saturates at approximately 300 Hz (Shannon, 1983a). Exceptions have been noted, as summarized by Kong et al. (Kong et al., 2009). This saturation imposes an upper limit on the use of any remaining periodicity information and on the use of the envelope as an additional cue to intonation. On the basis of experiments carried out on normal hearing subjects listening to vocoder processed speech simulating the output of CIS, Green et al. found that the temporal envelope provided cues to F0 perception, but only for an F0 of up to approximately 200 Hz (Green et al., 2002). Spectral cues for F0 depend on a sufficient number of perceptually independent channels allowing place code for pitch. However, the number of channels in a CIS filterbank is small compared to estimates of the number of independent channels in the normal auditory system (Moore, 2003). Additionally it is further limited to about eight across the speech frequency range, probably due to intracochlear current spread limiting spectral resolution (Berenstein et al., 2008; Friesen et al., 2001; Green et al., 2002; Henry and Turner, 2003; Shannon, 1983b). As a result, CI users show a threshold for spectral envelope differences insufficient to reliably resolve even the first harmonics of F0 (Green et al., 2002; Henry and Turner, 2003; Berenstein et al., 2008).

Voice pitch information makes an important contribution to understanding speech. Previous studies suggest that voice...
pitch helps a listener segregate speech in noise (Assmann and Summerfield, 1990; Brokx and Nooteboom, 1982). F0 variation is the physical correlate of intonation, which is a major component of prosody. The perception of pitch movements in non-tonal languages, such as English, Dutch, and German, enables a listener to follow the ‘discourse’ in terms of focus (‘what is important’) and to make distinctions between questions and affirmations. It also functions to identify speaker gender and speaker identity. These functions were found to be limited in subjects with a CI (Cleary et al., 2005; Fu et al., 2005; Green et al., 2005). In tonal languages, such as Mandarin and Cantonese, lexical distinctions are based on pitch patterns, implying that the recognition of pitch patterns is essential for a listener to recognize at least a subset of the words in tonal languages. CI recipients are less able to accurately identify lexical tones and consequently lexical meaning, again due to their poor ability to extract voice pitch information (Ciocca et al., 2002; Lee et al., 2002; Peng et al., 2004), even though a minority of subjects perform exceptionally well if natural speech stimuli are used, possibly by using extraneous cues such as duration (Peng et al., 2004).

Several attempts have been made to improve pitch processing in CIs. Geurts and Wouters (Geurts and Wouters, 2004) devised a strategy in which the frequency of the first harmonic was extracted and used to steer the output of two discriminable electrodes in order to produce pitch percepts in between that of each electrode separately. When F0 temporal cues were eliminated from the stimuli, this place-pitch strategy allowed improved F0 detection. This feasibility study did not test speech perception.

Another approach adopted in several studies is to enhance periodicity cues to F0. In one study, the original waveform envelope was replaced with a 100% modulated sawtooth-like envelope on all channels (Green et al., 2004, 2005). Although CI users improved slightly in their ability to identify intonation patterns, vowel perception was worse. In a follow-up study (Hamilton et al., 2007) the enhanced envelope was applied to only one channel to limit the disruption of spectral information thought to have caused the decrement in speech perception found earlier. Contrary to expectations, pitch perception was poorer when the enhancement was applied to the most apical channel, however, for some of the subjects slight improvements were found when the most basal channel was used. Although no significant effects on vowel perception were found, the authors caution that the limited number of subjects precluded definite conclusions. Vandali and associates compared several ways to enhance temporal information (Vandali et al., 2005). One strategy intended to code fine temporal detail in each channel by providing stimuli corresponding in time and amplitude to positive temporal peaks in the band-pass filtered signals. Three other strategies provided deeper F0 modulation cues than the clinical strategy, some of which with extra attention to align the modulation cues across all channels. The CI users better ranked the sung vowels in pitch with the experimental strategies than with the clinical strategy, while no decrement in speech perception was observed. Instead of a sawtooth envelope, Laneau et al. imposed a 100% modulated F0-related sinusoidal envelope across all channels (Laneau et al., 2006). Implant users showed improved musical interval perception with this strategy, but F0 discrimination of single-formant stimuli was the same as with the clinical strategy.

In conclusion, some experimental strategies enhancing F0 place or periodicity cues have shown improved aspects of pitch perception. Although some investigators note that these improvements come with a cost, namely a decrement in speech perception, others do not report such a decrement. This leaves hope for a clinically successful implementation of these experimental strategies. However, as long as the fundamental limits to place coding and rate coding referred to above are not alleviated, it is reasonable to expect only modest improvements in pitch perception.

A completely different approach to augmenting pitch perception is to combine electrical and acoustic stimulation for subjects with some residual hearing, either with a hearing aid (HA) in the same ear (referred to as Electrical Acoustic Stimulation or “EAS” (von Ilberg et al., 1999; Turner et al., 2004) or a hearing aid in the non-implanted ear, generally referred to as “bimodal fitting” (Ching et al., 2004; Kong et al., 2005; Kong and Carlyon, 2007). It is assumed that the auditory system is able to combine the low frequency input from the HA with the high frequency information from the CI, even when stimuli are presented in different ears as in bimodal fitting. Acoustic stimulation does not have the same limitations encountered with currently available cochlear implant systems, namely large intra-cochlear channel interactions and temporal saturation of the electrically stimulated nerve. This is not to say that F0 perception is normal once acoustic stimulation is provided. Cochlear hearing loss may cause a decrement in temporal and spectral resolution which alters the processing of voice pitch (Moore and Carlyon, 2005). As a result, a significant increase of the F0 difference limen measured with complex harmonic tones has been found in young moderately hearing-impaired subjects (Moore and Peters, 1992).

In normal hearing, F0 detection primarily depends on resolved lower harmonics. However, for severely hearing-impaired listeners these harmonics may be unresolved, meaning that these listeners may have to rely more on amplitude modulations that occur at a rate equal to the F0, caused by the beating between the unresolved higher harmonics within each analysis file (Oxenham, 2008).

Studies have shown that speech recognition in noise improves when a CI is combined with a contralateral hearing aid (Ching et al., 2004; Kong et al., 2005; Dorman et al., 2008). Kong et al. (2005) investigated the effect of bimodal fitting on listener recognition of English sentences in the presence of a competing talker. A significant bimodal advantage was observed, even when acoustic stimulation alone did not lead to speech recognition (Kong et al., 2005). One hypothesis for this advantage is that temporal F0 cues delivered by the hearing aid enable improved perceptual segregation of speech from competing noises (Oxenham, 2008). However, both the role of F0 periodicity and the segregation account of EAS, and the benefit of bimodal fitting are under investigation. Kong and Carlyon (2007) used vocoder simulations to obtain data on subjects with normal hearing and found an
improvement in simulated bimodal speech recognition in noise even after the F0 cue had been removed from the low-pass filtered speech presented in the other ear. This was taken as evidence that voicing or spectral cues to the first formant may offer “glimpses” into important parts of the running speech signal (Assmann and Summerfield, 1990). Moreover, it was concluded that F0 is neither necessary nor sufficient to explain the improved speech recognition in noise. In contrast, Brown and Bacon (2009) demonstrated that, under vocoded simulated conditions, both F0 and the low-frequency amplitude envelope are useful cues for implant users. Similarly, significant benefit was found when the output of the vocoder was combined with a tone carrying F0 cues, compared to vocoder alone (Brown and Bacon, 2009). This benefit remained, even with a competing tone following F0 of a background speaker, which was interpreted as evidence against segregation and in favor of the “glimping” account of EAS and bimodal fitting benefit.

The benefits of bimodal fitting have also been reported for children with CIs. Holt et al. showed improved speech perception scores in the bimodal condition compared to CI-only and HA-only, particularly for speech understanding in noise with a +5 dB signal-to-noise ratio (Hearing-In-Noise Test Children’s Version) (Holt et al., 2005). Similarly, Lee et al. reported improved speech performance in noise in the bimodal condition compared to the CI-only condition (Lee et al., 2008). Obviously, tone-languages heavily depend on the correct reception of F0 information. Luo and Fu showed that in CI vocoder simulations with Chinese adult listeners, low-frequency acoustic information below 500 Hz significantly improved lexical tone recognition in speech-shaped noise (Luo and Fu, 2006).

Until now, it has not been investigated if low-frequency acoustic information also improves the perception of intonation in bimodally fitted children. In the present study, results from two experiments are provided to clarify how additional acoustic information affects F0 detection in bimodally fitted children. Experiment 1 used a two-interval same/different task adapted from O’Halpin et al. (2005) to establish the just detectable pitch movement for a bisyllabic nonsense word in which F0 was manipulated. In an attempt to include meaningful material in a more or less ecologically valid task, experiment 2 assessed the children’s ability to distinguish between questions and affirmations in acoustically manipulated simple Dutch utterances. Our hypothesis was that in the bimodal condition (CI plus hearing aid, “CI+HA”) the children would detect smaller pitch movements than in the CI-alone condition. For reference purposes, normative data were obtained from a group of children with normal hearing.

II. MATERIALS AND METHODS

A. Subjects

Children with normal hearing (NH) and implanted children were recruited for the study. The two groups were matched for age. All participants were Dutch native speakers except for two CI users. Although Dutch was not considered their first language, they both had sufficient command of the language and could understand the task. Approval for the study protocol was received from the Ethical Committee of the Radboud University Nijmegen Medical Centre (number 2007/090). Written informed consent was obtained from the subjects or from their parents if the child was under 16 years of age.

1. Subjects with normal hearing

To obtain reference values and to establish whether children in the age range of our test group were able to perform the task, 14 NH children aged between 6.8 and 16.7 years (mean age 10.1 years) were recruited among colleagues and friends of the researchers. Normal hearing was defined as the presence of otoacoustic emissions or, in the absence of otoacoustic emissions, as pure-tone air conduction thresholds of ≤15 dB HL at octave frequencies from 125 to 8000 Hz.

2. Subjects using a cochlear implant and a hearing aid

The experimental group comprised 19 implanted children whose age at implantation was below 16 years and whose age at testing ranged from 6.0 to 19.8 years (mean 12.1 years). All the children had been using their implant for more than one year. Experience with bimodal fitting varied from 1.0 to 8.7 years. Demographic data on each subject and details about the implant and hearing aid use are shown in Tables I and II. A hearing aid had been fitted to the non-implanted ear of each child. The free-field unaided and aided hearing thresholds in the implant and the non-implanted ear are also given in Tables I and II. Fitting of both CI and HA was evaluated by the same audiologist. Word recognition based on the number of correct phonemes identified by a listener was tested using the “Nederlandse Vereniging voor Audiologie” (Dutch Society of Audiology) open-set monosyllable word recognition test (Bosman, 1989). The mean percentage of correct phonemes in the three conditions was: CI-only 69% (SD 23%), equivalent to a word score of 47%, HA-only 24% (SD 25%), equivalent to a word score of 10%, and CI+HA 75% (SD 19%), equivalent to a word score of 56%.

B. General procedure of Experiments 1 and 2

Each child was tested in a sound-treated double-walled room. At the beginning of each experiment, the task was explained and examples of the stimuli were presented. Stimuli were presented via a loudspeaker placed one meter in front of the subject. The presentation level was chosen to be comfortably loud. It was checked if this level exceeded the infinite compression threshold of the child’s speech processor.

We selected stimulus with the highest RMS level and looped the stationary part of the vowel that received either sentence accent, word stress, or both. The loudest stimuli were found to be 68 dB SPL as measured with a Brüel & Kjaer Investigator 2260 sound level meter held at the level of the subject’s ear. Assuming an dynamic input range of 30 dB for the older processors and 40 dB for the newer processors (Davidson et al., 2009), and given the aided thresholds in
Table I, it follows that in nine subjects the loudest stimulus may have reached the threshold for infinite compression.

All the children completed experiments 1 and 2 and experiment order was randomized across subjects. In addition to random experiment presentation, the order in which the CI and CI+HA conditions were tested was randomized in each experiment.

C. Experiment 1: Just noticeable difference of fundamental frequency (F0) of the non-word “baba”

1. Speech materials

The stimuli consisted of the bisyllabic nonsense-word “baba” ([ˈbaːbaː]) manipulated according to O’Halpin et al. (O’Halpin et al., 2005). Two Dutch native speakers (a male aged 26 years and a female aged 25 years) produced the

<table>
<thead>
<tr>
<th>Table I. Demographic data and details on the implants and unaided and aided thresholds of the implanted ear.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II. Demographic data and details on the hearing aids and unaided and aided thresholds of the non-implanted ear fitted with a hearing aid.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>19</td>
</tr>
</tbody>
</table>
A pitch accent was induced in the first syllable by creating a peak in the F0 contour at the midpoint of the vowel. The speech processing software PRAAT (Boersma and Weenink, 2006). PSOLA provides high quality manipulated speech signals (Moulines and Charpentier, 1990). The onset F0 values of the first syllable produced by the male and the female speaker were set at 100 Hz and 200 Hz respectively. To replicate the F0 declination in natural speech, a linear F0 fall of 2.8 semitones was added to the second syllable. This is in accordance with the rules incorporated into a Text-To-Speech-system for Dutch that are based on typical Dutch pitch movements (Gussenhoven et al., 1999; Gussenhoven and Rietveld, 2000; Kerkhoff and Rietveld, 1995; Hart et al., 1999).

A pitch accent was induced in the first syllable by creating a peak in the F0 contour at the mid-point of the vowel (known as H* in the “autosegmental description of intonation” (Gussenhoven et al., 1999). From the F0 at onset (“L”), the pitch contour rose linearly to the midpoint “H*,” followed by a linear fall to the end of the first syllable (“L”). The difference in F0 at onset and at H* ranged from 0.85 semitones to 22.1 semitones (almost two octaves), with a step size of 0.85 semitones (Fig. 1). The duration of the first syllable as produced by both speakers was approximately 200 ms. To create the pitch accent, the first 200 ms of the stimuli were manipulated. Total durations of the male utterance and the female utterance were 540 ms and 590 ms, respectively. Introducing pitch movements may have altered the loudness of the stimuli and thus introduced a cue into the discrimination task. To reduce this possibility, the peak amplitude of each manipulated utterance was scaled to the same value (95% of the maximum amplitude of the sound buffer of the computer system).

FIG. 1. Examples of manipulated F0 contours of the ‘baba’-bisyllables, recorded from native Dutch speaker (male). The black bold line represents the unstressed or base stimuli. The gray lines represent the different contours of the peak of the accentuated stimuli.

word 15 times, each time attempting a flat intonation. For each speaker, recordings were selected that sounded the most monotone. From this preselection, we selected recordings with only minimal differences in length between the first and second syllable to minimise word stress cues and sentence accent. The recordings were digitized at a sampling frequency of 44 kHz (16 bit resolution). Pitch contours of these bisyllabic words were manipulated using the Pitch Synchronous Overlap and Add (PSOLA) technique implemented in the speech processing software PRAAT (Boersma and Weenink, 2006). PSOLA provides high quality manipulated speech signals (Moulines and Charpentier, 1990). The onset F0 values of the first syllable produced by the male and the female speaker were set at 100 Hz and 200 Hz respectively. To replicate the F0 declination in natural speech, a linear F0 fall of 2.8 semitones was added to the second syllable. This is in accordance with the rules incorporated into a Text-To-Speech-system for Dutch that are based on typical Dutch pitch movements (Gussenhoven et al., 1999; Gussenhoven and Rietveld, 2000; Kerkhoff and Rietveld, 1995; Hart et al., 1999).

A pitch accent was induced in the first syllable by creating a peak in the F0 contour at the mid-point of the vowel (known as H* in the “autosegmental description of intonation” (Gussenhoven et al., 1999). From the F0 at onset (“L”), the pitch contour rose linearly to the midpoint “H*,” followed by a linear fall to the end of the first syllable (“L”). The difference in F0 at onset and at H* ranged from 0.85 semitones to 22.1 semitones (almost two octaves), with a step size of 0.85 semitones (Fig. 1). The duration of the first syllable as produced by both speakers was approximately 200 ms. To create the pitch accent, the first 200 ms of the stimuli were manipulated. Total durations of the male utterance and the female utterance were 540 ms and 590 ms, respectively. Introducing pitch movements may have altered the loudness of the stimuli and thus introduced a cue into the discrimination task. To reduce this possibility, the peak amplitude of each manipulated utterance was scaled to the same value (95% of the maximum amplitude of the sound buffer of the computer system).

2. Procedure

An adaptive two-alternative forced-choice same/different task was used. In each trial, two “baba” non-words were presented, separated by a silent interval of 500 ms. In half of the trials, only one of the words (randomly chosen) contained an accent, while in the other half of the trials neither of the words had a pitch accent. Each trial was run twice in a random order across the subjects: once with the male version and once with the female version. Subjects were asked to press the left or right arrow button on the keyboard that had been labeled with the word “same” and a drawing of two identical elephants (left) or “different” and a drawing of two dissimilar elephants (right). Visual feedback was provided. At the beginning of each test, a practice run of 16 trials was presented. All the implanted children used their cochlear implant and their hearing aid during the practice runs. Both devices were set at the normal everyday setting.

Thresholds were obtained using a 2-down 1-up staircase procedure (Levitt, 1971). After an incorrect response, the height of the peak of the accentuated stimulus was increased by one-step. After two correct responses, the height was decreased by one-step. A test run proceeded until one of the following occurred: (1) ten reversals, (2) eight successive incorrect responses at the maximum stimulus difference, or (3) eight successive correct responses at the minimum stimulus difference. The F0 threshold was estimated from the mean of the F0 differences of the final six reversals.

D. Experiment 2: Discrimination between questions and affirmations

Stimuli were six isolated bisyllabic or tri-syllabic words, all names of fruit, spoken with an affirmative intonation by a Dutch native speaker (a man aged 57 years) at an average F0 of 130 Hz. Interrogative versions of these words were created by imposing F0 patterns typical of questions in Dutch (and ‘exaggerated’ versions of them) on the last syllable of the utterance. None of the other markers of questions vs. affirmations were changed (e.g., duration, spectral pattern, overall F0 level).

1. Speech materials

To create the stimuli, three words were selected that carry stress on the first syllable: “aardbei” ([a:rtbɛi]; strawberry), “paprika” ([ˈpaːprɪkaː]; pepper) and “mango” ([ˈmaŋɡoː]; mango), and three words were selected that carry stress on the final syllable: “citroen” ([ˈsiːtrun]; lemon), “banaan” ([bəˈnaːnː]; banana) and “meloen” ([ˈmɔlunː]; melon). As described above, the pitch contours were manipulated using the PSOLA technique in PRAAT. All the original productions were provided with a linear falling pitch contour (“LL%”) of 6 semitones that is typical of Dutch utterances, which yielded the affirmative stimulus. While attempting to provide the affirmative stimulus with a rising intonation typical for questions (“LHH%”), it was found that the naturalness of the stimuli was best preserved when only the last part of the F0 contour was manipulated. Rules developed in the Text-To-Speech-system for Dutch, mentioned above, were adopted. They can be summarized as follows: F0 rise and fall...
have a duration of 100 ms and the final falling movement (to L%) and rising movement (peak at H%) have a duration of 20 ms; and the starting locations were determined from the end of the words. In compliance with the rules of the Text-To-Speech-system, the start of the rise or fall for words ending with an [n] and for those with a low amplitude in the last 120 ms was moved to the beginning of the word. This is because it is known to be more difficult to perceive the LHH% pitch rise when the amplitude is low or during a nasal resonance.

The total duration of the stimuli varied from 440 ms to 580 ms. Four questions were created (Q1, Q2, Q3, and Q4), each with increasingly higher F0 rises at the end of the utterances. Figure 2 shows an example of the pitch contours of the Affirmation and the four question variants. Table III shows the increases in F0 as expressed in semitones and in percent. The step size between the four categories of questions was three semitones. In Dutch, LHH% rises that signal questions range from 6 to 9 semitones. Thus, the peak amplitude of each manipulated utterance was scaled to the same value (95% of the maximum amplitude of the sound buffer of the computer system).

### 2. Procedure

The stimuli were presented in blocks of 48 items containing 24 affirmations (6 tokens repeated 4 times) and 24 questions (6 tokens each from the Q1, Q2, Q3 and Q4 categories). All six words were used equally often as an affirmation and as a question. Each word was also used equally often in each of the categories Q1-Q4. Each block of 48 stimuli was presented twice. The stimuli were randomly selected without replacement.

In each trial, the subjects heard a single word and were required to identify it as either a “question” or an “affirmation” by pointing at a drawing of a person who was gesturing a question, or a person who was making an affirmative gesture. Visual feedback was provided. Each test started with a practice run of 12 different stimuli. All the implanted children used their cochlear implant and their hearing aid during the practice runs. Both devices were set at the normal everyday setting.

### III. RESULTS

#### A. Experiment 1: Just noticeable difference in fundamental frequency (F0) of the non-word “baba”

1. **Subjects with normal hearing**

   Figure 3 shows the results of the NH subjects pooled over the male and female speaker. The mean F0 differences at threshold were 3.3 semitones (21% above baseline) with an SD of 2.7 semitones for the male speaker and 4.3 semitones (28% above baseline) with an SD of 4.2 semitones for the female speaker. A paired-samples t-test (SPSS version 13.0) did not show a significant difference in the responses to the male and female speakers ($t_{13}=2.7$, $p>0.12$, $\eta^2_{\text{partial}}=0.172$, observed power=0.332). The F0 difference at threshold, expressed in semitones, decreased with the logarithm of the age of the subject ($F_{1,13}=11.7$; $p<0.01$, $R^2=0.493$). Figure 3 indicates that the performance of children older than 10 years was better than that of most of the younger children. This suggests that the test may have been too difficult for some of the younger children. The average F0 difference at threshold in the children aged 10 years and older was 1.5 semitones (SD=0.4).

2. **Subjects using a cochlear implant and a hearing aid**

   Figure 4 shows the mean F0 difference at threshold in the CI-only and CI+HA conditions compared to the NH
group. In response to the male speaker, thresholds improved on average from 72% to 49% F0 difference. In response to the female speaker, thresholds improved from 9.4 semitones to 6.9 semitones (CI + HA) (i.e., a decrease from 72% to 49% F0 difference). Independent samples t-test showed a significant improvement in performance between the NH group compared to the implanted children in the CI-only condition (t_{31} = 3.2; p < 0.01; \eta^2_{partial} = 0.246). However, the scores of the NH group were not significantly better than the implanted children wearing their CI+HA (t_{31} = 1.9; p > 0.05; \eta^2_{partial} = 0.101; observed power = 0.439).

Figure 5 shows the individual F0 difference thresholds in the CI-only condition compared to the CI+HA condition. Most data points were on or under the diagonal line. Figure 5 shows that in the CI-only condition, the performance of 42% of the implanted subjects (8 out of 19) was within 1 SD of the average of NH subjects. In the CI+HA condition, this percentage improved to 74% (14 out of 19) and all but three CI subjects performed within 2 SD from the average of NH subjects. A repeated measures analysis of variance with two within-subject fixed factors, namely Condition (CI-only and CI+HA) and Speaker (male and female), showed significantly better F0 discrimination in the CI+HA condition than in the CI condition (F_{1,18} = 6.2; p < 0.02; \eta^2_{partial} = 0.26; observed power = 0.65). The effects of Speaker (F_{1,18} = 1.4; p > 0.24; \eta^2_{partial} = 0.07; observed power = 0.21) and the interactions between Speaker and Condition (F_{1,18} = 0.55; p > 0.46; \eta^2_{partial} = 0.03; observed power = 0.11) were not significant.

Linear correlations (Pearson), nonparametric correlations (Spearman & Kendall) and curvilinear correlations were calculated between the F0 difference thresholds and (1) age, (2) logarithmic function of age, (3) the pure tone average PTA, i.e., mean threshold at 0.5, 1, 2 and 4 kHz, (4) average aided threshold with the hearing aid, (5) average aided threshold with the CI, (6) aided thresholds at 0.25 kHz, (7) duration of deafness, (8) duration of CI use, (9) duration of hearing aid use, (10) age at implantation, (11) age of deafness, (12) speech perception scores in quiet (i.e., phoneme or word scores) in HA-only, CI-only, and the bimodal condition. Of these, only one nonparametric correlation was found to be significant. F0 difference thresholds were marginally better for patients that were older at implantation (Kendall’s tau: -0.283, p = 0.046).

B. Experiment 2: Discrimination between questions and affirmations

1. Subjects with normal hearing

All the normal hearing subjects participated in two blocks of 48 trials. The mean correct score was 81% (SD 15%). Their individual scores are shown in Fig. 6. A signifi-
significant Pearson correlation coefficient was found between age and intonation discrimination ($r_{11}=0.58; p<0.05$, one-tailed). However, it was clear that this correlation was mainly caused by the poor scores of the three youngest children (≤8.3 years). Based on the binomial distribution, each child’s performance was statistically compared to chance levels, with a criterion level of 64% correct ($\alpha=0.05$, one-sided). Performance was significantly better than chance for 11 subjects but not for the three youngest subjects (aged 6, 7 and 8 years).

Figure 7 shows the mean correct scores on the five categories of the manipulated utterances. As was expected, the higher the end-pitch rise, the higher the score. As the implanted group had missing values in experiment 2, a linear mixed model analysis was carried out with two within-subject fixed factors: End-pitch contour (affirmation, Q1, Q2, Q3 and Q4) and Block (first and second blocks). Utterances with higher end-pitch rises were more often correctly classified as questions ($F_{4,117}=26.2; p<0.001$). In contrast, the effect of Block ($F_{1,117}=1.8; p=0.158$) and the interaction between Block and End-pitch contour ($F_{4,117}=0.39; p>0.81$) were not significant. Pairwise comparisons of end-pitch contour using Bonferroni adjustment for multiple comparisons showed one significant contrast: Q1 yielded significantly fewer correct responses than all other categories ($p<0.001$ in each case).

2. Subjects using a cochlear implant and a hearing aid

An experienced pediatric audiologist judged that 2 out of the 19 children (both 6 years of age) lacked concentration during the practice run. The results of these two subjects were therefore excluded from analysis. Due to lack of time or concentration, 9 of the 17 remaining subjects were unable to participate in the second block of the test. A bimodal benefit of 62% (CI+HA) against 56% (CI-only) was found for the 17 subjects who completed the first block, and 73% (CI+HA) against 62% (CI-only) for the repeated presentation (8 subjects). The higher scores in the second block suggested a learning effect. In Fig. 8, the individual scores obtained in the CI-only condition are compared to the scores in the CI+HA condition. On the basis of the binomial distribution, performance was above chance for 5 of the 17 subjects (29%) in the CI-only condition and for 9 of the 17 subjects (53%) in the CI+HA condition (criterion level is 64% correct, $\alpha=0.05$, one-sided).

A linear mixed model analysis was carried out with three fixed within-subject factors: Condition (CI and CI+HA), Block (first and second blocks) and End-pitch contour (Affirmation, Q1, Q2, Q3 and Q4). Significantly better scores were found in the CI+HA condition compared to the CI-only condition ($F_{1,213.33}=6.5; p<0.02$). The effect of End-pitch contour were also significant ($F_{4,213.33}=12.3; p<0.001$). In contrast, the effects of Block and the interactions between Condition and Block ($F_{2,213.33}=0.313; p>0.57$), Condition and End-pitch ($F_{4,213.33}=0.517; p>0.72$), Block and End-pitch ($F_{4,213.33}=0.838; p>0.50$) and the three-way interaction between Condition, Block and End-pitch ($F_{4,213.33}=1.425; p>0.23$) were not significant.

When the 9 subjects who did not complete both blocks were excluded, approximately the same results were found. Again, significantly better scores were found in the CI+HA condition ($F_{1,133}=7.4; p<0.01$). The effect of End-pitch contour was also significant ($F_{4,133}=11.8; p<0.001$). As before, the effect of Block and the interactions between Condition and Block ($F_{1,133}=0.004; p>0.95$), Condition and End-pitch ($F_{4,133}=0.434; p>0.78$), Block and End-pitch ($F_{4,133}=0.526; p>0.72$) and the three-way interaction between Condition, Block and End-pitch ($F_{4,133}=1.40; p>0.24$) were not significant.

Additionally, an independent sample t-test comparing the 8 subjects who completed both blocks and the 9 subjects...
who did not, showed no significant differences with respect to: (1) age, (2) age at implantation, (3) age at deafness, and (4) question-statement identification scores of the first block. As a result of the similar scores obtained in the group which completed both blocks and the results of the entire group, the results were pooled for further analysis.

Linear correlations (Pearson), nonparametric correlations (Spearman & Kendall) and curvilinear correlations were calculated for the whole group between intonation identification and (1) age, (2) logarithmic function of age, (3) the pure tone average (PTA, i.e., mean threshold at 0.5, 1, 2 and 4 kHz), (4) average aided threshold with the hearing aid, (5) average aided threshold with the CI, (6) aided thresholds at 0.25 kHz, (7) duration of deafness, (8) duration of CI use, (9) duration of hearing aid use, (10) age at implantation, (11) age of deafness, (12) speech perception scores in quiet (i.e., phoneme or word scores) in HA-only, CI-only, and the bimodal condition. No significant correlations were found.

Pairwise comparisons of the end-pitch contour on the basis of Bonferroni adjustment for multiple comparisons showed that significantly fewer Q1 utterances were correctly identified than Affirmations ($t_{4,213.33} = 5.8; p < 0.001$), Q2 ($t_{4,213.33} = 5.2; p < 0.001$), Q3 ($t_{4,213.33} = 5.5; p < 0.001$) and Q4 utterances ($t_{4,213.33} = 5.5; p < 0.001$). This effect was significant in both the CI-only and the CI+HA conditions (Fig. 7).

In the CI-only condition, only 24% of the implanted children (4 of the 17) scored within 1 SD of the mean of the NH children. In the CI+HA condition, this percentage increased to 47% (8 out of the 17) and all but 3 implanted children performed within 2 SD of the mean of the NH children. Independent samples t-tests showed a significantly better performance by the NH group compared to the implanted children in the CI-only condition ($t_{10} = 5.4; p < 0.001; \eta^2_{partial} = 0.50$), and the CI+HA condition ($t_{10} = 3.5; p < 0.01; \eta^2_{partial} = 0.29$).

C. Correlation between F0 perception and discrimination of questions and affirmations

A one-tailed Pearson correlation test showed that subjects with better F0 thresholds in Experiment 1 were also better in classifying questions and affirmations in Experiment 2 (condition CI-only: $r_{17} = -0.45; p < 0.05$, condition CI+HA: $r_{17} = -0.44; p < 0.05$). There was also a significant correlation between the bimodal benefits (i.e., the difference between CI score and CI+HA score) in both experiments ($r_{17} = 0.53; p < 0.05$).

IV. DISCUSSION

In the present investigation, bimodal fitting significantly supported the perception of intonation compared to using CI-only. Both the discrimination of F0 excursions signaling pitch accents and the ability to use F0 patterns to identify questions versus affirmations improved significantly. In both tasks, an F0 excursion of about 6 semitones was required to be detectable in the CI+HA condition, which was similar to the excursion needed by the poorer performing NH control subjects. The performance in the CI+HA condition was sufficient to detect only the largest F0 excursions normally produced by Dutch speakers. It is reasonable to expect decreased performance in less than optimal listening situations when reverberation or (low frequency) noise may mask F0 (Lavandier and Culling, 2008).

A. Just noticeable difference in fundamental frequency

The first experiment assessed the discrimination of fundamental frequency excursions used to signal a pitch accent. The addition of the hearing aid was of significant benefit to the implanted children. The JND improved from 9.4 (CI) to 6.9 semitones (CI+HA) in the test with the male speaker, versus 11.3 (CI) to 7.5 semitones (CI+HA) in the test with the female speaker. The effect of Speaker was not significant, but it was in the same direction as that reported by others (Green et al., 2004; Laneau and Wouters, 2004; Chatterjee and Peng, 2008). Discrimination thresholds of the NH children (male speaker: 3.3 semitones, female speaker: 4.3 semitones) were almost half those of the CI subjects. In other words, the overall performance of the implanted children was poor.

Several studies on cochlear implant users reported JNDs for F0 changes that were considerably smaller than those found in the present study (Rogers et al., 2006; Cleary et al., 2005; Geurts and Wouters, 2001; Laneau et al., 2004), but considerable differences may well explain the discrepancies. In the first place, there were differences in subject population (i.e., adults vs. children). Psychophysical studies have shown that frequency discrimination thresholds in young children are higher than in adults. Halliday et al. (2008) measured frequency discrimination thresholds in adults and children. The performance of the children was clearly poorer than that of the adults (Halliday et al., 2008). In the present study better F0 difference thresholds were found for subjects that were older at implantation. This seems to be an unexpected finding. However, age at implantation was highly correlated with age at testing, possibly indicating that older subjects were better at this test, as expected from the results for frequency discrimination. Furthermore, this correlation was weak and non-significant after a Bonferroni correction. Additionally, differences in speech materials may also explain the discrepancies in the results. For instance, Geurts and Wouters (2001) measured JNDs of 0.7 to 2.2 semitones in four adult cochlear implant users with an F0 reference stimulus of 150 Hz in steady-state synthetic vowels (for a reference stimulus of 250 Hz, results were considerably poorer). Instead of a steady state F0, we used stimuli with rapidly changing F0 contours. The JND for F0 differences between steady state complex harmonic stimuli can be less than 1% in normal hearing subjects (Klatt, 1973). If rapidly changing F0 contours are presented to normal hearing subjects, a JND of 19% or 2.3 semitones can be found (‘t Hart, 1981), which is similar to the JND of 1.6 semitones for normal hearing adults reported by van de Sandt (2008) who used the same materials and procedure as in the present study.

The results of our NH subjects suggest that children younger than 10 years could not reliably perform the F0
discrimination task. Surprisingly, this was not observed in the implanted group, which may have been due to their greater familiarity with auditory testing in general. Additionally, it must be taken into account that this effect might be absent due to the variability in the implanted group in duration of deafness, duration of CI use, age at implantation, age of deafness, amount of residual hearing, or a combination of these variables. However, no correlations were found between these possible sources of variation and test performance.

B. Discrimination between questions and affirmations

In the second experiment, the average results of the implanted children show a modest but significant benefit of bimodal fitting for discriminating between questions and affirmations. The percentage correct improved significantly from 56% (CI) to 62% (CI+HA) in the first presentation and from 62% (CI) to 73% (CI+HA) in the repeated presentation.

Our results differ somewhat from those reported by Green et al. (2005) who reported mean percentage correct scores of 69% and 81% in the CI-only condition in response to a male speaker and a female speaker, respectively. In Green et al.’s experiment, the subjects were all postlingually deaf adults instead of the prelingually deaf children that we tested. A second explanation for the difference in results is that in the study by Green et al., the pitch rises were slower (at least 200 ms) than our study (120 ms). This difference relates to the discrimination of steady state versus rapidly changing F0 contours. Indeed, White and Plack (White and Plack, 1998) found that F0 discrimination improved with increasing duration of the pitch rise.

Another important difference is that Green et al. used naturally spoken affirmations and questions; whereas, our study manipulated the last part of the F0 contour. Therefore, other cues may have been available to their subjects besides F0, such as differences in formant spectrum, intensity, and duration. We found significant correlations in the range of 0.4 and 0.5 between F0 discrimination and question/statement identification in the CI-only and the CI+HA condition. In support of our hypothesis, the results suggest that subjects used F0 information rather than other cues in the question versus affirmation task.

In the current study, question/affirmation identification was assessed using five different utterances (viz. one affirmation and four different question categories: Q1, Q2, Q3, Q4). The increasing final rise in pitch across the question categories Q1-Q4 resulted in increasingly improved identification, but the largest difference in performance was between Q1 and the other stimuli. This effect was present in the NH and implanted children. It appears that the rising F0 contour that spanned 6–9 semitones in Q1 was too small to be perceived as a question by most of the subjects and was therefore frequently identified as an affirmation. This is in agreement with the JND of 6.9 semitones in the implanted subjects in experiment 1. As a rising contour of approximately 6 semitones is typical of Dutch questions, this result suggests that even when implanted children use a contralateral hearing aid, they require exaggerated F0 excursions to identify a question, particularly when other cues, such as duration, are lacking. In contrast, NH subjects should have been able to hear the rising contour of Q1 as they had a mean JND of 3.3 semitones. It is not clear why Q1 was more often perceived as an affirmation, but it may have been due to either a lack of duration cues or the complexity of the task.

Chatterjee et al. (Chatterjee and Peng, 2008) also reported that implanted adults required exaggerated F0 excursions to identify a question, and even with the most exaggerated F0 excursion, not all utterances were identified as questions by their subjects. The overall performance of their group was better than in our study, however the subjects in Chatterjee et al.’s study were all adults and the majority of their subjects (8 out of the 10) were postlingually deaf.

C. The effect of age on the perception of intonation

In the implanted children, no significant correlations were found between age at testing and performance in the two experiments. In contrast, these correlations were present in the group of NH children. They showed particularly large spread in F0 discrimination thresholds at the age of 9 years and younger. This spread ranged from nearly as good as the older children to much poorer. On the question/affirmation identification task, the three youngest children in the NH group (≤8.3 years) performed below chance level. This may be related to not understanding the task, lack of concentration, or a small attention span. A possible explanation for the lack of correlation between age and performance in the CI group is their greater familiarity with auditory testing. Alternatively, perhaps not all implanted children, even the older ones, could master the task; any bimodal benefit, if present, would only show in older children after they have developed sufficient skills for the task. Furthermore, comparison of Figs. 3 and 4 shows more variability in the results of implanted children compared to those of older NH children. This may have obscured any effect of age.

The results for implanted children in the first experiment indicate no significant correlations between F0 JND on the one hand, and the aided thresholds, duration of deafness, duration of CI use, and duration of HA use on the other. In addition, no significant correlations were found between F0 JND and age at implantation or age of deafness. These findings suggest that possibly a combination of these factors, or factors currently not investigated are responsible for the variability in performance [e.g., language ability or neural survival, placement of the electrodes, educational or learning variables (Cleary et al., 2005; Peng et al., 2004)]. Future research should consider these issues.

D. Benefit of the hearing aid for the perception of intonation

In contrast with previous studies that assessed prosody perception in implanted children who depended solely on their CI, the present study investigated the effect of bimodal fitting. Compared to the CI-only condition, performance in the CI+HA condition improved significantly in experiments 1 and 2. As discussed in the introduction, acoustic stimula-
tion provides access to several F0 cues that are not available to the same extent when using a CI-only. These cues include additional low-frequency periodicity information, resolved lower harmonics of F0, and amplitude modulations due to the beating of unresolved higher harmonics within channels. A prerequisite for the use of these cues is audibility. The average aided thresholds were about 50 dB SPL up to 1000 Hz, 55 dB SPL at 2000 Hz, and 75 dB SPL at 4000 Hz. The range between subjects suggests that after amplification, the louder speech segments were audible for most subjects, however, for many subjects frequencies of 1000 Hz and higher may have been relatively soft in the acoustically stimulated ear. This suggests that the severely hearing-impaired subjects may have relied on F0 periodicity information, the presence of resolved lower harmonics, or both periodicity information and lower harmonics. However, we cannot exclude the beating of higher harmonics of up to 1000 Hz as an additional cue to F0. The fact that we did not find a correlation between aided thresholds and bimodal benefit provides some support to the notion that low frequency cues, rather than beating within higher frequency channels, dominate performance.

Part or all of the superior F0 discrimination and question/affirmation identification in the bimodal condition may have been simply due to the performance of the ear with the hearing aid. Due to time constraints, we did not include a hearing aid only condition and we are therefore unable to assign the bimodal benefit to either the true integration of information from both ears, or the subject’s use of the ear best equipped for this task. However, in a study with eight postlingually deaf subjects, F0 discrimination was tested with the same task as in Experiment 1. Thresholds improved from 16 (CI-only) to 12 semitones both in the CI+HA and in the HA-only condition, indicating that performance in the CI+HA condition solely depended on the ear provided with the hearing aid (van de Sandt, 2008). A “better-ear” effect was also observed for melody recognition with EAS (Kong et al., 2005).

It is possible that part of the bimodal benefit found in our study is the result of decreased performance by subjects in the CI-only condition, because this is not the natural listening condition for these subjects. However, we assume that the latter effect is important in experiments where subjects are required to recognize words or sentences, but it is less important in our study where subjects were required to complete psychoacoustic tests with little or no semantic or syntactic load. One way to shed light on this issue would be to test a representative control group of unilateral CI users who are deaf in the opposite ear and do not use a hearing aid. However, a large number of subjects would be required to achieve sufficient power given the interindividual variability observed in this study.

E. F0 perception and speech perception in quiet

Although recent studies showed that F0 may provide benefit in speech understanding, especially in noise (Dorman et al., 2008; Holt et al., 2005), we found no correlation between the speech perception task and F0 perception. F0 also contributes to speech perception in quiet, suggesting that F0 might be an important cue for several linguistic features, such as consonant voicing (Holt et al., 2001), lexical boundaries (Spitzer et al., 2009; Spitzer et al., 2007), and contextual emphasis (Fry, 1955). However, the stimuli we used were monosyllabic words. Thus, features such as lexical boundaries or contextual emphasis were not available to the subjects. If we had used multisyllabic words, a correlation between F0 perception and speech perception may have been possible due to the top-down processes of word preselection based on the F0 contour (Marslen-Wilson, 1987).

V. CONCLUSION

The present study evaluated the benefits of using a cochlear implant and a hearing aid in the non-implanted ear in children. For the implanted group, performance significantly improved when the hearing aid was added. This benefit was found in the discrimination of pitch movements in bisyllabic non-words and in the ability to distinguish between questions and affirmations. However, even with a hearing aid, the implanted group required exaggerated F0 excursions to perceive a pitch accent and to identify a question. These exaggerated excursions are close to the maximum excursions typically used by Dutch speakers. Nevertheless, the results of this study showed that compared to the implant only condition, bimodal fitting improved the perception of intonation.

\[ H_0: y^2 \leq \chi^2_{(1), 0.05} \]

\[ H_1: y^2 > \chi^2_{(1), 0.05} \]

It is possible that part of the bimodal benefit found in our study is the result of decreased performance by subjects in the CI-only condition, because this is not the natural listening condition for these subjects. However, we assume that the latter effect is important in experiments where subjects are required to recognize words or sentences, but it is less important in our study where subjects were required to complete psychoacoustic tests with little or no semantic or syntactic load. One way to shed light on this issue would be to test a representative control group of unilateral CI users who are deaf in the opposite ear and do not use a hearing aid. However, a large number of subjects would be required to achieve sufficient power given the interindividual variability observed in this study.