PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Spatial release from masking in bilateral cochlear implant users listening to the Temporal Limits Encoder strategy

Alan KAN[1]; Qinglin MENG[2]

[1] University of Wisconsin-Madison, USA

[2] South China University of Technology, China

ABSTRACT

Binaural hearing benefits with bilateral cochlear implants (CI) are usually smaller than with normal hearing (NH) in the same tasks. This gap in performance has typically been attributed to a lack of coordinated stimulation between ears and the high stimulation rates used in clinical processors. These factors hinder sensitivity to interaural timing differences (ITDs); an important binaural cue for NH listeners. The Temporal Limits Encoder (TLE) strategy was originally designed to encode unilateral low-frequency temporal fine structure pitch cues into the signal envelope of CIs. However, TLE also lowers the stimulation rate on some channels which may potentially provide useable ITD cues. Here, we measured spatial release from masking (SRM) in bilateral CI users listening to TLE vs Advanced Combinational Encoder (ACE) strategy to determine if TLE provides a benefit. The CCiMobile research platform was used for testing. Results from eight listeners showed comparable word recognition performance in quiet and co-located conditions for TLE and ACE, even with a short acclimatization period with TLE. In the spatially-separated condition, performance across the group was more similar with TLE than ACE, and more listeners showed SRM. These results indicate that TLE has the potential for improving binaural hearing benefits for CI users.

Keywords: Cochlear Implants, Sound Coding, Spatial Release from Masking

1. INTRODUCTION

Cochlear implants (CIs) have been largely successful in restoring speech understanding in patients who are profoundly deaf in both ears, especially in quiet situations (1,2). However, in noisy situations, the ability to understand speech with one CI is very poor (1), and bilateral cochlear implantation is fast becoming the standard of care (3). While speech understanding in noise has been shown to improve with bilateral CIs (4), the benefits of having two ears for listening are still much smaller than that of normal hearing (NH) listeners (5,6). To assess the benefits of bilateral hearing on speech understanding in noise with devices such as CIs or hearing aids, spatial release from masking (SRM) is often measured (e.g., 4,7). SRM is defined as the difference in signal-to-noise (SNR) needed to understand speech in noise due to the spatial separation of a target signal from masking interferers. The benefit of spatial separation is a combination of monaural and binaural effects, namely head shadow, squelch and summation (6). Head shadow is largely a monaural effect and arises because the head acts as an acoustic shadow to improve the SNR in one ear. Squelch and summation are binaural effects; squelch is dependent on the auditory systems ability to utilize binaural cues such as interaural time and level differences (ITD and ILD, respectively), and summation is the benefit of having access to two signals that can be combined to enhance the audibility of a target signal in front of the listener.

For CI users, the benefits of bilateral CIs have largely been due to the monaural head shadow effect and the binaural effects of squelch and summation have been much smaller than that of NH listeners (5). Smaller binaural benefits in bilateral CI users have largely been attributed to the lack of access to ITDs because of the way sound is encoded by CI sound coding strategies (6). CIs encode an incoming acoustic signal by filtering it into a small number (between 12 and 24 depending on manufacturer) of bandlimited channels. The signal in each bandlimited channel can be decomposed into a slow-varying envelope and a faster temporal fine structure (TFS). In CIs, only the envelope is used to

[1] ahkan@waisman.wisc.edu
[2] mengqinglin@scut.edu.cn

amplitude-modulate a high rate electrical pulse train (≥900 pulses per second) while the TFS is often discarded. At the low frequencies, the TFS provides important cues used by NH listeners for sound localization (8,9).

Attempts to improve speech-in-noise understanding for bilateral CIs have focused on developing sound coding strategies that encode TFS (10–12). To encode TFS, the pulse rate of low frequency channels is usually reduced and firing of electrical pulses are timed to certain features of the acoustic TFS, such as the signal peak or zero crossing. The reduction in pulse rate at these channels is thought to be necessary because psychophysical studies have shown the ITD sensitivity is better at rates around 100 pulses per second (13,14). However, these approaches have only led to small improvement in binaural benefits (15).

Recently, the temporal limits encoder (TLE) strategy was proposed to try to enhance TFS pitch cues in CIs (16,17). The TLE strategy is different to other TFS-encoding strategy in that it does not explicitly lower the stimulation rate, nor does it rely on encoding a "feature" of the TFS such as a peak or zero crossing. Instead, the TLE strategy transposes each bandlimited channel to an intermediate, lower frequency range (e.g., 150-350 Hz) which is within the range of good temporal sensitivity for both pitch and ITD. The transposed signal maintains the overall envelope but also introduces a slower modulation that is related to the TFS into the signal envelope. Experiments in NH listeners listening to CI vocoder simulations of the TLE strategy has shown some advantages in pitch and unilateral speech-in-noise tasks (16,17). Further, TLE provided a ~4 dB spatial unmasking benefits in NH listeners in a task measuring binaural intelligibility level differences (18). In this work, we measured SRM in bilateral CI users when listening with the TLE strategy vs the Advanced Combinational Encoder (ACE) strategy.

2. METHODS

2.1 Listeners

Eight (7 females and 1 male), postlingually-deafened bilateral CI users with Cochlear Ltd implants were tested. All listeners used ACE as their everyday sound coding strategy. Listeners were between the ages of 51 and 78 (Mean age: 62 years old) and had at least 7 years of bilateral CI experience. All listeners had prior experience with psychophysical testing and had measurable ITD sensitivity when measured with direct electrical stimulation at low pulse rates (100 pulses per second). Listeners were tested at the University of Wisconsin-Madison and were paid a daily stipend for their time. Experimental procedures conformed to the regulations set by the National Institutes of Health and were approved by the University of Wisconsin-Madison's Health Science Institutional Review Board.

2.2 Strategy Implementation

Testing was conducted using the CCiMobile research platform (19) connected to a Windows Surface Pro. The CCiMobile is the second generation of a bilaterally-synchronized, real-time research platform developed by UT-Dallas for Cochlear Ltd internal devices. A description of the hardware can be found in (19). A real-time implementation of ACE written in MATLAB is provided by UT-Dallas (20). We adapted the MATLAB code to implement the TLE strategy for this experiment, as well as creating a new user interface that allows us to easily switch between strategies.

Figure 1 shows the signal block diagram. The same processing is applied to the right and left ears. The CCiMobile samples the incoming audio at a 16 kHz rate and buffers the audio for 8 ms before passing to MATLAB. Overlap-add processing is applied in 128-point sample frames and frames are shifted according to the stimulation rate set by the patients clinical MAP (usually 900 pulse per second, two listeners had 1200). A 128-point Fast Fourier Transform (FFT) is applied to each frame. The FFT bins are then combined into 22 channels according to the frequency allocation set in the patient MAP. For ACE, the N highest maxima are chosen from the 22 channels in each ear, where N is set by the patient MAP. All listeners had N=8 except one who had N=10. The N maxima were then compressed to fit into the patient's dynamic range and added to the stimulation buffer. Once 8 ms worth of pulses had been processed, the stimulation buffer was then passed to the CCiMobile.

For TLE, a frequency transposition step is added prior to peak-picking. Frequency transposition is achieved in the FFT domain by multiplying the FFT bins within the range of 300 to 1500 Hz by $e^{j\pi(N-1-2ft)}$, where N is the index of the FFT frequency bin, t is the time index, and f is the down-modulator frequency. In this experiment, f was set to the lower frequency cutoff of the channel minus 150 Hz. The down-modulated FFT bins are then combined into 22 channels the same way as that of ACE.

The output of the two strategies are shown in Figure 2 for the word "Jane" which has been filtered by a head-related transfer function corresponding to a source 90° to the right-hand side. It can be seen that stimulation on the lower frequency channels of the TLE strategy are sparser but the overall signal envelope is maintained. Further, there is an ITD introduced into the lower frequency channels in the TLE strategy (lower panels of Figure 2).

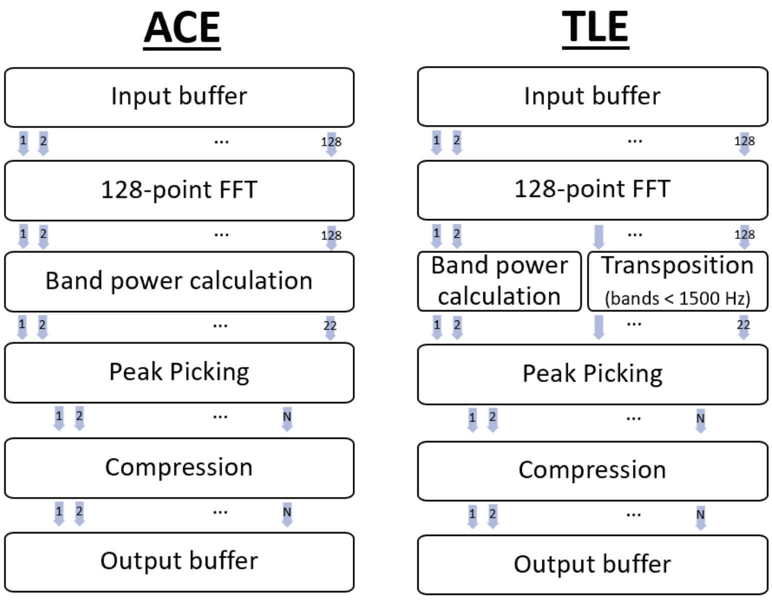

Figure 1 – Signal block diagram of ACE and TLE strategies

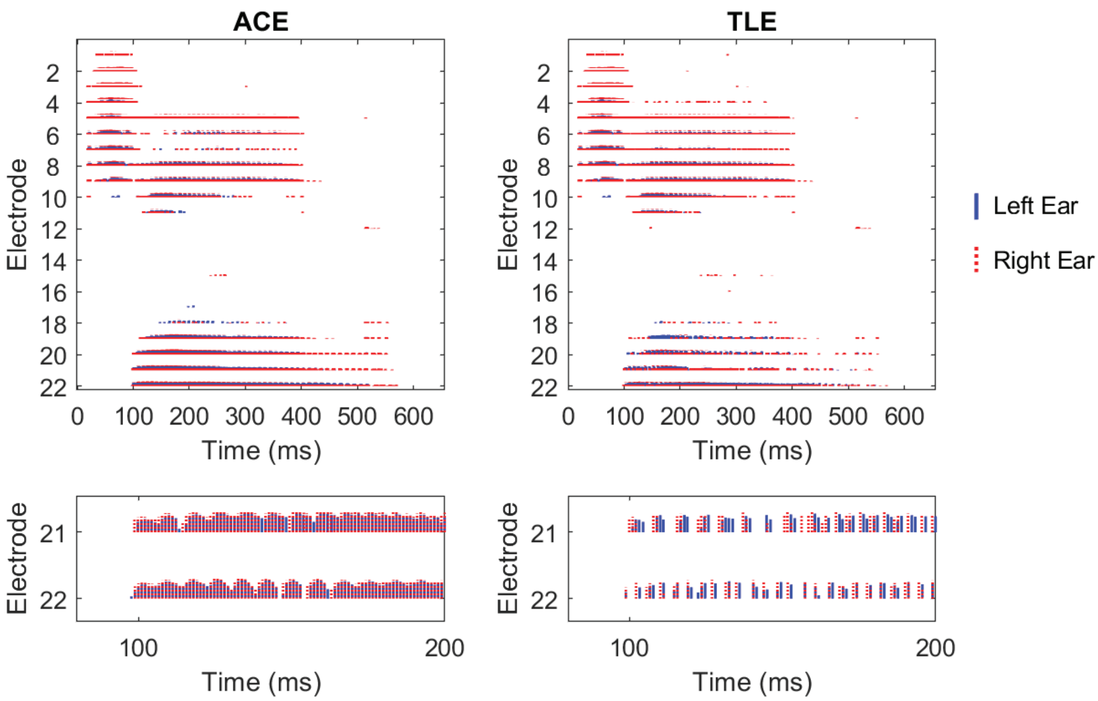

Figure 2 – Electrodogram of the word "Jane" spoken by a source located on the right-hand side. Each blue and red line represents an electrical pulse in the left and right ear, respectively. Upper panels show the entire electrodogram, while lower panels show a magnified portion of the stimulation on electrodes #21 and #22 (low-frequency electrodes).

2.3 Testing Procedure

Experiments were conducted in a single-walled soundproof booth. The booth had sound absorbing foam attached to the inside walls to reduce reflections. Inside the booth was a semi-circular array of loudspeakers (Cambridge SoundWorks). The loudspeakers used in this experiment were at 0° (front), -90° (left) and +90° (right) on the horizontal plane. The listener sat in the middle of the array and was 1.2 m from the loudspeakers.

For the experiment, the listener's task was to identify the word presented from the loudspeaker in front of the listener. A choice of 50 words were shown on a touchscreen in front of the listener. The listener indicated their response by choosing the button corresponding to the word they heard.

Three conditions were tested: (1) word recognition in quiet; (2) co-located target talker and interferers – in this condition, both target and interferers were presented from the 0° loudspeaker; (3) spatially-separated target talker and interferers – in this condition, the target talker was presented from 0°, and an interferer was presented from each of the left and right loudspeakers.

The target talker was a male voice speaking a mono-syllable word, while the interferer was a female voice speaking sentences. In the quiet condition, words were presented at a fixed level of 60 dBA, and listeners was tested on 50 words divided into two blocks of 25 words each. In each block of trials, the choice of words shown on the screen was different. In the conditions with interferers, the interferer was presented at a fixed level of 50 dBA and always started/ended 250 ms before/after the target word. The presentation level of the target word was changed on each trial using a two-down, one-up adaptive staircase with twelve turnarounds. The staircase started at a +5 dB SNR and moved in 3 dB steps for the first three turnarounds. For the remaining turnarounds, the SNR changed in 2 dB steps.

Prior to testing, listeners were given at least 1 hour of practice listening with the TLE strategy. During this time, the CI user wore the CCiMobile device and engaged in conversation with the first author. No listener reported having difficulty with understanding the conversation when listening with the TLE strategy, though all listeners reported a noticeable difference in voice quality (lower pitch). After the 1 hour of practice, listeners would report that the first authors voice sounded more "normal", though there was still a noticeable difference in their own voice.

Word recognition in quiet was tested first and always began with the TLE strategy followed by the ACE strategy. Once testing of word recognition in quiet was complete, listeners were tested in the conditions with interferers. The order of spatial configuration and strategy was randomized for each listener, and two adaptive tracks were collected for each strategy/spatial configuration combination.

3. RESULTS

Percent correct word recognition in quiet is shown in Figure 3. On average, word recognition was quite comparable between the two strategies (86.5% vs 90.25% correct when listening with the TLE and ACE strategies, respectively). A paired *t*-test conducted on the percent correct scores revealed no significant difference between the two strategies ($t(7)=1.46$, $p=0.19$).

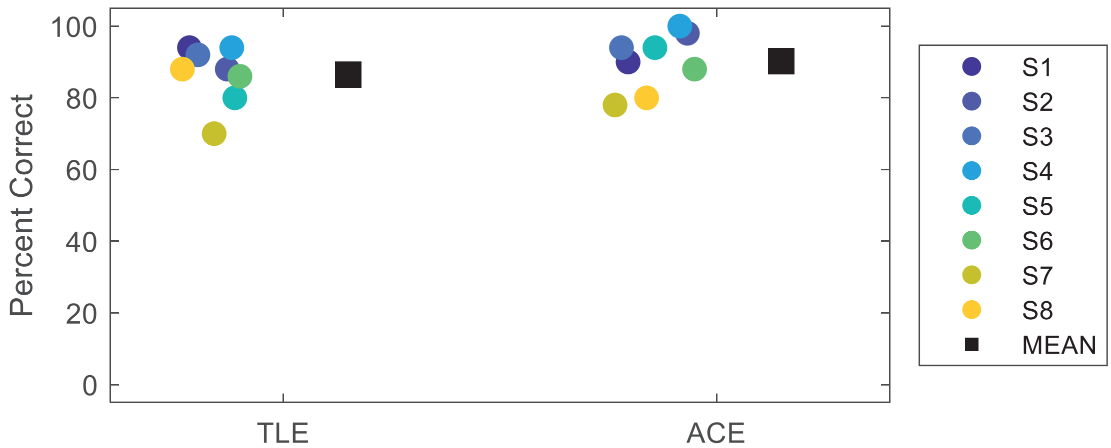

Figure 3 – Word recognition in quiet

For the word recognition in noise conditions, a 70.7%-correct speech reception threshold (SRT) was obtained for each condition using a Bayesian estimation of the psychometric function (21). Figure 4 shows the SRTs obtained in each condition for the two strategies. Substantial inter-subject differences can be observed. Given the resolution of the adaptive track is 2 dB, then in the co-located condition, four listeners (S1, S6, S7, S8) had SRTs that were similar between the two strategies, two had TLE SRTs that were more than 2 dB lower than ACE (S3, S4), and the remaining two had TLE SRTs that were more than 2 dB higher than ACE (S2, S5). In the spatially-separated condition, four listeners had TLE SRTs that were more than 2 dB lower than ACE (S1, S4, S6, S7), three had TLE SRTs that were more than 2 dB higher than ACE (S2, S3, S5), and one listener had SRTs that were within 2 dB (S8). As a group, mean SRTs in the co-located condition were 9.49 dB and 8.65 dB for TLE and ACE, respectively. Mean SRTs in the separated condition were 5.75 dB and 6.18 dB for TLE and ACE, respectively. Two-way, repeated-measures analysis of variance revealed a significant effect of spatial configuration ($F(1)=8.417$, $p=0.02$) but not strategy ($F(1)=0.026$, $p=0.88$).

SRM computed from the SRTs of the co-located and separated conditions are shown in Figure 5. Five listeners obtained an SRM greater than 2 dB when using the TLE strategy vs three listeners for ACE. One listener found it more difficult to listen with the TLE strategy when target and interferers were spatially separated. On average, SRM was 3.74 dB and 2.47 dB for TLE and ACE, respectively.

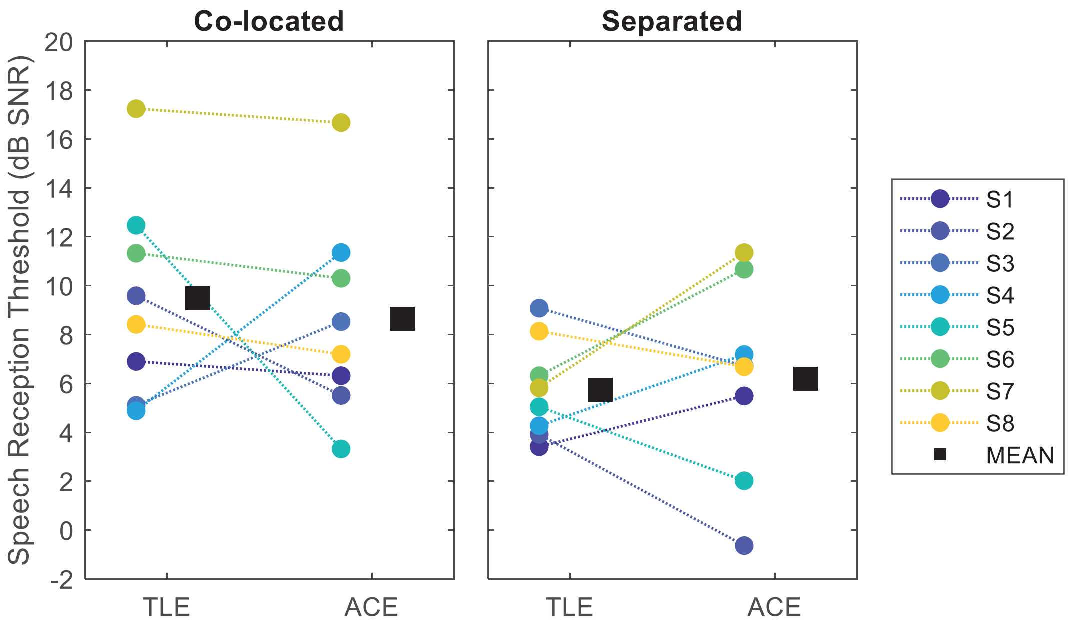

Figure 4 – Speech reception thresholds for the TLE and ACE strategies when tested with target-interferers co-located and spatially-separated.

4. DISCUSSION

This work evaluated whether the TLE strategy could provide a binaural benefit for CI users when trying to recognize words in a noisy environment. The TLE strategy was implemented on the CCiMobile research platform which allowed real-time live testing of the strategy. Informal comments obtained from listeners indicated that the TLE strategy was perceptually different to ACE and lowered the pitch of voices. This was somewhat to be expected given that the original goal of TLE was to convey pitch cues more saliently by reduction of stimulation rate. Prior psychoacoustic work has shown that rate of stimulation is able to convey a pitch cues (22,23). Most listeners reported that after an hour of listening with the TLE strategy, voices began sounding more “normal”, suggesting that given a longer duration of acclimatization, TLE can be an acceptable listening strategy. Formal testing of word recognition in quiet found no significant difference between TLE and ACE. This is a strong positive because it shows that TLE can maintain important speech envelope cues despite the lowered stimulation rate.

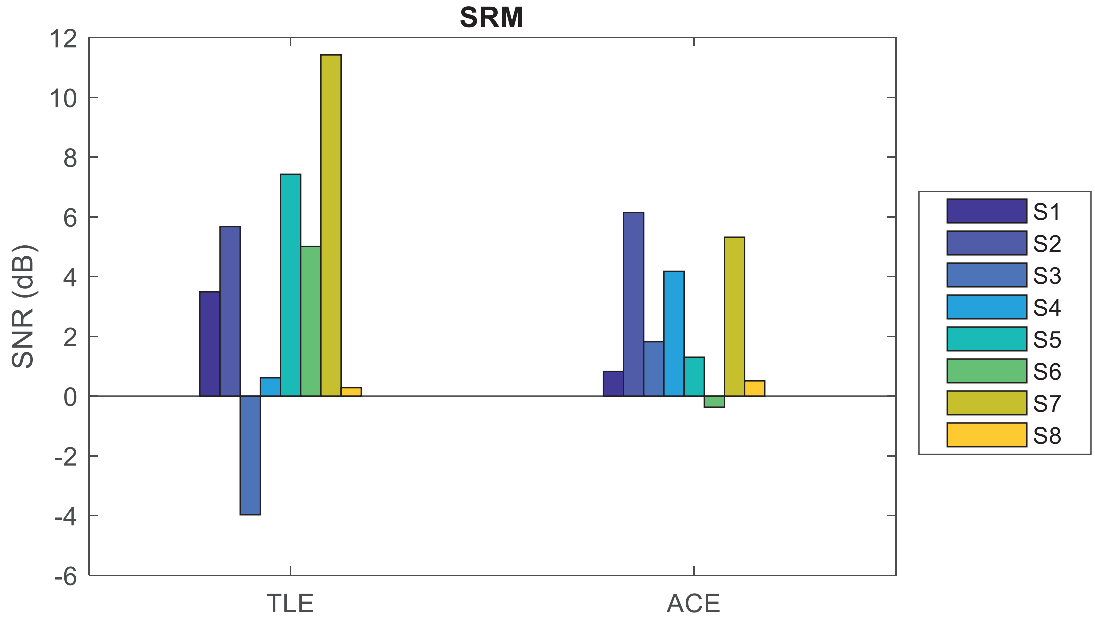

Figure 5 – Spatial release from masking for each listener when listening with the TLE and ACE strategies.

When listening in noise, TLE appears to provide reduced the variability in performance across bilateral CI listeners compared to ACE when target and interferers were spatially separated (see Figure 4). In addition, more listeners were able to benefit from SRM when listening with TLE compared to ACE. These are promising results and demonstrates a potential benefit of the TLE strategy for enabling the binaural benefits of spatial unmasking in bilateral CI users. More work is needed to understand the impact of choice of TLE parameters; in particular, the down-modulator bandwidth. Also, it is likely that if listeners were given a longer period of listening with TLE, a more consistent benefit across the group may emerge.

5. CONCLUSIONS

The TLE strategy was evaluated for the first time in CI users. Results showed that TLE can maintain speech understanding performance that is comparable to the ACE strategy and has the potential to provide a binaural benefit for speech understanding in noisy environments in a greater number of bilateral CI users.

ACKNOWLEDGEMENTS

The authors would like to thank the listeners for traveling to the University of Wisconsin-Madison to participate in this research. This work was funded by the National Institutes of Health-National Institute on Deafness and Other Communication Disorders (R03DC015321 to Kan), and in part by the National Institutes of Health-Eunice Kennedy Shriver National Institute of Child Health and Human Development (U54HD090256 to Waisman Center) and the National Natural Science Foundation of China (Grant No. 11704129 and 61771320 to Meng).

REFERENCES

1. Firszt JB, Holden LK, Skinner MW, Tobey EA, Peterson A, Gaggl W, et al. Recognition of Speech Presented at Soft to Loud Levels by Adult Cochlear Implant Recipients of Three Cochlear Implant Systems. Ear Hear. 2004 Aug;25(4):375–87.
2. Wilson BS, Dorman MF. The Surprising Performance of Present-Day Cochlear Implants. IEEE Trans Biomed Eng. 2007 Jun;54(6):969–72.
3. Peters BR, Wyss J, Manrique M. Worldwide trends in bilateral cochlear implantation. Laryngoscope. 2010 May;120 Suppl(5):S17-44.
4. Litovsky RY, Parkinson A, Arcaroli J. Spatial Hearing and Speech Intelligibility in Bilateral Cochlear

Implant Users. Ear Hear. 2009 Aug;30(4):419–31.
5. Loizou PC, Hu Y, Litovsky R, Yu G, Peters R, Lake J, et al. Speech recognition by bilateral cochlear implant users in a cocktail-party setting. J Acoust Soc Am. 2009 Jan;125(1):372–83.
6. Litovsky RY, Goupell MJ, Misurelli SM, Kan A. Hearing with Cochlear Implants and Hearing Aids in Complex Auditory Scenes. In: Middlebrooks JC, Simon J, Popper A, Fay R, editors. The Auditory System at the Cocktail Party Springer Handbook of Auditory Research, Vol 60. Cham, Switzerland, Switzerland: Springer; 2017. p. 261–91.
7. Neher T, Behrens T, Carlile S, Jin C, Kragelund L, Petersen AS, et al. Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition. Int J Audiol. 2009;48(11):758–74.
8. Wightman FL, Kistler DJ. The dominant role of low-frequency interaural time differences in sound localization. J Acoust Soc Am. 1992 Mar;91(3):1648–61.
9. Macpherson EA, Middlebrooks JC. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. J Acoust Soc Am. 2002 May;111(5 Pt 1):2219–36.
10. van Hoesel RJM. Exploring the Benefits of Bilateral Cochlear Implants. Audiol Neurotol. 2004;9(4):234–46.
11. Hochmair I, Nopp P, Jolly C, Schmidt M, Schösser H, Garnham C, et al. MED-EL Cochlear implants: state of the art and a glimpse into the future. Trends Amplif. 2006 Dec;10(4):201–19.
12. Churchill TH, Kan A, Goupell MJ, Litovsky RY. Spatial hearing benefits demonstrated with presentation of acoustic temporal fine structure cues in bilateral cochlear implant listeners. J Acoust Soc Am. 2014 Sep;136(3):1246–56.
13. Kan A, Litovsky RY. Binaural hearing with electrical stimulation. Hear Res. 2015 Apr 2;322:127–37.
14. Laback B, Egger K, Majdak P. Perception and coding of interaural time differences with bilateral cochlear implants. Hear Res. 2015 Apr 19;322:138–50.
15. Zirn S, Arndt S, Aschendorff A, Laszig R, Wesarg T. Perception of Interaural Phase Differences With Envelope and Fine Structure Coding Strategies in Bilateral Cochlear Implant Users. Trends Hear. 2016 Sep 22;20.
16. Meng Q, Zheng N, Li X. A temporal limits encoder for cochlear implants. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2015. p. 5863–7.
17. Meng Q, Zheng N, Li X. Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants. J Acoust Soc Am. 2016 Jan;139(1):301–10.
18. Meng Q, Wang X, Zheng N, Schnupp JWH, Kan A. Binaural Hearing Measured with the Temporal Limits Encoder using a Vocoder Simulation of Cochlear Implants. Acoust Sci Technol. 2019;in press.
19. 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 Nov;60(11):3060–73.
20. CILab:CRSS. CCi-MOBILE [Internet]. 2019. Available from: https://crss.utdallas.edu/CILab/CCi-MOBILE.html
21. Schütt HH, Harmeling S, Macke JH, Wichmann FA. Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vision Res. 2016 May;122:105–23.
22. Townshend B, Cotter N, Van Compernolle D, White RL. Pitch perception by cochlear implant subjects. J Acoust Soc Am. 1987 Jul;82(1):106–15.
23. Kong Y-Y, Deeks JM, Axon PR, Carlyon RP. Limits of temporal pitch in cochlear implants. J Acoust Soc Am. 2009 Mar;125(3):1649–57.