



HHS Public Access

Author manuscript

Ear Hear. Author manuscript; available in PMC 2015 September 01.

Published in final edited form as:

Ear Hear. 2014 ; 35(5): 580–584. doi:10.1097/AUD.0000000000000044.

Binaural Enhancement for Bilateral Cochlear Implant Users

Christopher A. Brown

Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Abstract

Bilateral cochlear implant (BCI) users receive limited binaural cues and, thus, show little improvement to speech intelligibility from spatial cues. The feasibility of a method for enhancing the binaural cues available to BCI users is investigated. This involved extending interaural differences of levels, which typically are restricted to high frequencies, into the low-frequency region. Speech intelligibility was measured in BCI users listening over headphones and with direct stimulation, with a target talker presented to one side of the head in the presence of a masker talker on the other side. Spatial separation was achieved by applying either naturally occurring binaural cues or enhanced cues. In this listening configuration, BCI patients showed greater speech intelligibility with the enhanced binaural cues than with naturally occurring binaural cues. In some situations, it is possible for BCI users to achieve greater speech intelligibility when binaural cues are enhanced by applying interaural differences of levels in the low-frequency region.

Keywords

Bilateral cochlear implants; Speech intelligibility; Binaural cues; Interaural time differences; Interaural level differences

INTRODUCTION

Implanting both cochleas of hearing-impaired listeners has become more common in recent years. However, implantation is invasive, costly, and can potentially destroy any residual hearing in the ear to be implanted. A loss of residual hearing can be detrimental to speech reception, even if the amount of hearing is extremely limited (Brown & Bacon 2010; Zhang et al. 2010). Thus, it is crucial that clear benefits of bilateral cochlear implants (BCIs) be established over a single device, with or without the addition of residual hearing, to justify the decision to implant the second ear. One often-cited benefit of BCI is the ability of such users to perceive and use binaural cues, and the potential outcome is most often stated or implied to be improved speech reception through their use, such as the spatial release from masking observed in listeners with normal hearing. However, BCI users have thus far shown relatively poor localization abilities (Grantham et al. 2008) and limited spatial release from masking (Loizou et al. 2009).

Copyright © 2014 by Lippincott Williams & Wilkins

Address for correspondence: Christopher A. Brown, Department of Communication Science and Disorders, University of Pittsburgh, 4033 Forbes Tower, Pittsburgh, PA 15260, USA. cbrown1@pitt.edu.

This is likely because BCI users receive limited access to binaural cues. They do perceive interaural differences of levels (ILDs), but they have shown poorer sensitivity to interaural differences of time (ITDs; Grantham et al. 2008), even with envelope ITDs, which are generally reasonably well preserved in BCIs (e.g., Laback et al. 2004). Robust ILDs are generally restricted to frequencies above about 1500 to 2000 Hz (Fig. 1) because the longer wavelengths at lower frequencies are not shadowed by the head. Thus, the availability of binaural information to BCI users is strongly frequency dependent. It has been shown that sensitivity to binaural cues declines when they are inconsistent across frequency (Francart & Wouters 2007; Brown & Yost 2011). In addition, any ILDs that BCI users receive will be subjected to large amounts of compression. This includes automatic gain control on the processing front end, which essentially limits the level of more intense sounds, likely reducing ILDs as a result. There is also the compression that occurs to map the input dynamic range (which is typically 60 dB or less; Spahr et al. 2007) to the electric dynamic range (typically 10–20 dB; Zeng et al. 1998).

The goal of this article is to examine the efficacy on speech intelligibility for BCI users of enhancing the ILD cue by extending it into the low-frequency region. Although naturally occurring ILDs are very small at low frequencies (Fig. 1), headphone experiments in which low-frequency ILDs are applied manually have shown that normal-hearing listeners are as sensitive to low-frequency ILDs as they are to those at high frequencies* (Yost & Dye 1988). Francart et al. (2009) have shown that adding larger-than-normal ILDs in the lower frequencies can improve localization for simulated bimodal listeners.

Binaural enhancement was achieved in the current study by estimating instantaneous ITDs in the low-frequency region, which are present naturally yet poorly represented by current CI technology, and converting them to low-frequency ILDs, which are not present naturally but should be useable by BCI patients. In the case of a single stationary talker, the instantaneous ITD does not change over time. However, when there are two spatially separated modulated sound sources, such as two talkers speaking concurrently, the instantaneous ITD will change, depending on the relative levels of each talker at a given moment (Yost & Brown 2013).

The ultimate goal of this work is to implement this ITD-to-ILD conversion algorithm in a real-time device, in order to allow BCI users to adapt to this type of processing. Although work has begun on this goal, all processing occurred offline for the current study.

MATERIALS AND METHODS

Subjects

Eight BCI patients participated (see Table 1 for demographic information). All procedures were approved by the institutional review board at the University of Pittsburgh. When users' CIs were used along with headphone stimulus delivery, no changes to their everyday

*Interaural differences of levels can be delivered via headphones by simply attenuating the left or right channel, and thus can be imposed on any type of stimulus, including those that contain only low-frequency energy. In the same way, interaural differences of times can be applied to any signal by simply applying a delay to the signal in either the left (image to the right) or right (image to the left) ear.

programs were made. When a research processor was used with a direct-connect setup, the subjects' maps were used.

Speech Materials

The target speech was drawn from the CUNY set (Boothroyd et al. 1985), produced by a female talker, and the maskers were produced by a different female (IEEE 1969). When needed for duration, multiple masker tokens were concatenated, and the signal-to-noise ratio was +2 dB. No target or masker token was heard more than once.

Processing

As stated previously, all processing occurred offline, prior to testing. This means that both the binaural cues and the enhancement were applied and the signals were then delivered to the speech processors, either acoustically via headphones to the microphones of the CIs or via a direct-connect setup with a research processor.

Natural Binaural Cues—Prior to testing, a broadband noise was recorded through the left and right ears of a KEMAR, from 90 degrees and at a distance of about 5.47 ft, in an 11 × 13 ft sound-treated room (broadband RT60 = 90 msec). Each recording was then filtered into 32 ERB-wide bands having center frequencies between 128 and 8192 Hz, and an ILD was computed in each band (Fig. 1). This ILD-by-frequency function was applied to a target or masker token by filtering the token into the same 32 bands, and in each band, applying the computed frequency-specific ILD value. After summing the bands, an ITD of 600 μs was applied. Cues were applied so that the apparent location of target tokens was to the left, and maskers to the right. Finally, the left and right channels of the target and masker were summed.

Binaural Enhancement—Binaural enhancement was achieved using the following procedure. A target-plus-masker mixture carrying the natural binaural cues described previously was filtered into 32 ERB-wide bands as mentioned earlier. Enhancement was applied to the lowest 20 bands (band 20 had a cf of 2248 Hz). Each band was segmented into 20 msec bins with no overlap. In each bin, a sliding lag cross-correlation algorithm was applied, with a 1.2-ms (±600 μs) window size and no overlap. For each 20-ms bin, the between-channel delay that produced the largest correlation was taken as the instantaneous ITD. Thus, an ITD was computed every 20 msec. The ITD values were then converted to ILDs using a simple conversion table in which an ITD of ±600 μs corresponded to an ILD of ±30 dB and an ITD of 0 μs corresponded to a 0 dB ILD. Intermediate values were linearly interpolated (e.g., an ITD of ±300 μs corresponded to an ILD of ±15 dB). The instantaneous ILDs were then applied to the signal within the particular band, and all bands were summed. Thus, the enhanced stimulus contained both the natural cues (ITD and ILD information), and the enhanced (low-frequency ILD information). ILDs were always achieved by reducing the level at an ear, never by increasing the level.

Informal listening showed that analysis bin sizes of less than 20 msec did not lead to noticeable improvements in perceived lateralization, but it did increase the amount of transients (clicks) generated in the resulting signal, due to large changes in level that

occurred most often at moments of time when the levels of the two sources were relatively similar. In these cases, the ITD estimates were often spurious, which led to the large-level changes. Although future studies will directly examine the impact of parameters such as bin size and the ITD-to-ILD table on performance, the goal of the current study is to demonstrate a proof of concept, and, thus, they were not manipulated systematically here.

Conditions

In one condition, a target-plus-masker mixture was created with no binaural cues applied and presented diotically (diotic). In another condition, the natural binaural cues described earlier were applied (natural). In the third condition, both the natural cues, and the binaural enhancement described earlier were applied (enhanced).

Procedure

Participants were first presented with 10 unprocessed target sentences for stimulus familiarization and to achieve a comfortable listening level. They were then instructed to listen for that talker, to repeat what she said as best they could, and to ignore the masker. They were told that in some conditions the target talker might appear to be to the left of midline and the masker to the right. The presentation order of the conditions was randomized across subjects, and participants heard the same order of conditions twice, with 10 target sentences (50 keywords) per block. For example, a presentation order might be “enhanced, diotic, natural, enhanced, diotic, natural.” Presenting the random order twice was done in an attempt to control for any short-term adaptation that might occur. Thus, scores were derived from 20 sentences or 100 keywords per condition. No target or masker token was heard more than once by a participant. For six patients, stimuli were delivered via Sennheiser HD250 headphones, which were placed comfortably over each device’s microphone, and were not moved or removed until the completion of the experiment. Each of these participants’ everyday programs were used. For two patients, the UT-Dallas research processor was used in offline mode, which allowed streaming of stimuli from a computer in a manner equivalent to using an accessory cable in “direct-connect” mode with commercially available speech processors. For more information about the UT-Dallas processor, see the study by Ali et al. (2013). All procedures were approved by the institutional review board at the University of Pittsburgh.

RESULTS

The results are shown in Figure 2. Data are plotted as percent correct speech intelligibility in the presence of each type of binaural cue (none, natural or enhanced). Data from patients receiving direct stimulation are plotted with dashed lines and filled symbols. Mean speech intelligibility was 19% correct when no binaural cues were available, 34% correct when natural cues were available, and 65% correct with enhanced cues. Although the sample size is too small for inferential statistics, the pattern of results is clear and demonstrates that speech intelligibility can be improved for BCI users by enhancing the binaural cues available to them, at least when the signal is on one side of the head and the masker is on the other side and the levels of the signal and masker are different.

DISCUSSION

Participants averaged 15 percentage points of benefit to speech intelligibility with natural cues, and about 46 percentage points of benefit with the enhanced cues. Stated another way, binaural enhancement provided an additional 31 percentage points of speech intelligibility over the binaural cues that are typically available to BCI users.

Although the sample size for the current study is small ($n = 8$), all subjects show the same pattern of results; that is, all eight participants benefited from the enhancement. This was the case even for the two patients for whom the natural cues alone provided no benefit (BCI patients 2 and 5). The enhancement was also effective for the two prelingually deafened participants, although these two participants showed the smallest benefit from enhancement (18 percentage points for BCI patient 3, and 14 percentage points for patient 5), which is not surprising. The other six patients averaged 36 percentage points of benefit due to enhancement.

Anecdotally, all participants preferred listening to the enhanced stimuli, stating that it was easier to attend to either talker. For normal-hearing listeners listening informally, the enhanced processing produces apparent sound sources that are more lateralized than the natural cues.[†]

In a sense, the proposed method works by taking advantage of (and in some ways, enhancing) the amplitude modulation inherent in the speech signals (Yost & Brown 2013). When the level of the left talker is higher than the level of the right talker, the algorithm is more likely to detect an instantaneous ITD that tracks the left sound source. As a consequence, the signal in the right ear is attenuated. Likewise, the right talker being louder than the left talker will result in attenuation of the signal presented to the left ear. If we consider only the left ear as an example, the result is that when the level of the left talker is low, the signal to the left ear is low even if the level of the right talker is high, resulting in an improved signal-to-noise ratio for the left talker in the left ear. And the same improvement occurs for the right talker in the right ear. However, the fact that the algorithm, at least in theory, benefits from modulation present in the masker is an interesting one, given that modulated maskers (competing talkers in particular) are usually the most difficult to deal with for typical noise reduction techniques.

Nevertheless, this concurrent benefit for both talkers may not hold for spatial configurations other than the one tested; that is, discussion of the results must be qualified in that they were obtained when the target was to the left and the masker was to the right. It remains to be seen what the nature of the benefit will be in other spatial configurations or with more than one masker. One particular advantage of the current technique, however, is that no a priori knowledge of the sound sources is required. More specifically, in the case of the two

[†]Pilot data were collected with four normal-hearing listeners, who were presented tokens processed in a similar way to the natural and enhanced conditions in the current study. They were asked to judge the laterality of the two concurrent talkers, such that a value of -10 indicates a talker is fully lateralized to the left, $+10$ fully to the right, and 0 indicates a talker is centered in the head. Subjects thus made two estimates per token, one for each talker. For the natural condition, mean scores were -3.4 and $+3.8$. In the enhanced condition, mean scores were -10 and $+10$. These pilot data indicate that, at least for normal-hearing listeners, the enhancement algorithm produces greater lateralization than the natural cues.

concurrent talkers used in the current configuration (one to the left and one to the right), it does not matter which is the target and which is the masker: The enhancement is achieved for both sound sources, and the listener is free to attend to either or to switch attention as desired. There may be less benefit with other configurations. Indeed, given the apparent reliance of the current approach on amplitude modulation, it seems reasonable to expect that as the number of maskers (competing talkers) increases, the amount of benefit might decrease. There may be some configurations in which the proposed approach is actually detrimental.

The algorithm has not been optimized in any way, and produces audible clicks and pops, as a result of occasional large-level changes that occur most often at moments when the energy from each token is equal, which can produce spurious instantaneous ITD estimates. While a few BCI participants noted that they could hear this noise when asked about it, none volunteered this information, and all said that it was reasonably tolerable and not very distracting. Others reported not hearing the clicks. When the enhanced stimuli were vocoded (which simulates CI processing) using both noise-band and sinusoidal carriers, the clicks were inaudible to several listeners with normal hearing. One patient also commented that the overall percept seemed to be jumping back and forth in his head. This makes sense given the way the processing works. Nevertheless, this was not deemed by the patient to be too distracting, and the patient volunteered that he thought he would be able to get used to it given more exposure.

One possible limitation of the current study is the use of headphones for stimulus delivery rather than the use of direct stimulation. This may have allowed such things as headphone placement to influence the actual ILDs delivered to the user. However, care was taken in headphone placement to ensure a comfortable fit, and all subjects receiving headphone delivery reported (and were observed to have) no need to adjust headphones throughout the experiment. The study was also designed to be as brief as possible to this end, and all comparisons in the current study are within subjects. Finally, stimulus delivery to two of the eight participants occurred via the UT-Dallas research processor in a direct-connect mode, which is equivalent to using accessory audio cables with commercial devices. While a comparison between groups of such small sizes comes with a number of caveats, there are no obvious differences in the patterns of results between the two groups. Thus, while not ideal, it seems reasonable to conclude that the results of the current study are valid despite the use of headphones for the majority of participants.

Another possible limitation of the current study involves the use of simulated binaural cues. This was necessary because a real-time implementation of the algorithm does not yet exist, so the experiment could not be conducted in a real room. It may be that the “natural” ILD and ITD information applied in the current study somehow did not accurately simulate or perhaps underrepresented the cues that occur in a real room. However, pilot data indicated that the natural cues were a reasonable simulation of the binaural cues that occur naturally.‡

It is important to note that synchronization between BCIs would be necessary for the current binaural enhancement algorithm to be implemented. Although certainly technically feasible, no such synchronization is currently available commercially.

As mentioned, one goal of this work is to implement binaural enhancement in real time. The algorithm is computationally rather simple and, thus, feasible for a real-time application. Work has already begun to achieve this. Additional work is also needed to optimize the various parameters of the algorithm used in the current study. For example, the ITD-to-ILD map used is likely less than ideal. Perhaps somewhat counterintuitively, it may be beneficial to use smaller ILDs for a given ITD than those used currently. This may prove helpful in situations in which both sound sources are separated in space, but on the same side of midline. In this configuration, imposing a larger-than-necessary ILD would actually decrease the perceptual separation between the sources. Analysis bin size (which was 20 msec in the current implementation) is another candidate for optimization. Although shorter bin sizes were piloted, larger bins should also be examined to further reduce the computational requirements. The goal would be to find the largest bin size that does not negatively affect performance.

While the sample sizes are small and the present data were obtained under relatively restricted conditions (signal and masker on opposite sides of the head and a signal-to-masker ratio of +2 dB), this study represents a proof of concept of how binaural enhancement, and subsequent improved benefit to speech intelligibility, may be achievable in BCI users.

Acknowledgments

The author thanks Kate Helms Tillery and Hussnain Ali for invaluable technical assistance. Louise Loiseau is also thanked for contributing the realroom spatial release from masking data referred to in footnote 3, which were collected as part of her doctoral dissertation. Finally, the author thanks Philip Loizou for his helpful advice and guidance with this work. Philip will be dearly missed.

This research was supported by an NIDCD grant (R01 DC008329) awarded to the author and has emerged from ongoing collaborative work with William A. Yost.

References

- Ali H, Lobo AP, Loizou PC. Design and evaluation of a personal digital assistant-based research platform for cochlear implants. *IEEE Trans Biomed Eng.* 2013; 60:3060–3073. [PubMed: 23674422]
- Boothroyd, A.; Hanin, L.; Hnath, T. A Sentence Test of Speech Perception: Reliability, Set Equivalence, and Short Term Learning. New York, NY: Speech & Hearing Sciences Research Center, City University of New York; 1985. (Internal No. RCI 10)
- Brown CA, Bacon SP. Fundamental frequency and speech intelligibility in background noise. *Hear Res.* 2010; 266:52–59. [PubMed: 19748564]
- Brown CA, Yost WA. Interaural spectral asymmetry and sensitivity to interaural time differences. *J Acoust Soc Am.* 2011; 130:EL358–EL364. [PubMed: 22088041]

‡Four of the bilateral cochlear implant (BCI) users from the current study were brought back to the lab, and spatial release from masking (SRM) was tested using two conditions having slightly different configurations that in the current study. In a diotic condition, the target talker was combined with two different female distractor talkers at a signal-to-noise ratio of -2, with no binaural cues applied. In a second condition, the two distractors had the natural cues applied as in the current study, while the target was diotic. Thus, the target was always along midline, and each talker was to one side (one to the left, one to the right). This configuration was chosen because real-room SRM data were recently collected from 11 BCI subjects under the same conditions in the listening room in which the interaural differences of level function was derived. Thus, the efficacy of the simulated cues can be directly assessed by comparing SRM using those cues to SRM derived in the actual room. The average SRM was 6.5 percentage points for the 11 BCI patients in the actual room, and 10.3 percentage points for the three BCI patients hearing simulated binaural cues (“natural” cues). The results thus confirm that the natural cues used in the current study were not under-estimating what listeners receive typically in a real room.

- Francart T, Van den Bogaert T, Moonen M, Wouters J. Amplification of interaural level differences improves sound localization in acoustic simulations of bimodal hearing. *J Acoust Soc Am.* 2009; 126:3209–3213. [PubMed: 20000934]
- Francart T, Wouters J. Perception of across-frequency interaural level differences. *J Acoust Soc Am.* 2007; 122:2826–2831. [PubMed: 18189572]
- Grantham D, Ashmead D, Ricketts T, Haynes D, Labadie R. Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using CIS+ processing. *Ear Hear.* 2008; 29:33–44. [PubMed: 18091105]
- IEEE. IEEE recommended practice for speech quality measurements. *IEEE Trans Audio Electroacoust.* 1969; 17:225–246.
- Laback B, Pok S, Baumgartner W, Deutsch WA, Schmid K. Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors. *Ear Hear.* 2004; 25:488–500. [PubMed: 15599195]
- Loizou PC, Hu Y, Litovsky R, Yu G, Peters R, Lake J, Roland P. Speech recognition by bilateral cochlear implant users in a cocktail- party setting. *J Acoust Soc Am.* 2009; 125:372. [PubMed: 19173424]
- Spahr AJ, Dorman MF, Loisel LH. Performance of patients using different cochlear implant systems: Effects of input dynamic range. *Ear Hear.* 2007; 28:260–275. [PubMed: 17496675]
- Yost WA, Brown CA. Localizing the sources of two independent noises: Role of time varying amplitude differences. *J Acoust Soc Am.* 2013; 133:2301–2313. [PubMed: 23556597]
- Yost WA, Dye RH Jr. Discrimination of interaural differences of level as a function of frequency. *J Acoust Soc Am.* 1988; 83:1846–1851. [PubMed: 3403800]
- Zeng FG, Galvin JJ, Zhang C. Encoding loudness by electric stimulation of the auditory nerve. *NeuroReport.* 1998; 9:1845–1848. [PubMed: 9665613]
- Zhang T, Dorman MF, Spahr AJ. Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear Hear.* 2010; 31:63–69. [PubMed: 20050394]

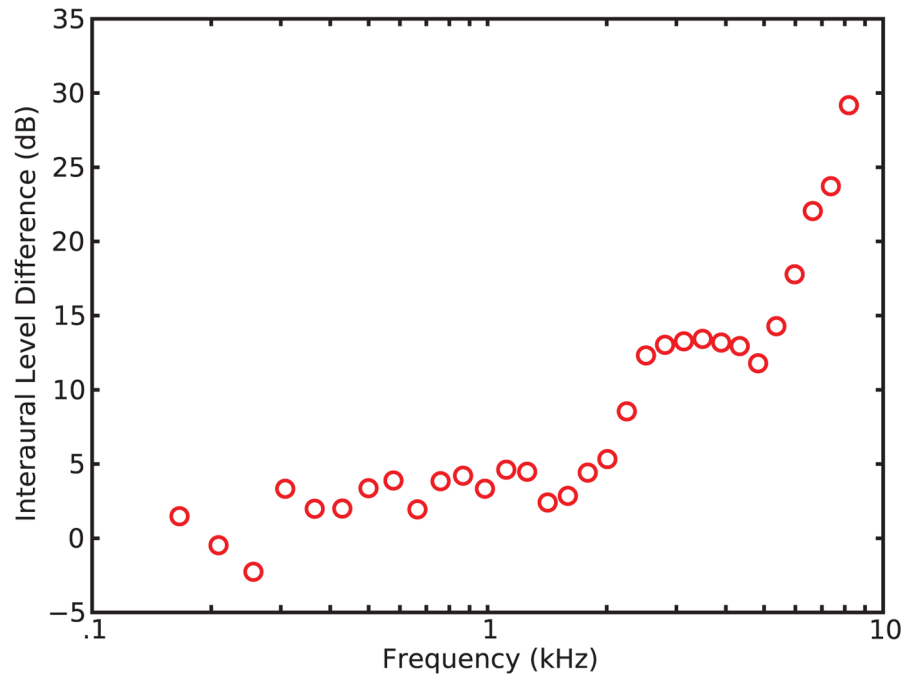


Fig. 1. Interaural differences of levels as a function of frequency for a broadband noise presented 90 degrees to one side, and recorded with a KEMAR.

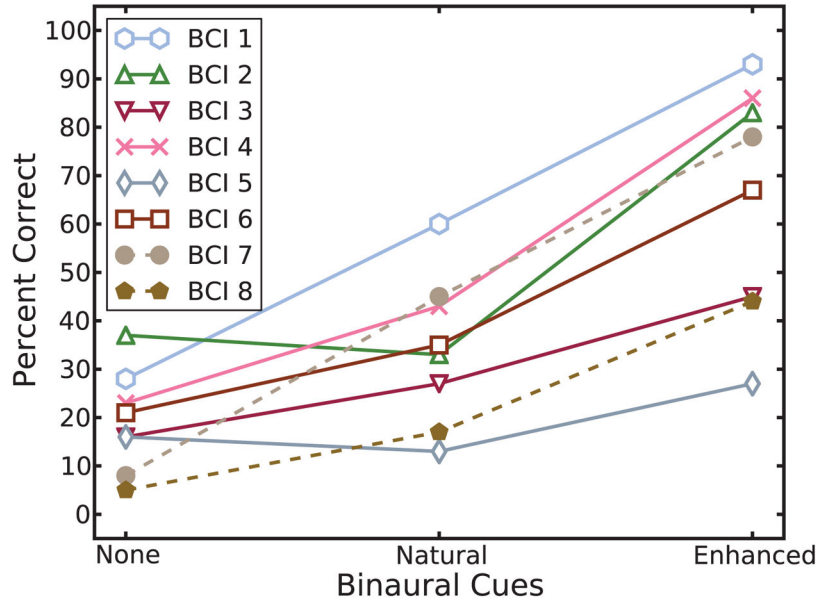


Fig. 2. Speech intelligibility for five BCI listeners in three conditions. In the “diotic” condition, no binaural cues were applied to either the target or the masker. In the “natural” condition, binaural cues were applied, which included the ILD-by-frequency function shown in Figure 1, and a whole-waveform ITD of 600 μ s. In the “enhanced” condition, the stimulus mixture created in the natural condition was subjected to the ITD-to-ILD algorithm, which applied low-frequency ILD cues based on the instantaneous ITD information. The difference in performance between the diotic and natural conditions represents the SRM that the participants would receive typically. The difference between the natural and enhanced conditions represents the additional SRM provided by the binaural enhancement. BCI indicates bilateral cochlear implant; ILD, interaural differences of levels; ITD, interaural differences of time; SRM, spatial release from masking.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 1

Demographic participant data

S. No.	Age	Left Ear		Right Ear		Etiology	Stimulus Delivery
		Device	Duration Implanted	Device	Duration Implanted		
1	58	Cochlear Nucleus 5	1 yr, 10 mo	Cochlear Nucleus 5	1 yr, 7 mo	Menieres/postlingual	Headphones
2	68	AB Harmony	12 yr, 8 mo	AB Harmony	5 yr, 5 mo	Hereditary/perilingual	Headphones
3	55	AB Harmony	6 yr, 2 mo	AB Harmony	11 yr, 6 mo	Maternal Rubella/prelingual	Headphones
4	81	Cochlear Nucleus 5	3 yr, 0 mo	Cochlear Nucleus 5	2 yr, 0 mo	Noise exposure/postlingual	Headphones
5	59	AB Neptune	1 yr, 0 mo	AB Neptune	1 yr, 0 mo	Meningitis/prelingual	Headphones
6	63	Cochlear Nucleus 5	9 yr, 4 mo	Cochlear Nucleus 5	8 yr, 8 mo	Meningitis/postlingual	Headphones
7	53	Cochlear Nucleus 5	4 yr, 3 mo	Cochlear Nucleus 5	2 yr, 5 mo	Unknown/postlingual	Direct stim
8	60	Cochlear Nucleus 5	2 yr, 5 mo	Cochlear Nucleus 5	1 yr, 2 mo	Hereditary/postlingual	Direct stim