

# Investigating Interaural Frequency-Place Mismatches via Bimodal Vowel Integration

Trends in Hearing  
2014, Vol. 18: 1–10  
© The Author(s) 2014  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2331216514560590  
tia.sagepub.com



François Guérit<sup>1</sup>, Sébastien Santurette<sup>1</sup>, Josef Chalupper<sup>2</sup>, and Torsten Dau<sup>1</sup>

## Abstract

For patients having residual hearing in one ear and a cochlear implant (CI) in the opposite ear, interaural place-pitch mismatches might be partly responsible for the large variability in individual benefit. Behavioral pitch-matching between the two ears has been suggested as a way to individualize the fitting of the frequency-to-electrode map but is rather tedious and unreliable. Here, an alternative method using two-formant vowels was developed and tested. The interaural spectral shift was inferred by comparing vowel spaces, measured by presenting the first formant (F1) to the nonimplanted ear and the second (F2) on either side. The method was first evaluated with eight normal-hearing listeners and vocoder simulations, before being tested with 11 CI users. Average vowel distributions across subjects showed a similar pattern when presenting F2 on either side, suggesting acclimatization to the frequency map. However, individual vowel spaces with F2 presented to the implant did not allow a reliable estimation of the interaural mismatch. These results suggest that interaural frequency-place mismatches can be derived from such vowel spaces. However, the method remains limited by difficulties in bimodal fusion of the two formants.

## Keywords

cochlear implant, pitch-matching, two-formant vowels, bimodal, residual hearing

## Introduction

In recent years, an increasing number of patients with residual contralateral hearing have received a cochlear implant (CI). This population is therefore combining the neural excitation coming from the CI with that from the ear stimulated acoustically. This has been shown to improve speech perception in noise, an effect likely to come from better access to the low-frequency content of the speech in the ear with preserved acoustic hearing (Dorman & Gifford, 2010). However, the extent to which patients benefit from the combination of electric and acoustic stimulation is highly variable, with some cases of interference between the modes of stimulation (for a review, see Ching, van Wanrooy, & Dillon, 2007). Several factors have been suggested to explain this variability, such as differences in the amount of residual hearing, the devices used, and their fitting. In particular, due to the variability in electrode placement in the cochlea and in cochlear duct length among patients, it is difficult to activate nerve fibers with the same frequency-to-place map as in the contralateral ear. Typically, a standard frequency-to-electrode allocation

is used across subjects for the clinical fitting, assuming that the brain can adapt to a mismatch. The evolution of speech perception over time after implantation supports the theory of accommodation to a frequency shift (e.g., Skinner et al., 2002). However, a complete adaptation might not be possible in the case of large mismatches. Rosen, Faulkner, and Wilkinson (1999) showed that even after a long-term training period with a vocoder system simulating a 6.5-mm basalwards shift, speech recognition was worse than for the unshifted condition. Also in normal-hearing (NH) listeners, Siciliano, Faulkner, Rosen, and Mair (2010) used a six-channel vocoder and presented odd channels in the right ear, shifted 6 mm basally, while keeping the even channels unshifted in the left ear. After 10 hr of training, subjects

<sup>1</sup>Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark

<sup>2</sup>Advanced Bionics European Research Center GmbH, Hanover, Germany

## Corresponding author:

François Guérit, Hearing Systems, Department of Electrical Engineering, Building 352, Technical University of Denmark, DK-2800 Lyngby, Denmark.  
Email: fguerit@elektro.dtu.dk

showed poorer speech perception in this condition than when presented with the three unshifted channels only, suggesting that they did not benefit from combining the mismatched maps. More recently, and based on bilateral CI users' data, Kan, Stoelb, Litovsky, and Goupell (2013) suggested that the salience of interaural time and level differences was hampered for shifts greater than 3 mm. These binaural cues are essential for auditory scene analysis, and a decreased salience would imply difficulties, for example, in the segregation of speech from several noise sources.

The above findings suggest that the electrode-array location is important for adequate fitting of, and optimal benefit from, the CI. Although electrode location can theoretically be determined from computer tomography (CT) scans, these are often unavailable in audiological practice and require an additional dose of radiation. For patients having residual hearing in the opposite ear, behavioral pitch-matching has been suggested but is rather difficult because of the different percepts elicited by the implant and the acoustic stimulation. Carlyon et al. (2010) also showed that results for behavioral pitch-matching experiments are strongly influenced by nonsensory biases and that the method is tedious and time-consuming. Other behavioral methods have been suggested such as a contralateral masking paradigm (James, Blamey, Shallop, Incerti, & Nicholas, 2001; Lin, Lu, & Zeng, 2013) or an interaural time difference detection task (Francart, 2011; Francart, Brokx, & Wouters, 2009), but the results from these two methods are not very precise and are also very time consuming to obtain. More recently, the use of the binaural interaction component of the auditory brainstem response has been proposed, based on data from cats (He, Brown, & Abbas, 2010). However, the preliminary results showed no significant correlation between the amplitude of the component and interaural pitch comparisons in humans (He, Brown, & Abbas, 2012).

In the present study, based on the ability to fuse vowel formants across ears (Broadbent & Ladefoged, 1957; Cutting, 1976), an alternative method using synthesized two-formant vowels was developed and tested. This method is potentially clinic-friendly, using stimuli that are similar to those CI users deal with in their everyday lives. The question addressed was the following: Can the second formant (F2) of a two-formant vowel be used as an indicator of interaural frequency-to-place mismatch by presenting it either to the aided/normal-hearing side or to the implanted side? If the implant is perfectly fitted, the perceived vowel distributions should not depend on the ear to which F2 is presented, when fixing the first formant (F1) on the acoustic side. In the presence of an interaural mismatch, vowel distributions should show differences when presenting F2 to the acoustic versus the electric side. To test this hypothesis, an experiment with NH listeners using a vocoder system and simulated

interaural mismatches was implemented. Then, the procedure was tested with bimodal (BM) and single-sided deafness (SSD) CI users. Along with this procedure, speech-in-noise reception thresholds of the CI listeners were collected for each ear and both ears combined.

## Methods

### Subjects

Eight NH subjects were tested in Denmark, all of them native German speakers. Their hearing thresholds were below 20 dB HL at all audiometric frequencies, and the mean age was 25.4 years, ranging from 22 to 30 years. The experimental procedure was approved by the Danish Science-Ethics Committee (ref. number H-3-2013-004), and written informed consent was obtained from all subjects before data collection.

Eleven implant users were tested in the ENT department of the Unfallkrankenhaus in Berlin (UKB) and were all native German speakers. Five BM and six SSD implant users took part in the experiment. Detailed information can be found in Table 1. All BM and SSD subjects were postlingually, unilaterally deafened and had similar duration of experience with their implant (mean = 18 months,  $SD = 2.2$  months). The SSD group was, on average, younger (46 years) than the BM group (62 years) and had a shorter duration of deafness (6 years) when they received the CI, compared with the BM group (15 years). Individual pure-tone audiometry thresholds are shown in Figure 1. Aided thresholds were also measured to ensure that all stimuli were audible to the patients. The BM subjects were wearing their hearing aids (HAs) during all tests described here. The experimental procedure was approved by the ethics committee of Charité Berlin (ref. number EA4/069/12), and written informed consent was obtained from all subjects before data collection.

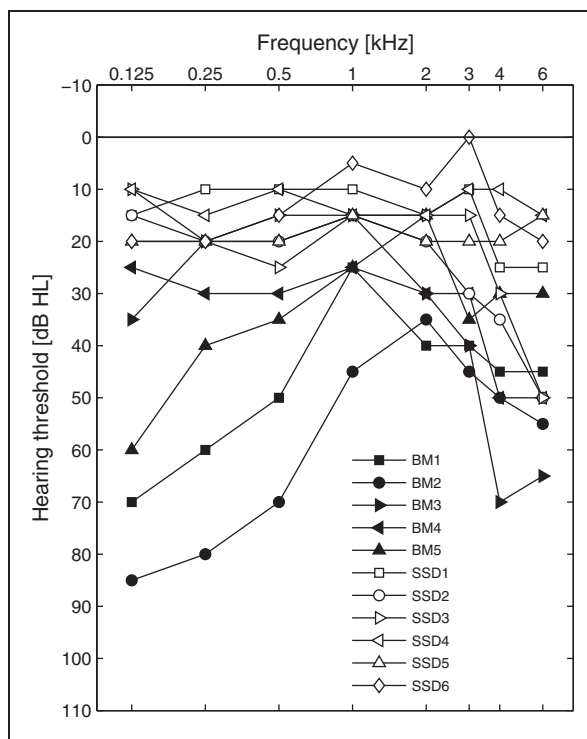
### Two-Formant Vowels Test: Stimuli and Setup

Two-formant vowels were generated using a MATLAB-based Klatt synthesizer (Klatt, 1980) and embedded between the consonants /t/ and /k/. The fundamental frequency (F0) was fixed at 110 Hz (male speaker), and the bandwidth of the first and second formant was 90 and 110 Hz, respectively. The duration of the vowels was slightly longer than normal ( $\approx 350$  ms) for ease of recognition in CI users, and the stimuli were presented at 60 dB SPL. F1 was set to 250 Hz and 400 Hz and F2 between 600 Hz and 2200 Hz in 200 Hz steps. With these settings, six different German vowels could be elicited when progressively increasing F2 with fixed F1: [u:]/[y:]/[i:] with F1 at 250 Hz and [o:]/[ø:]/[e:] with F1 at 400 Hz (cf. Table 2).

**Table 1.** Demographics of the CI Listeners.

Subject ID	Age at surgery (years)	Duration of deafness at surgery (years)	Duration of implant use (months)	Electrode; Strategy	CI side	Insertion depth angle (deg)	PTA of the nonimplanted ear (dB HL)
BM1	74	3	15	Helix; HiRes Fidelity 120	Right	350	38
BM2	73	20	21	Ij; HiRes S	Right	452	50
BM3	65	11	15	Ij; HiRes Fidelity 120	Left	355	20
BM4	60	37	19	Ij; HiRes S	Left	355	28
BM5	37	5	20	Helix; CIS	Right	332	25
SSD1	59	5	16	Helix; HiRes Fidelity 120	Right	375	12
SSD2	62	2	19	Helix; HiRes Fidelity 120	Right	353	18
SSD3	53	6	20	Helix; HiRes Fidelity 120	Right	355	18
SSD4	31	2	19	Ij; HiRes Fidelity 120	Left	375	13
SSD5	36	2	19	Helix; HiRes Fidelity 120	Right	360	18
SSD6	35	20	15	Ij; HiRes S	Right	337	10

Note. The distinction between SSD and BM was based on whether they wore a hearing aid (BM) or not (SSD). PTAs were calculated from the thresholds at 0.5, 1, and 2 kHz and were all below 20 dB HL for the SSD patients. Insertion depth angles were measured from postoperative computer tomography scans. CI = cochlear implant; SSD = single-sided deafness; BD = bimodal; PTA = pure-tone threshold average.



**Figure 1.** Individual pure-tone thresholds of the nonimplanted side for the CI listeners. Thresholds for the bimodal listeners (filled symbols) were obtained without help of the hearing aid.

A *monaural* version (F1 and F2 in the left channel) and a *dichotic* version (F1 in the left and F2 in the right channel) were created for each stimulus. For the study with NH listeners, the right channel was processed using a vocoder mimicking Advanced Bionics CI processing

**Table 2.** Possible Vowel Choices for the Normal-Hearing Subjects During the Categorization Task.

Possible choice	TUK	TÜK	TIK	TOK	TÖK	TEK
Phonetic equivalent	[u:]	[y:]	[i:]	[o:]	[ø:]	[e:]
Typical F1 (Hz)	320	301	309	415	393	393
Typical F2 (Hz)	689	1569	1986	683	1388	2010

Note. Phonetic equivalent as well as typical F1 and F2 values (Strange, Bohn, Trent, & Nishi, 2004) are indicated; 250 Hz was chosen rather than 300 Hz for F1 when synthesizing the vowels to make sure that subjects would differentiate stimuli having two different F1.

(Litvak, Spahr, Saoji, & Fridman, 2007). Fifteen-channel noise excitation was used for this vocoder, with noise bands having 25 dB/octave of attenuation. Three different settings were used: “Voc1,” “Voc2,” and “Voc3.” For the condition Voc1, the synthesis filters were identical to the analysis filters in order to simulate ideal place pitch. For the Voc2 and Voc3 conditions, the idea was to simulate a typical shift (about a fifth, or seven semitones) and a worst-case shift (more than an octave). Therefore, the synthesis filters were shifted, simulating either a slight mismatch in terms of electrode placement (Voc2, 2.5 mm shift) or a larger mismatch (Voc3, 5 mm shift at the apex). For the Voc3 condition, the mismatch was smaller toward the base for not losing the high-frequency content, as it would be in common CI processor settings.

For the NH listeners, Sennheiser HDA 200 headphones were used, ensuring good interaural attenuation (Brännström & Lantz, 2010). Test procedures were implemented in MATLAB, and all tests were conducted in a double-walled sound-attenuating listening booth.

For the implant users, the right channel was connected to the implant using the Advanced Bionics Direct Connect<sup>®</sup> system, that is, bypassing the microphone, but using the clinical speech processor. The left channel was connected to a loudspeaker, placed 1 m to the left or right side of the subjects, to stimulate their nonimplanted ear. Subjects indicated their responses orally to the experimenter, who was operating the customized MATLAB-based interface from outside the booth.

### *Two-Formant Vowels Test: Procedure for NH Listeners*

NH subjects were asked to categorize each stimulus using one of six possibilities, chosen to match with the frequency range of the stimuli (Table 2). They could listen to each stimulus up to three times if needed. No feedback was provided after choosing one of the possible choices. The different combinations of F1 and F2 resulted in two blocks of 18 stimuli each: a monaural and a dichotic block.

The first part of the test was performed using the monaural stimuli and was organized as follows: (a) two repetitions of the stimulus block were presented for training only, and (b) five repetitions were recorded ( $5 \times 18 = 90$  presentations). All stimuli were presented in a random order, and subjects were aware of the number of remaining presentations.

After this first test, the subjects were trained to fuse stimuli that were nonvocalized on one side and vocalized on the other. This was done by listening to 8 min of an audiobook, from which the right channel had been vocalized (with the Voc1 settings) and the left channel low-pass filtered at 500 Hz to mimic a typical audiogram of bimodal listeners. Subjects were asked to listen carefully to both sides, with the aim to train them to combine the nonvocalized and vocalized percepts. This training was successful, as changing the frequencies of F1 and F2 elicited different vowels for all subjects. In a pilot test without listening to the audiobook, three out of four subjects based their response on F1 only (nonvocalized), and therefore, changing the frequency of F2 had no effect on their vowel perception.

After this training, nine dichotic subtests (three for each vocoder setting, presented in a random order) were administered, following the same protocol as for the monaural test: (a) two repetitions of the dichotic stimulus block were presented for training only and (b) five repetitions of the block were recorded.

### *Two-Formant Vowels Test: Procedure for CI Users*

The same categorization task was used, but to reduce the duration of the experiment, only stimuli with F1 at 250 Hz were presented. Accordingly, only “TUK,” “TÜK,” and “TIK” were possible responses during the

task. The experiment was divided into two subtests, the first one with the monaural stimulus set and the second with the dichotic set. For each subtest, the stimulus set was repeated twice for training only, and then 10 repetitions were recorded, all stimuli being randomly presented.

### *Speech Perception of the Implant Users*

Two weeks prior to the vowel test, speech reception thresholds (SRTs) of the CI users were measured with the Oldenburg sentence test (OLSA) and the International Female Fluctuating Masker (IFFM) as interferer that was always fixed at 65 dB SPL (Holube, 2012). Subjects were seated in the booth, with the loudspeaker 1 m away from the nonimplanted ear. First, the SRT was measured for the nonimplanted ear only (CI removed, with the HA on for the BM users). Then, the SRT was obtained for the implanted ear through the Direct Connect<sup>®</sup> system. Every measurement was carried out twice using a different sentence list from the OLSA corpus.

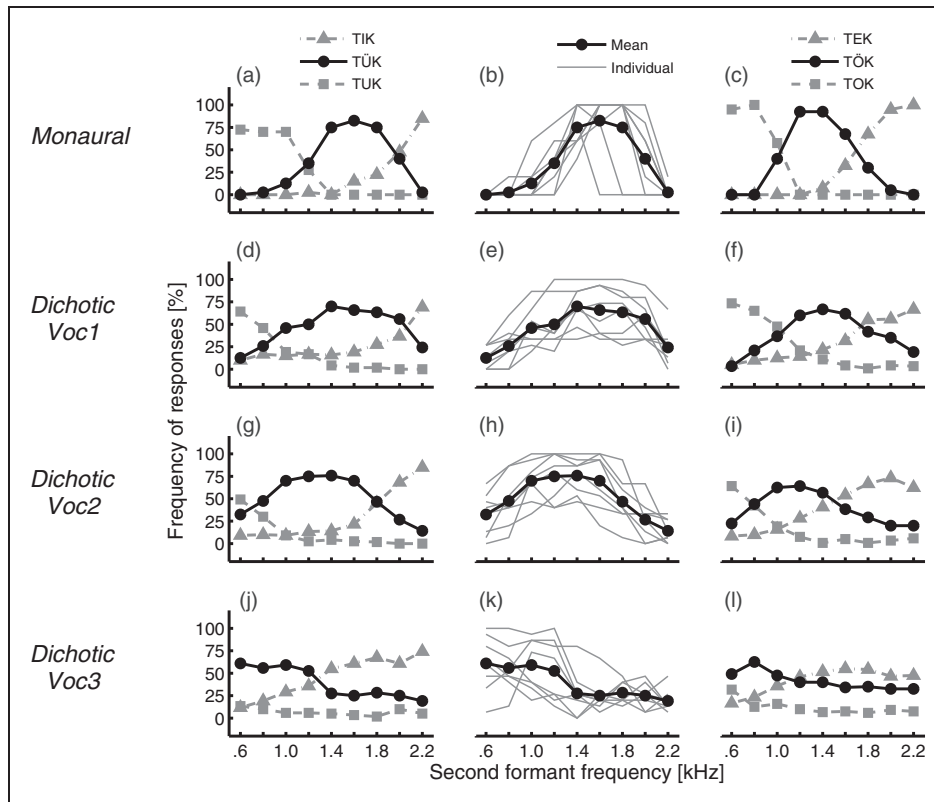
From these two conditions, the difference  $\Delta_{EI-Ac}$  between the electric and acoustic (nonimplanted ear) SRT was calculated. Then, the SRT was measured with the stimuli presented to both sides.  $\Delta_{EI-Ac}$  was added to the speech level on the electric side in order to provide cues from both ears around the SRT. Otherwise, the combined SRT would rely mainly on the better ear. A 500-ms delay was also added to the IFFM on the electric side, reducing the interaural correlation of the interferer. Therefore, cues obtained by listening in the dips would not be accessible at the same time, further limiting the effect of having a better ear.

## **Results**

### *NH Listeners*

Figure 2 shows the vowel categorization results for the eight NH listeners. In the top panel (Figure 2, first row), the results of the monaural test are plotted (F1 and F2 in the left channel). When F1 is fixed at 250 Hz (Figure 2(a)), changing F2 from 600 Hz to 2200 Hz evokes clearly different vowels: [u:] for  $F2 \approx 800$  Hz, [y:] for  $F2 \approx 1500$  Hz, and [i:] for  $F2 \approx 2000$  Hz. Individual distributions of the vowel [y:], obtained with only five repetitions, are shown in Figure 2(b). These patterns are consistent with previously reported North-German vowel maps, for example, in Strange et al. (2004). For  $F1 = 400$  Hz (Figure 2(c)), similar distributions are observed but with the three vowels [o:], [ø:], and [e:].

For the dichotic condition, when presenting F2 to the right ear, vocalized without any mismatch (Voc1), the three vowel distributions are broader (Figure 2,



**Figure 2.** Mean ( $N=8$ ) and individual results of the categorization test for the NH listeners. The number of occurrences (in %) is indicated for each vowel as a function of the frequency of F2. For the monaural condition, only five repetitions of the stimuli were presented, whereas 15 repetitions were used for the dichotic condition. Left column (a, d, g, and j): Mean results with F1 fixed at 250 Hz; therefore, only the occurrence of the choices TUK, TÜK, and TIK is shown. Middle column (b, e, h, and k): Individual results (gray lines) when F1 is fixed at 250 Hz, for the mid-F2 vowel TÜK. The mean is also shown in black. Right column (c, f, i, and l): Mean results when F1 is fixed at 400 Hz; therefore, only the occurrence of the choices TOK, TÖK, and TEK is shown. First row (a–c): Monaural condition, F1 and F2 are presented in the left channel. Second row (d–f): Dichotic condition, F1 is presented in the left channel while F2 is in the right channel, being processed with an unshifted vocoder (Voc1). Third row (g–i): Dichotic condition but with a vocoder slightly shifted (Voc2). Fourth row (j–l): Dichotic condition with a vocoder more pronouncedly shifted (Voc3). It can be seen with the mid-F2 vowel distribution (black circles) that the distribution is shifting toward the left, due to the simulated shift of the vocoder.

second row). This broadening is a direct consequence of the broadening of the individual distributions (Figure 2(e)), rather than an increased variability of the peak location. This was expected, as the noise-vocoder creates a spread of excitation. However, the distributions still reflect the three different vowels centered at similar values of F2 to without the vocoder, both for F1 at 250 and 400 Hz. For example, the mid-F2 vowel (black curve) has its distribution centered at 1400 Hz (TÜK) for both conditions. It should be noticed that two subjects exhibit a rather flat distribution of the mid-F2 vowel (Figure 2(e)). This indicates that changing F2 does not have an effect on their vowel perception, highlighting the difficulty to fuse F1 (nonvocoded) and F2 (vocoded). These subjects may have based their choice mostly on F1 perception, rather than achieving spectral fusion.

When simulating a shift with the vocoder (Voc2 and Voc3), vowel distributions were affected, as shown in the

third and fourth rows of Figure 2. The low-F2 vowels (TUK and TOK) progressively disappeared. Shifting the vocoder basally assigns channels to higher frequencies. Therefore, F2 frequencies at 600 Hz in the original signal are shifted, evoking vowels having a higher F2 frequency. The high-F2 vowels (TIK and TEK) are represented at more frequencies, and the mid-F2 vowels (TÜK and TÖK) have their distribution shifted toward the left using this representation. Looking at Figure 2(e), (h), and (k) that shows individual distributions of the mid-F2 vowel when F1 is fixed at 250 Hz (TÜK), it can be observed that changing F2 has an effect for most of the subjects. However, for the larger shift (Voc3), a higher number of subjects show a flat distribution. This can be explained both by the difficulty to achieve spectral fusion and by the fact that subjects could have been confused by perceiving only the low- and mid-F2 vowels during this test condition. These individual flat



distributions broaden the mean distribution (Figure 2(j) and (l)).

### CI Listeners

Vowel distributions measured for the implant users are shown in Figure 3. The top panels show the results of the monaural condition, in which both formants were presented acoustically, and the bottom panels present the dichotic results, for F2 presented to the implant. For the monaural condition, these are very similar to the NH listeners' distributions: The three categories (TUK, TÜK, and TIK) are similarly distributed over the F2 frequency range (Figure 3(a)). Moreover, individual distributions of the mid-F2 vowel for the same condition (Figure 3(b)) show a very good agreement across subjects, even though five of them wore a HA, and some of the SSD subjects had mild hearing losses in the nonimplanted ear.

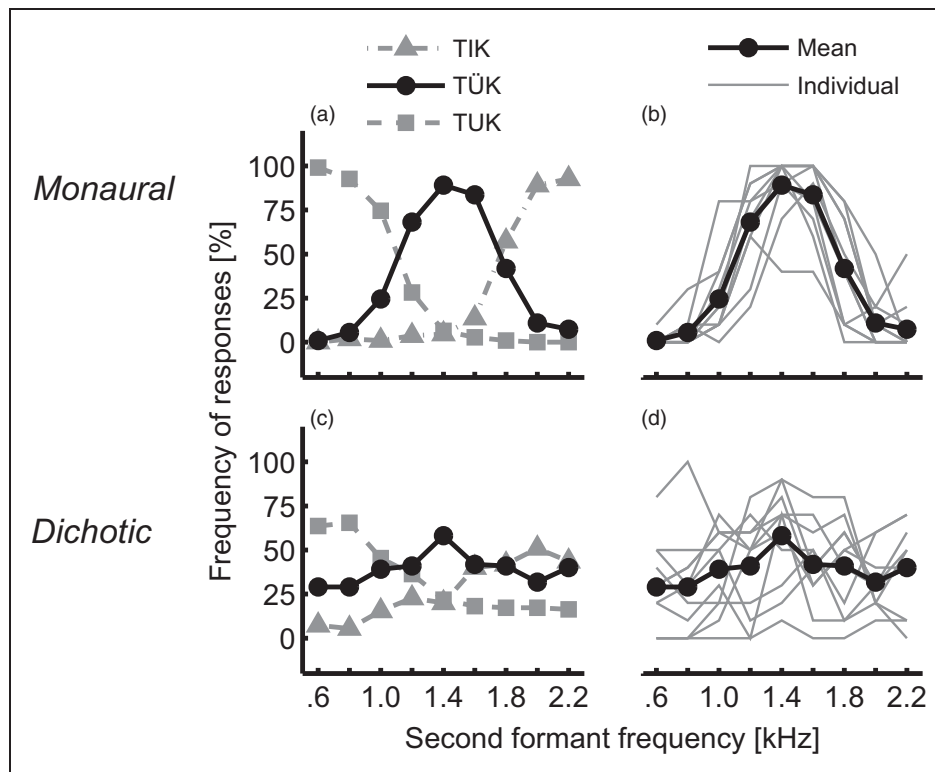
For the dichotic condition, when F2 is presented to the implant while F1 is kept on the acoustic side, the variability across subjects increases dramatically (Figure 2(d)). To highlight this variability, a subset of

five subjects' dichotic responses is shown in Figure 4. Subject SSD1 was the only one with a clear pattern for the three vowels, centered at values similar to the monaural condition. Other subjects never perceived the low-F2 (Figure 4(d)), mid-F2 (Figure 4(b)), or high-F2 vowel (Figure 4(e)), and some subjects confused vowels, for example, the low- and mid-F2 vowel for subject BM2 (Figure 4(c)).

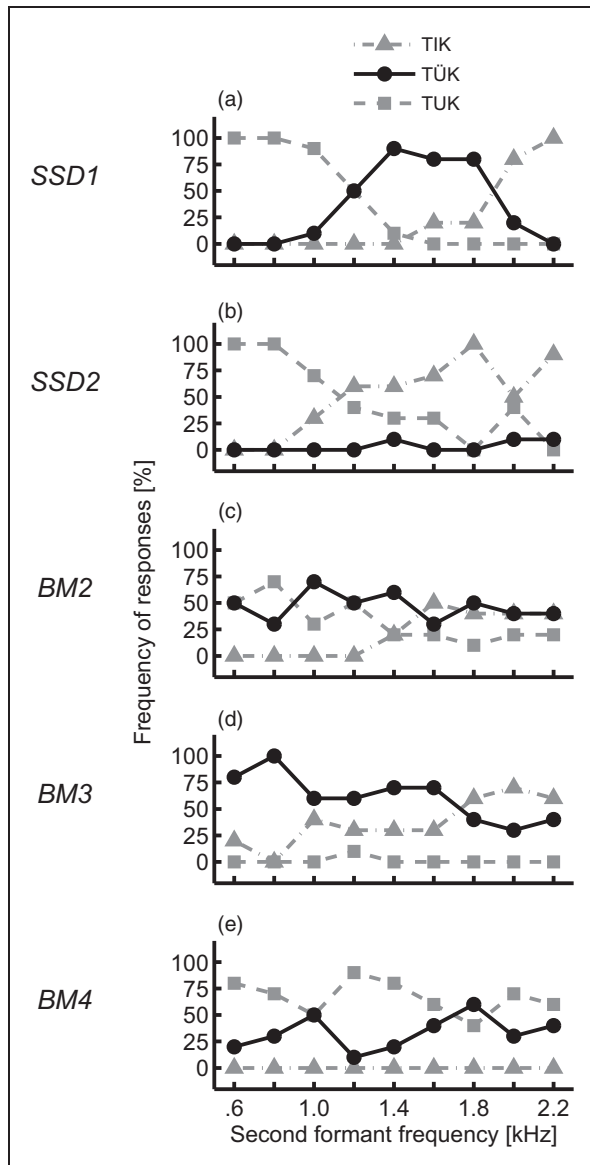
The mean distributions for the 11 CI users (Figure 3(c)) are broader than in the monaural condition. This broadening results from the individual variability, rather than from broad individual distributions, as seen in the results obtained in the NH subjects. It is also interesting to notice that the mean distributions of the dichotic condition, despite being shallower, are centered at F2 values similar to the monaural condition, especially for the low- and high-F2 vowels.

### Speech Perception Results of the CI Listeners

SRTs measured with the OLSA test in an IFFM background are shown in Figure 5. For each subject, the gray squares indicate the SRT measured in the nonimplanted

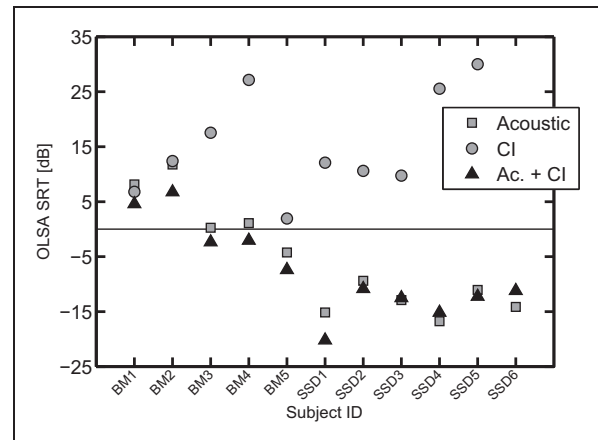


**Figure 3.** Mean ( $N = 11$ ) and individual results of the of the categorization test for the CI listeners. Left panels (a and c): Mean results of the occurrence of the three possible choices: TUK, TÜK, and TIK. Right panels (b and d): Individual (gray lines) and mean (dark circles) results for the mid-F2 vowel TÜK. Top panels (a and b): Monaural condition, F1 and F2 are presented acoustically. Bottom panels (c and d): Dichotic condition, F1 is presented acoustically and F2 electrically. In the bottom panels, the large variability when presenting F2 to the CI can be seen.



**Figure 4.** Individual results of the categorization test for a subset of CI listeners for the dichotic condition (F1 presented acoustically and F2 electrically). Very different patterns can be observed, with only one (a) resembling the results of the monaural condition, while the others (b–e) do not.

ear (*Acoustic*), the gray circles the SRT measured in the CI ear, and the dark triangles show the combined acoustic-electric thresholds. The acoustic ear was significantly better than the electric ear for the whole group ( $p < .001$ , Wilcoxon rank-sum test) but not for the BM group ( $p = .095$ , Wilcoxon). For the condition where acoustic and CI stimulation were combined (*Ac. + CI*), the speech level was adjusted and the IFFM was uncorrelated between the ears (cf. Methods section) to encourage subjects to use cues from both sides. Using this particular setup, no significant change was observed between the *Acoustic* and *Ac. + CI* conditions for the whole group



**Figure 5.** Speech reception thresholds of the CI listeners, using the Oldenburg sentence test (OLSA) in an International Female Fluctuating Masker (IFFM) background. Better speech perception is indicated by lower SRTs and is mainly driven by audibility for the acoustic presentation of the stimuli (gray squares), as the SSD subjects have lower thresholds than the BM subjects. When tested with the CI only (gray circles), the thresholds are often higher than for the acoustic condition, especially for the SSD subjects. For the combined condition (dark triangles), thresholds are similar to the ones obtained with the acoustic presentation. Subject SSD6 reached the upper limit of the dynamic range (clipping at the CI) for the CI condition.

( $p = .84$ , Wilcoxon) or for each group (BM,  $p = .42$ ; SSD,  $p = .94$ ). Despite this, the difference was always toward an improvement when adding the CI for the BM group (up to 5 dB, compare triangles with squares in Figure 5). For the SSD group, no such trend could be seen, with the effect of adding the CI being either positive or negative across subjects. This indicates that BM listeners benefited more than the SSD listeners from having the CI, with this particular setup. These results suggest that the population tested here relies mainly on the information from the acoustically stimulated ear, as previously reported with a similar cohort in Vermeire and Van de Heyning (2009).

## Discussion

The results of the NH listeners showed that fusing dichotic vowels vocoded in one ear is possible, and that the two-formant vowel method can in principle be used to derive an interaural mismatch in place of stimulation. CI listeners could perform the monaural task reliably, but their dichotic results showed a large individual variability. These aspects are discussed in more detail below.

### Design of the Experiment Using NH Listeners

The binaural fusion of different frequency bands to form an object, referred to as spectral fusion in Cutting (1976),

has been shown using vowels synthesized with the same system for each ear (Broadbent & Ladefoged, 1957; Takanen, Raitio, Santala, Alku, & Pulkki, 2013), and therefore with a similar percept on each side. Here, NH listeners were able to achieve spectral fusion when F1 was created with a pulse-excited system (Klatt synthesizer) and F2 was noise-vocoded when presented to the opposite ear. During the first pilot tests, subjects often perceived two different auditory events, typically perceiving a vowel on the nonvocoded side, and noise on the opposite, vocoded, side. Prior training with an audiobook having the left channel low-pass filtered and the right channel vocoded was sufficient to overcome this issue (at least for the Voc1 and Voc2 conditions). A noise-vocoder rather than a sine-vocoder was chosen in order to simulate the difference in perceptual quality between electric and acoustic stimulation. The efficiency of this short training (8 min) with an audiobook might indicate that it is easier to fuse the two percepts in the NH procedure than fusing the electric and acoustic percepts for CI listeners.

Setting F1 at 250 Hz, three different vowels could be perceived by changing F2 from 600 to 2200 Hz ([u:], [y:], and [i:]). A similar pattern was observed having F1 set at 400 Hz, with the vowels [o:], [ø:], and [e:]. Due to this three-vowel distribution obtained by varying F2 only, an indirect measure of frequency perception can be derived when comparing the monaural and dichotic conditions, with the monaural condition acting as a reference. When simulating a frequency mismatch by shifting the synthesis filters of the vocoder, an effect could be seen in the vowel distributions, both individually and in the group average (Figure 2): The F2 center frequencies were shifted for each vowel, reflecting the interaural mismatch. However, this effect may be expected to be more salient in such NH listeners, who did not have time to adapt to the mismatch, than in CI listeners, who might have acclimatized to a potential interaural mismatch.

### *Monaural Results in CI Listeners*

Implant users were able to perform the task reliably in the monaural condition, with low inter- and intrasubject variability (Figure 3 (a) and (b)). This part of the test was achieved in 15 min (30 min with the dichotic condition) and was easy to explain to the subjects. In comparison to classical pitch-matching experiments, where training is necessary for the subject to perform the task, this more ecological approach thus seems promising, as the population wearing implants is far from the cohort of young students usually tested in such psychoacoustic experiments.

The bimodal population tested here was atypical, as most BM subjects had a severe low-frequency hearing

loss, which can lead to a distorted perception of pitch. Interestingly, the bimodal subjects (wearing their HA during the experiment) performed well in this monaural vowel discrimination task (see the low across-subjects variability in Figure 3(b)), even though their OLSA SRT was significantly higher than that of the SSD group ( $p < .01$ , Wilcoxon rank-sum test). This does allow conclusions about their pitch perception ability, but it indicates that formant discrimination is preserved when assessed with synthesized two-formant vowels.

### *Variability in the Dichotic Results of CI Listeners*

Individual results of the implant users for the dichotic condition (Figure 3(d)) showed large differences across subjects. Only one subject had a similar distribution for both conditions (SSD1, Figure 4(a)), whereas the others showed very different patterns (a few examples can be seen in Figure 4). All subjects had a similar experience with their implant (18 months) and a similar insertion depth (mean = 363°,  $SD = 31^\circ$ ). In the study from Harnsberger et al. (2001), no systematic shift was reported in the monaural, individual, vowel maps they recorded, and the variability was attributed to individual differences in formant frequency discrimination. Here, only one formant was presented to the CI, therefore reducing the variability stemming from individual differences in formant frequency discrimination. The large difference in the individual results could also be caused by difficulties to fuse F1 and F2 when they have different perceptual qualities, as reported in classical pitch-matching experiments (Carlyon et al., 2010). Abnormal binaural spectral integration has also been shown in bimodal subjects, which could account for the difficulties in integrating information from both ears (Reiss, Ito, Eggleston, & Wozny, 2014). Here, some subjects showed a flat distribution (Figure 4(c)), suggesting that they based their response on F1 only, presented acoustically. Moreover, subjects having speech thresholds similar to subject SSD1 on the implanted side (suggesting an equivalent vowel perception) showed very different results in the dichotic test condition. Taken together, these findings suggest an insufficient fusion between the electric and acoustic percepts as the main reason for the variability seen in the results. Whether matching the place of excitation in the presence of these very different perceptual qualities would help in terms of speech perception remains unknown, as other factors (e.g., binaural spectral integration) might also be contributing (Francart & McDermott, 2013; Reiss et al., 2014).

Shallow distributions can be seen in the mean results of the CI listeners (Figure 3(c)), given the large individual variability. The distribution of the mid-F2 vowel (TÜK) is flat, but the low-F2 and high-F2 vowels have mean distributions with patterns centered at values similar to



the monaural condition. This suggests that, after 18 months of implantation, implanted listeners may be acclimatized to the new tonotopic organization given by the implant, consistent with previous studies (McDermott, Sucher, & Simpson, 2009; Reiss, Turner, Erenberg, & Gantz, 2007).

### Further Investigations

The two-formant vowel method described in this study was intended to estimate interaural frequency-place mismatches. However, the results suggest that abnormal bimodal vowel integration is prominent in the BM and SSD subjects tested here, which limits the possibility of estimating this interaural mismatch. Further investigations on such vowel integration would be relevant, especially regarding the growing implanted population with residual hearing in the contralateral ear. This could include a study of binaural spectral integration considering dichotic pitch fusion (Reiss et al., 2014) and formant fusion.

Furthermore, regarding the use of this method for mismatch evaluation, the effects of training, adaptive procedures, a comparison with a classical pitch-matching experiment, and testing with both formants presented to the CI should be considered to better understand the sources of the individual variability. Training with an audiobook appeared to be very efficient with the NH subjects but might be underestimating the difficulties of bimodal spectral integration. Testing with both formants stimulated electrically would provide information on individual formant frequency discrimination, which could be a potential cause of individual variability.

The CI population tested here had the acoustic ear as the better ear, which is not typical of a BM population. This raises some challenges, for example, in terms of assessing the speech-in-noise benefit of the combined electric-acoustic stimulation, where not only the SRTs but also the percepts differ between the two ears. Methods to test both ears simultaneously at their own SRT on these listeners have not yet been described.

If the two-formant vowel procedure was tested in a more typical BM population, with only very low frequencies preserved, F2 could be too high in frequency to be perceivable. Instead of measuring the monaural condition, a reference acoustic vowel map could be used for the estimation of the mismatch. It is also known that pitch perception changes over time after implantation (Reiss et al., 2007) or after a change in the frequency-to-electrode map (Svirsky, Silveira, Neuburger, Teoh, & Suarez, 2004). These are points that were not tested here but should be replicated with the two-formant vowel method for its validation. Finally, only bimodal German-speaking listeners participated in this experiment, but this protocol could also be

applied to other languages with a few modifications in the stimuli, as well as listeners with residual hearing in the implanted ear or bilateral implant users.

### Conclusions

NH listeners' results and mean results of the CI listeners suggest that place mismatches can be derived from vowel spaces obtained when presenting two-formant vowels monaurally and dichotically in SSD and BM listeners. This test is also easier and less time consuming to perform for subjects than a classical pitch-matching paradigm. However, the method's reliability remains very limited by the individual variability, and results mostly indicate an abnormal bimodal vowel integration within the CI population tested here.

### Authors' Note

Part of this work was presented at the Fourth International Symposium on Auditory and Audiological Research held in Nyborg, Denmark, August 2013, and at the 37th Annual Midwinter Meeting of the Association for Research in Otolaryngology held in San Diego, California, February 2014.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Technical University of Denmark.

### Acknowledgments

The authors would like to thank Franziska Barthold for conducting the tests with the CI patients in Berlin, as well as Ingo Todt for analyzing CT scans to estimate insertion depth.

The authors also thank the editor, Andrew Oxenham, and two anonymous reviewers for helpful comments on an earlier version of this paper.

### References

- Brännström, K. J., & Lantz, J. (2010). Interaural attenuation for Sennheiser HDA 200 circumaural earphones. *International Journal of Audiology, 49*(6), 467–471. doi:10.3109/14992021003663111.
- Broadbent, B. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America, 29*(6), 708–710.
- Carlyon, R. P., Macherey, O., Frijns, J. H. M., Axon, P. R., Kalkman, R. K., Boyle, P., . . . , & Dauman, R. (2010). Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear. *Journal of the Association for*

- Research in Otolaryngology: JARO*, 11(4), 625–640. doi:10.1007/s10162-010-0222-7.
- Ching, T. Y. C., van Wanrooy, E., & Dillon, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends in Amplification*, 11(3), 161–192. doi:10.1177/1084713807304357.
- Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, 83(2), 114–140.
- Dorman, M. F., & Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *International Journal of Audiology*, 49(12), 912–919. doi:10.3109/14992027.2010.509113.
- Francart, T. (2011). Sensitivity of bimodal listeners to interaural time differences with modulated single- and multiple-channel stimuli. *Audiology & Neuro-Otology*, 16(2), 82–92.
- Francart, T., Brokx, J., & Wouters, J. (2009). Sensitivity to interaural time differences with combined cochlear implant and acoustic stimulation. *Journal of the Association for Research in Otolaryngology: JARO*, 10(1), 131–141. doi:10.1007/s10162-008-0145-8.
- Francart, T., & McDermott, H. J. (2013). Psychophysics, fitting, and signal processing for combined hearing aid and cochlear implant stimulation. *Ear and Hearing*, 34(6), 685–700. doi:10.1097/AUD.0b013e31829d14cb.
- Harnsberger, J. D., Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., & Meyer, T. A. (2001). Perceptual “vowel spaces” of cochlear implant users: Implications for the study of auditory adaptation to spectral shift. *The Journal of the Acoustical Society of America*, 109(5), 2135. doi:10.1121/1.1350403.
- He, S., Brown, C. J., & Abbas, P. J. (2010). Effects of stimulation level and electrode pairing on the binaural interaction component of the electrically evoked auditory brain stem response. *Ear and Hearing*, 31(4), 457–470. doi:10.1097/AUD.0b013e3181d5d9bf.
- He, S., Brown, C. J., & Abbas, P. J. (2012). Binaural interaction component of the electrically evoked auditory brainstem response and interaural pitch comparisons in bilateral cochlear implant recipients. *Ear and Hearing*, 33(1), 57–68. doi:10.1097/AUD.0b013e31822519ef.Preliminary.
- Holube, I. (2012). Speech intelligibility in fluctuating maskers. In T. Dau, M. L. Jepsen, T. Poulsen, & J. C. Dalsgaard (Eds.), *Proceedings of ISAAR 2011: Speech Perception and Auditory Disorders. 3rd International Symposium on Auditory and Audiological Research*. (pp. 57–64). August 2011, Nyborg, Denmark: The Danavox Jubilee Foundation.
- James, C., Blamey, P., Shallop, J. K., Incerti, P. V., & Nicholas, A. M. (2001). Contralateral masking in cochlear implant users with residual hearing in the non-implanted ear. *Audiology & Neuro-Otology*, 6(2), 87–97. doi:10.1159/000046814.
- Kan, A., Stoelb, C., Litovsky, R. Y., & Goupell, M. J. (2013). Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. *The Journal of the Acoustical Society of America*, 134(4), 2923–2936. doi:10.1121/1.4820889.
- Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67(3), 971–995.
- Lin, P., Lu, T., & Zeng, F.-G. (2013). Central masking with bilateral cochlear implants. *The Journal of the Acoustical Society of America*, 133(2), 962–969. doi:10.1121/1.4773262.
- Litvak, L. M., Spahr, A. J., Saoji, A. A., & Fridman, G. Y. (2007). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *The Journal of the Acoustical Society of America*, 122(2), 982–991. doi:10.1121/1.2749413.
- McDermott, H., Sucher, C., & Simpson, A. (2009). Electro-acoustic stimulation. Acoustic and electric pitch comparisons. *Audiology & Neuro-Otology*, 14(Suppl. 1): 2–7. doi:10.1159/000206489.
- Reiss, L., Ito, R., Eggleston, J. L., & Wozny, D. R. (2014). Abnormal binaural spectral integration in cochlear implant users. *Journal of the Association for Research in Otolaryngology: JARO*, 15(2), 235–248. doi:10.1007/s10162-013-0434-8.
- Reiss, L., Turner, C. W., Erenberg, S. R., & Gantz, B. J. (2007). Changes in pitch with a cochlear implant over time. *Journal of the Association for Research in Otolaryngology: JARO*, 8(2), 241–257. doi:10.1007/s10162-007-0077-8.
- Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. *The Journal of the Acoustical Society of America*, 106(6), 3629–3636.
- Siciliano, C. M., Faulkner, A., Rosen, S., & Mair, K. (2010). Resistance to learning binaurally mismatched frequency-to-place maps: Implications for bilateral stimulation with cochlear implants. *The Journal of the Acoustical Society of America*, 127(3), 1645–1660. doi:10.1121/1.3293002.
- Skinner, M. W., Ketten, D. R., Holden, L. K., Harding, G. W., Smith, P. G., Gates, G. A., . . . , & Blocker, B. (2002). CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients. *Journal of the Association for Research in Otolaryngology: JARO*, 3(3), 332–350. doi:10.1007/s101620020013.
- Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, 115(4), 1791–1807. doi:10.1121/1.1687832.
- Svirsky, M. A., Silveira, A., Neuburger, H., Teoh, S. W., & Suarez, H. (2004). Long-term auditory adaptation to a modified peripheral frequency map. *Acta Otolaryngologica*, 124(4), 381–386. doi:10.1080/00016480310000593.
- Takanen, M., Raitio, T., Santala, O., Alku, P., & Pulkki, V. (2013). Fusion of spatially separated vowel formant cues. *The Journal of the Acoustical Society of America*, 134(6), 4508. doi:10.1121/1.4826181.
- Vermeire, K., & Van de Heyning, P. (2009). Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiology & Neuro-Otology*, 14(3), 163–171. doi:10.1159/000171478.