Interaural bimodal pitch matching with two-formant vowels

François Guérit[1,*], Josef Chalupper[2], Sébastien Santurette[1], Iris Arweiler[2] and Torsten Dau[1]

[1] *Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2] *Advanced Bionics European Research Center GmbH, D-30625 Hanover, Germany*

For bimodal patients, with a hearing aid (HA) in one ear and a cochlear implant (CI) in the opposite ear, usually a default frequency-to-electrode map is used in the CI. This assumes that the human brain can adapt to interaural place-pitch mismatches. This 'one-size-fits-all' method might be partly responsible for the large variability of individual bimodal benefit. Therefore, knowledge about the location of the electrode array is an important prerequisite for optimum fitting. Theoretically, the electrode location can be determined from CT scans. However, these are often not available in audiological practice. Behavioral pitch matching between the two ears has also been suggested, but has been shown to be tedious and unreliable. Here, an alternative method using two-formant vowels was developed and tested with a vocoder system simulating different CI insertion depths. The hypothesis was that patients may more easily identify vowels than perform a classical pitch-matching task. A spectral shift is inferred by comparing vowel spaces, measured by presenting the first formant in the HA and the second either in the HA or the CI. Results suggest that pitch mismatches can be derived from such vowel spaces. In order to take auditory adaptation in individual patients into account, the method is tested with CI patients with contralateral residual hearing.

INTRODUCTION

In the last years, an increased number of patients having residual contralateral hearing received a cochlear implant (CI). This population is therefore combining the neural excitation coming from the CI and that from the ear stimulated acoustically. However, 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 *et al.* (1999) showed that even after a long-term training period with a vocoder

*Corresponding author: fguerit@elektro.dtu.dk

Proceedings of ISAAR 2013: Auditory Plasticity – Listening with the Brain. 4th symposium on Auditory and Audiological Research. August 2013, Nyborg, Denmark. Edited by T. Dau, S. Santurette, J. C. Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. ISBN: 978-87-990013-4-7. The Danavox Jubilee Foundation, 2014.

system simulating a 6.5-mm basalwards shift, speech recognition was worse than for the unshifted condition. More recently, Siciliano *et al.* (2010) used a 6-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 hours of training, subjects 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.

The above findings suggest that the electrode-array location is crucial for adequate fitting and optimal benefit from the CI. Although electrode location can theoretically be determined from computed-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 of behavioral pitch-matching are strongly influenced by nonsensory biases and that the method is tedious and time-consuming. Here, based on the ability to fuse vowel formants across ears (Carlson *et al.*, 1975), an alternative method using two-formant vowels was developed and tested. This method is thought to be clinic-friendly, using stimuli that the CI users are dealing with in their everyday life.

The question addressed in the present study is the following: Can the second formant (F2) of a two-formant vowel be used as a pitch-matching stimulus by presenting it either on the aided/normal-hearing side or on 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 vs the electric side. To test this hypothesis, an experiment with normal-hearing (NH) listeners using a vocoder system and simulated interaural mismatches was implemented. In order to take auditory adaptation into account, as well as the difficulties regarding the fusion between electric and acoustic percepts, the method was also tested with bimodal (BM) and single-sided-deaf (SSD) CI users.

METHODS

Subjects

Eight NH listeners were tested, 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.

Eleven implant users were tested in the ENT department of the Unfallkrankenhaus (UKB) in Berlin, and were all native German speakers. Five BM and six SSD implant users took part in the experiment. The mean age was 55.6 years, ranging from 33 to 78 years. The subjects were post-lingually deafened and had a a similar experience with their implant (mean: 19.9 months, SD: 2.1 months). All were equipped with

Advanced Bionics electrode arrays and processors.

Stimuli and equipment

Two-formant vowels were generated using a Matlab-based Klatt synthesizer (Klatt, 1980), and embedded into the consonants /t/ and /k/. The duration of the vowels was slightly longer than normal (≈350 ms) for ease of recognition in CI users. The stimuli were presented at 60 dB SPL. F1 was set at 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.

A monaural (F1 and F2 in the left channel) and a dichotic (F1 in the left and F2 in the right channel) version were created for each stimulus. For the study with NH listeners, the right channel was vocoded using a vocoder mimicking Advanced Bionics CI processing (Litvak *et al.*, 2007). 16-channel noise excitation was used for this vocoder, with noise bands having 25 dB/octave of attenuation. Three different settings were used: 'Voc1' (perfect fitting), 'Voc2' (slight basal shift, ≈ 0.45 octave), and 'Voc3' (larger mismatch, ≈ 0.85 octave). For the NH listeners, Sennheiser HDA 200 headphones were used, ensuring a good interaural attenuation (Brännström and Lantz, 2010). Test procedures were implemented in Matlab and all testings were conducted in a double-walled sound-attenuating listening booth.

For the implant users, the right channel was connected to the implant processor, using the Advanced Bionics Direct Connect® system. Subjects were seated in a booth, and the left channel was connected to a loudspeaker, placed 1 meter to the left or right side of the subjects, to stimulate their non-implanted ear. Subjects indicated their responses orally to the audiologist in charge of the experiment, who was using the custom Matlab-based interface outside the booth.

Procedure for NH listeners

NH subjects were forced to categorize each stimulus using one of six possibilities, chosen to match with the frequency range of the stimuli (Table 1). They could listen to each stimulus up to three times if needed. 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 organized as follows: (1) two repetitions of the stimulus block were presented for training only, (2) 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 non-vocoded on one side and vocoded on the other. This was done by listening to 8 minutes of an audio-book, from which the right channel had been vocoded and the left channel lowpass-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

non-vocoded and vocoded percepts. After this training, nine dichotic sub-tests (three for each vocoder setting, presented in a random order) were administered, following the same protocol as for the monaural test: (1) two repetitions of the dichotic stimulus block were presented for training only, (2) five repetitions of the block were recorded.

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

Table 1: Possible vowel choices for the NH subjects during the categorization task. Phonetic equivalent as well as typical F1 and F2 values (Strange *et al.*, 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.

Procedure for implant 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 sub-tests, the first one with the monaural stimulus set, and the second one with the dichotic set. For each sub-test, the stimulus set was repeated twice for training only, and then 10 repetitions were recorded, all stimuli being randomly presented.

RESULTS

NH listeners

Figure 1 shows the vowel categorization results for the 8 NH listeners. In the left panel (A/E), results of the 'monaural' test are plotted. For F1 = 250 Hz as well as for F1 = 400 Hz, changing F2 from 600 Hz to 2200 Hz evokes clearly different vowels: [u:]/[o:] for F2≈800 Hz; [y:]/[ø:] for F2≈1500 Hz; [i:]/[e:] for F2≈2000 Hz. These patterns are consistent with previously reported North-German vowel maps (e.g., Strange *et al.*, 2004).

When presenting F2 to the right ear vocoded without any mismatch ('Voc1'), the three vowel distributions are broader (panels B and F in Fig. 1). 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. For example, the mid-F2 vowel (*black curve*) has its distribution centered around 1400 Hz ('TÜK') and 1600 Hz ('TÖK') for both conditions.

When simulating a shift with the vocoder ('Voc2' and 'Voc3'), vowel distributions are affected, as seen in panels C, D, G, and H in Fig. 1. The low-F2 vowels (TUK and TOK) progressively disappear. Shifting the vocoder basally assigns channels to

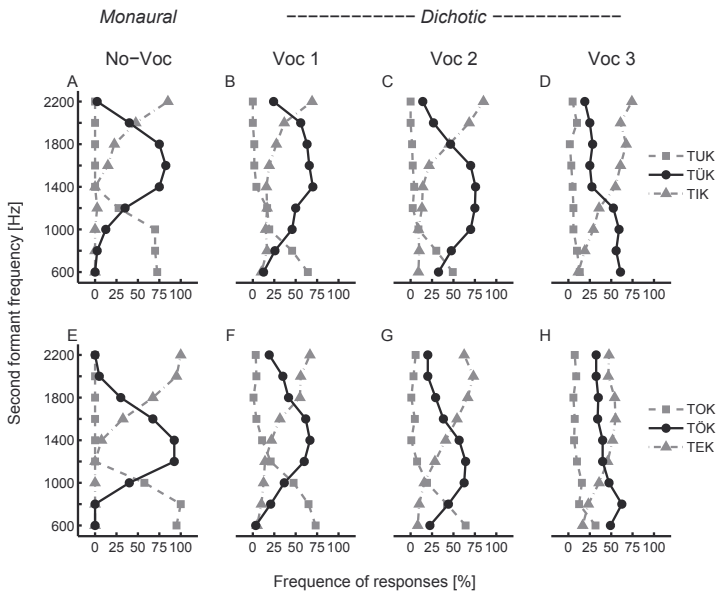

Fig. 1: Mean results (N = 8) of the categorization test for the NH listeners. The number of occurrences (in %) for each vowel is indicated as a function of the frequency of F2. *Top panel:* F1 is fixed at 250 Hz, therefore only the occurrence of the choices TUK, TÜK, and TIK is shown. *Bottom panel:* F1 is fixed at 400 Hz, only the occurrence of the choices TOK, TÖK, and TEK is shown. *Left panel (A/E):* F1 and F2 are presented in the left channel. *Mid-left panel (B/F):* F1 is in the left channel while F2 is in the right channel, processed with an unshifted vocoder. *Mid-right panel (C/G):* F1 is in the left channel while F2 is in the right channel, processed with a slightly shifted vocoder. *Right panel (D/H):* F1 is in the left channel while F2 is in the right channel, processed with a more pronouncedly shifted vocoder.

higher place-frequencies. Therefore, F2 frequencies at 600 Hz in the original signal are shifted, evoking vowels having a higher F2 frequency. In a similar way, the high-F2 vowels (TIK and TEK) are more and more represented, and the mid-F2 vowels (TÜK and TÖK) have their distribution shifted downwards in frequency using this representation.

To assess the simulated shift quantitatively, the F2 distribution of the mid-F2 vowels (categories TÜK and TÖK) are fitted by means of a Gaussian distribution. Fitted center frequencies (mean of the Gaussian distribution) are shown in Fig. 2. The expected center frequencies (dashed gray lines in Fig. 2) are calculated using the

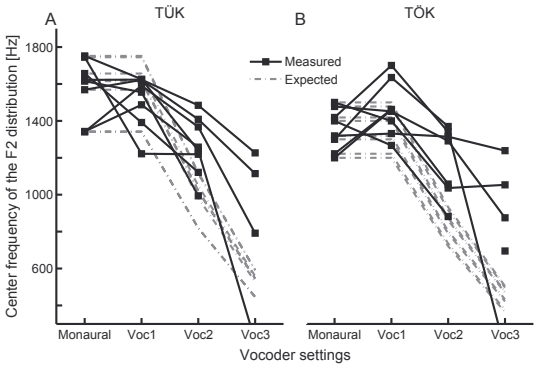

Fig. 2: Fitted center frequencies for individual NH listeners' (N = 8) mid-F2 vowel distributions. (*A*) Fitted center frequencies of the category 'TÜK' (F1 = 250 Hz). (*B*) Fitted center frequencies of the category 'TÖK' (F1 = 400 Hz). For panels (*A*) and *B*), a Gaussian fit was applied for the F2 distribution, and the center is plotted (black squares) for the different conditions. Expected centers for each individual were calculated from the results of the 'monaural' condition and the vocoder settings, and are indicated in dashed gray lines. Center frequencies reaching the frequency limits (100-4000 Hz) of the fitting procedure were removed.

vocoder settings and the fitted center frequency of the 'monaural' condition of each subject. Even though there is a trend of these fitted center frequencies to follow the expected shift from the vocoder, variability is high across subjects, especially for the largest mismatch ('Voc3'). Moreover, for the larger mismatch, some subjects showed a rather flat distribution, indicating a difficulty to fuse the two percepts: No effect of changing F2 indicates that they based their response on F1 only.

CI listeners

Vowel distributions for the monaural condition for both the SSD and BM implant users were very similar to the NH listeners' distributions: the three categories (TUK, TÜK, and TIK) were similarly distributed over the F2 frequency range. An example of one subject's monaural distribution is shown in Fig. 3 (panel A). Assuming that the brain would adapt to mismatches, similar vowel maps would be expected when presenting the second formant either in the implanted or non-implanted ear, as shown for NH listeners in Fig. 1. This was only observed for one of the eleven subjects (panel B in Fig. 3). For the other subjects, various patterns could be observed, and three of them are shown in panels C to E. Some subjects showed a pattern resembling a basal shift (C), others showed a rather flat distribution (D), and one subject even never perceived the mid-F2 vowel (E). This variability was seen for both groups (SSD and BM) and does not imply that these subjects have a mismatch, as discussed later.

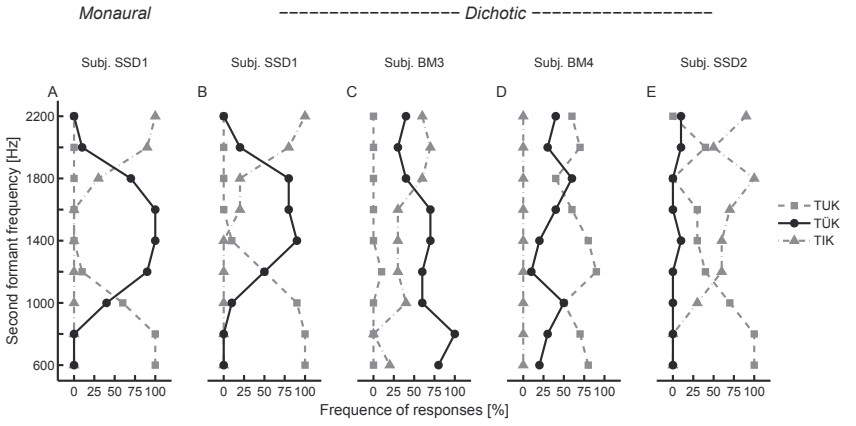

Fig. 3: Five examples of individual CI listeners categorization results. (*A*) 'Monaural' condition results of one subject. The mid-F2 vowel is highlighted in black. (*B*) 'Dichotic' results where the subject has a similar distribution to the 'monaural' condition. (*C*) 'Dichotic' results resembling a basal shift of the electrode array. (*D*) 'Dichotic' results where the subject showed uniform categorization. (*E*) 'Dichotic' results where the subject almost never perceived the mid-F2 vowel.

DISCUSSION AND CONCLUSIONS

NH listeners were able to fuse formants of two-formant vowels when presenting them dichotically with F2 vocoded. Fusion was challenging with the two different percepts, but this was overcome by a careful training and description of the test. The effect of simulating a shift could be seen in the vowel distributions. The low-F2 vowels ([u:] and [o:]) were less represented as the shift was increased. Estimates of the shifts from the mid-F2 vowels ([y:] and [ø:]) were overall smaller than their theoretical value, with high across-subjects variability, and might not represent the best way to estimate a shift. Overall, the NH listeners' results suggest that this new procedure could be a tool to indicate the existence of a mismatch, but that it remains challenging to evaluate this mismatch quantitatively.

Vowel distributions could be derived for all CI users in the monaural acoustic condition, indicating an ability to perform the task reliably. Despite this, large individual differences were observed for dichotic bimodal stimulation, with listeners showing either basal or apical shifts, or generally-poor vowel discrimination. This could be due to the difficulty to fuse percepts more than to possible mismatches. Indeed, for some NH subjects having difficulty to fuse non-vocoded and vocoded percepts, similar distributions could be seen, where the subjects would focus mainly on F1. This was overcome for NH subjects by training them to fuse percepts before

categorizing two-formant vowels. Adequate training should be investigated for CI patients in order to obtain vowel distributions based on the fusion of both formants.

CT-scan insertion depth evaluation should be compared to the vowel distributions of the CI patients to look for a possible correlation and shed light on the large variability observed. Moreover, speech perception results using either the CI stimulation only, the non-implanted side only, or both, will be collected for the tested patients. It might be interesting to look at a potential effect of having a dominant ear or a good combination of information across ears. As a general conclusion, the two-formant task is reliable and straight-forward in NH listeners and has potential to detect a mismatch in bimodal CI patients. However, it is difficult to obtain a quantitative estimate of the mismatch with this method and fusion issues should be overcome.

REFERENCES

Brännström, K.J., and Lantz, J. (**2010**). "Interaural attenuation for Sennheiser HDA 200 circumaural earphones", Int. J. Audiol., **49**, 467-471.

Carlson, R., Fant, G., and Granström, B. (**1975**). "Two-formant models, pitch and vowel perception", in *Auditory analysis and perception of speech*. Edited by G. Fant, pp 55-82.

Carlyon, R.P., Macherey, O., Frijns, J.H.M., Axon, P.R., Kalkman, R.K., Boyle, P., Baguley, D.M., Briggs, J., Deeks, J.M., Briaire, J.J., Barreau, X., and Dauman, R. (**2010**). "Pitch comparisons between electrical stimulation of a cochlear implant and acoustic stimuli presented to a normal-hearing contralateral ear", J. Assoc. Res. Oto., **11**, 625-640.

Klatt, D.H. (**1980**). "Software for a cascade/parallel formant synthesizer", J. Acoust. Soc. Am., **67**, 971-995.

Litvak, L.M., Spahr, A.J., Saoji, A.A., and Fridman, G.Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners", J. Acoust. Soc. Am., **122**, 982-991.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants", J. Acoust. Soc. Am., **106**, 3629-3636.

Siciliano, C.M., Faulkner, A., Rosen, S., and Mair, K. (**2010**). "Resistance to learning binaurally mismatched frequency to place maps: implications for bilateral stimulation with cochlear implants", J. Acoust. Soc. Am., **127**, 1645-1660.

Skinner, M.W., Ketten, D.R., Holden, L.K., Harding, G.W., Smith, P.G., Gates, G.A., Neely, J.G., Kletzker, G.R., Brunsden, B., and Blocker, B. (**2002**). "CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients", J. Assoc. Res. Oto., **3**, 332-350.

Strange, W., Bohn, O.-S., Trent, S.A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels", J. Acoust. Soc. Am., **115**, 1791-1807.