INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Cochlear-implant Listeners Listening to Cochlear-implant Simulated Speech

Fanhui Kong[1], *Nengheng Zheng*[1], *Xianren Wang*[2], *Hao He*[3], *Jan W. H. Schnupp*[4], *Qinglin Meng*[5]

[1]Guangdong Key Lab of Intelligent Information Processing, CEIE, Shenzhen University, China
[2]The First Affiliated Hospital, Sun Yat-Sen University, China
[3]College of Psychology and Sociology, Shenzhen University, China
[4]Department of Neuroscience, City University of Hong Kong, Hong Kong SAR of China
[5]School of Physics and Optoelectronics, South China University of Technology, China

nhzheng@szu.edu.cn, jan_hk@schnupp.net, mengqinglin@scut.edu.cn

Abstract

Channel vocoders with noise and sine-wave carriers are widely used to simulate modern multi-channel cochlear implants (CIs) in psychoacoustic experiments with normal hearing (NH) subjects. NH subjects perceive vocoded speech as impoverished and unnatural, but how CI listeners perceive vocoded sounds has not been systematically investigated. This letter reports that CI listeners could equally recognize both noise and sine-wave vocoded speech, albeit less well than NH listeners, and the recognition performance would not significantly increase beyond 8 channels. Nevertheless, they can easily discriminate up-to-80-channel vocoded speech from the original natural speech.

Index Terms: Cochlear implant, vocoder, temporal envelope, electric hearing

1. Introduction

Cochlear implants (CIs) are by now widely used as prosthetic devices to provide a sensation of hearing to patients with severe-to-profound hearing-impairment. Current CIs typically process incoming sounds via a bank of band-pass filters, and only the temporal envelope from each filter is preserved for electric stimulation. It is not possible to replicate the stimulation pattern delivered to auditory nerve fibers by CIs via acoustic stimulation in the normally hearing ear, but it can nevertheless be useful for research and demonstration purposes to simulate CI processing using envelope-based vocoders, with either bandlimited noise carriers [1] or sine-wave carriers [2]. Such vocoders degrade input sounds in ways which preserves envelopes in subbands but discards temporal fine structure (TFS) information, and are thus thought to deliver information of a similar quality to normal hearing (NH) subjects as CI users would experience. Vocoder simulation experiments have been reported in a very large number of papers (e.g. see Kong et al. [3] and Table I of Stone et al.[4]), to investigate numerous parameters of CI design and use, including channel number [5, 6], modulation rate [7], intensity resolution [8], electrode insertion depth [9, 10] and frequency allocation [11, 12], spatial hearing [13, 14], fundamental frequency discrimination [15, 16] and its contribution to speech segregation [17, 18], lexical tone perception [19, 20, 21], natural sound distortion perception [22, 23], or the evaluation of novel envelope-based strategies [24, 25]. A demonstration can be found in [26].

Vocoded speech not only lacks TFS, but also spectral fine structure because the number of subbands is usually not large enough to resolve much spectral detail, such as fine harmonics. How much of spectral fine structure is preserved can be made to vary by adjusting the number and bandwidth of subbands used in the vocoder. This makes vocoding a useful tool for sound perception research which has found applications beyond simulating CI hearing in NH subjects, including, for example in studies of perceptual and cognitive aging [27], toddlers' language learning [28], and even animal auditory perception [29]. Nevertheless, by far the most widely used application of vocoding remains to "simulate hearing with CI". Curiously, there appear to be few studies that have asked how CI users perceive vocoded sounds. NH subjects invariably find that vocoded speech sounds highly unnatural and it can be hard to understand, but if vocoded sound is a "good simulation of CI hearing", one might expect that CI subjects should find vocoded speech to be as easily intelligible, and perhaps also to sound as natural as, normal speech, provided that the vocoder is set up with appropriate parameters, such as a sufficiently large number of subband channels. Here we investigate whether this expectation is justified by measuring how well CI users can understand vocoded speech with a range of different parameters (Experiment 1) and how well they can discriminate natural from vocoded speech (Experiment 2). For comparison, an NH control group performed the same experiments.

Presenting vocoded "CI-simulation" sounds to CI users may at first glance seem an odd thing to do. We did not find any other published experiments on this topic in the literature. One tangentially related study investigated temporal envelope cues recovered from speech frequency modulation (e.g, Won et al. [30]), which both CI and NH listeners were reportedly able to do, and individual differences in the performance of that task may partly explain speech perception performance variability in CI users. Our vocoder experiments reported here are somewhat analogous, in that they also measure the ability of NH and CI listeners to use temporal envelopes cues to process speech. But that aspect of our study is not particularly novel. What motivated this work was rather that we felt that the near-ubiquitous use of vocoded sounds to simulate CI hearing ought to be "validated", in as far as it is possible to do so, by examining to what extent the perception of these "simulated" speech sounds, be it by NH or by CI subjects, is indeed "similar" to the way CI subjects perceive real speech. To an extent this validation was successful, in that speech recognition scores for CI and NH listeners showed similar asymptotes toward normal speech recognition scores for increasing channel number, with a "knee" at around 8 channels, and CI and NH listeners recognition scores were similarly independent of carrier type. However, if vocoded speech was a "highly similar simulation" of the way CI listeners perceive speech, then one might expect that CI users might not find it easy to distinguish vocoded speech from natural speech. However, our results show that, even though they do not match the near perfect and seemingly effortless performance of NH subjects, most CI users nevertheless find distinguishing normal speech from vocoded speech relatively easy in most cases.

10.21437/Interspeech.2023-554

Table 1: *CI user demographic information, hearing history, and device information*

Subject	Gender	Age	CI experiences (year)	CI Processor	Etiology	Exp.1	Exp.2
C1	F	21	17	R: Cochlear N5	Congenital	no	yes
C2	M	24	15	L: Cochlear Kanso	Drug-induced	yes	yes
C18	M	25	21	R:Cochlear N5	Congenital	yes	no
C20	M	10	8	R: Cochlear Freedom	Congenital	yes	yes
C21	F	34	7	R: Cochlear CP900	Drug-induced	yes	yes
C28	F	40	10	R: Cochlear N5 L: Cochlear Kanso	Drug-induced	yes	yes
C30	F	23	1	R: Cochlear Freedom	Sudden deafness	yes	yes
C31	F	12	10	L: Med-El OPUS 2	Congenital	yes	no
C32	F	11	9	R: AB Harmony	Congenital	yes	no
C33	M	16	10	R: AB Harmony	High fever	yes	no
C34	M	15	13	R: Med-El	Congenital	yes	yes
C35	M	15	11	R: Cochlear Sprint	Congenital	yes	no
C37	M	12	10	R: Cochlear N5	Jaundice	no	yes
C41	F	15	9	R: Cochlear N5	Unknown	no	yes
C49	M	11	8	R: Cochlear N5	Congenital	no	yes
C50	M	21	17	R: Cochlear N5	Gentamicin allergy	no	yes

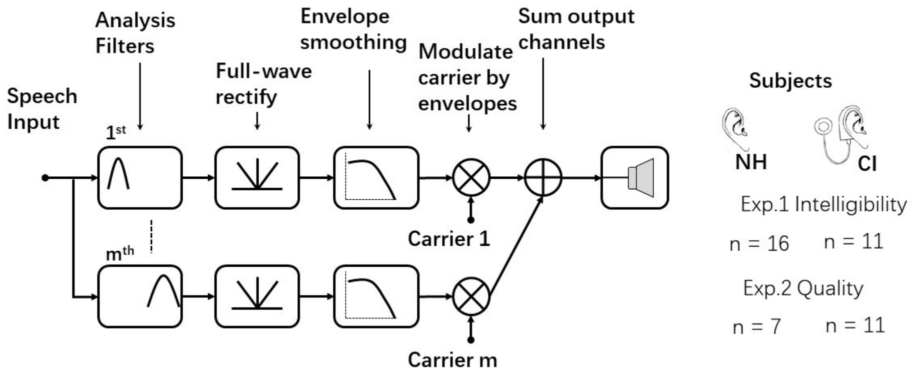

Figure 1: *Block diagram showing signal processing in the vocoders.*

2. Methods

2.1. Subjects

Twenty NH subjects (college students) and 16 CI subjects (listed in Table 1) were recruited. Of these, 16 NH subjects (S1-16) and 11 CI subjects participated in Experiment 1; 7 NH subjects (S14-20) and 11 CI subjects participated in Experiment 2. Participation was compensated and all subjects gave informed consent. All procedures were performed according to protocols approved by the Shenzhen University's ethical review board.

2.2. Vocoders

Fig.1 shows the vocoder processing procedure we used. Speech sounds were bandpass filtered (6th-order Butterworth in Experiment 1; 4th-order Butterworth in Experiment 2) in m channels covering the frequency range of 80 to 7999 Hz. The filter cutoff frequencies were defined by equally dividing the frequency range according to Greenwood function [31]. The envelope in each channel was computed by full-wave rectification and low-pass filtering at 125 Hz (8th-order Butterworth). A carrier signal was multiplied by the envelope, and the modulated outputs were summed to synthesize the vocoded speech sounds which were presented via a loudspeaker (see below). The carrier signal was either a bandlimited white noise (filtered by the corresponding bandpass filter for each channel) or a sine tone whose frequency is at the center of the corresponding channel and whose initial phase was random.

2.3. Stimuli and procedure

We used sentence material from two published speech databases for Mandarin as it is spoken in mainland China: the Mandarin speech perception (MSP) corpus [32] and the Mandarin hearing in noise test (MHINT) corpus [33]. MSP includes 10 lists, each with 10 sentences, each with 7 monosyllabic words recorded by a female speaker; MHINT includes 12 test lists and 2 training lists, each with 20 sentences, each with 10 monosyllabic words recorded by a male speaker.

In Experiment 1 (Intelligibility), vocoded speech (channel number m = 4, 8, 16, and 22) with sine-wave and noise carriers, as well as original speech, was presented to the CI listeners. Vocoded speech (channel number m = 4, 8, and 16) with sine-wave and noise carriers was presented to the NH listeners. Examples of electrodograms (based on one Cochlear product user's Advanced Combination Encoding strategy) and of spectrograms for 16 and 4 channel vocoded speech are shown in Fig.2. For each condition, a MHINT test list (20 sentences) was used to measure the word recognition scores. All conditions were tested in a randomized order. In each trial, one sentence was played, and could be repeated up to three times. The subject's score for each sentence was the highest number of correctly identified words over the up to three presentations of the sentence. No sentence was used in more than one trial for any one subject.

In Experiment 2 (Quality), a two-alternative forced choice task was used. In each trial, the subjects were instructed to click

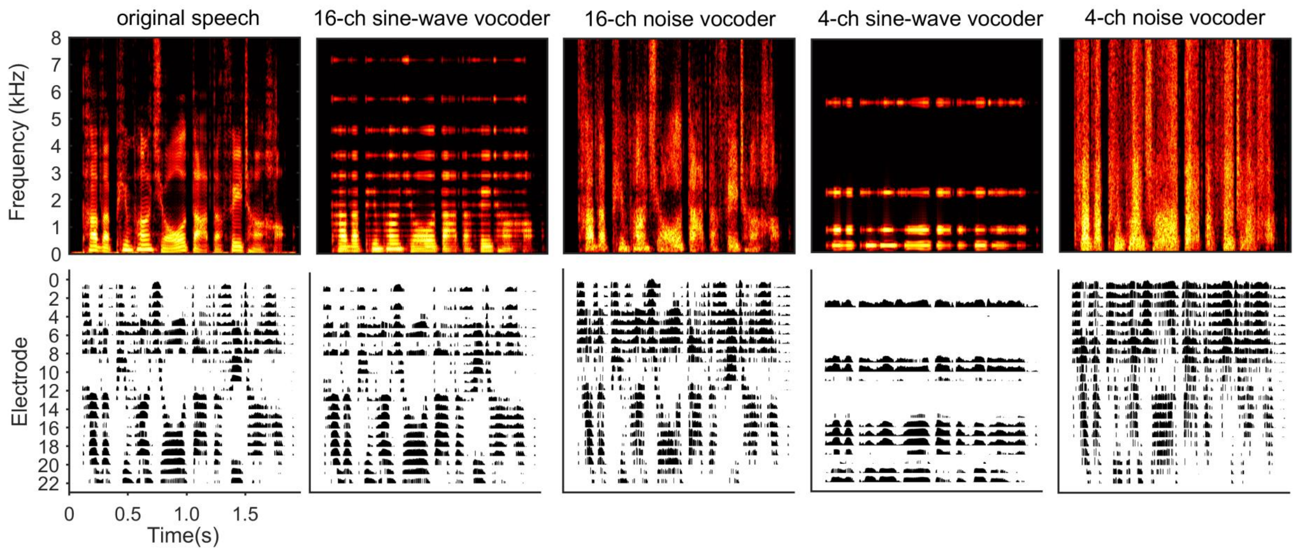

Figure 2: *Examples of spectrograms (top) and electrodograms (bottom) of original and vocoded speech*

on two buttons on a computer screen to play two sentences, one of which was an original sentence, and the other was a vocoded sentence. The subjects then had to judge which sentence sounded more like natural speech. There were 8 comparison pairs (2 carriers $\times$ 4 channel counts) including sine-wave carrier and noise carrier version of original-vs-80ch, original-vs-40ch, original-vs-22ch, original-vs-16ch. To reduce the potential small influence of the bandpass filtering on the temporal domain features of the band signal, the "original" speech was also filtered by the corresponding vocoder's analysis filters and then the outputs were directly recombined to resynthesize the"original" speech stimulus for these experiments. For each condition, 20 sentences were used, yielding 160 sentences for each subject. The sentences were randomly selected from a combined corpus consisting of 380 sentences from MHINT and MSP. For all pair conditions and sentences, the presentation order was randomized. All stimuli were presented at a comfortable level (approximately 70 dB A) through an audio interface (Focusrite Scarlett 2i4) and a loudspeaker (YAMAHA HS5I) placed in front of the subject with about 1 m distance from the head in a sound-proof room.

2.4. Statistical analysis methods

For Exp.1, we performed repeated measures analysis of variance (rm-ANOVA). Because NH subjects did not perform tasks under the conditions of vocoding with 22 channels and the original condition, we conducted ANOVAs in two steps to address the asymmetrical data-set. First, in order to compare the group performance, we selected the data on the conditions with 4, 8, and 16 channels, in which both of the two groups performed the tasks, to perform a 2 (group)$\times$ 3 (channel) $\times$ 2 (vocoder) ANOVA. Group (NH vs. CI) served as a between-subject factor. Channel (4, 8, 16 channels) and vocoder (sine-wave vs. noise) served as within-subject factors. Second, we performed a one-way rm-ANOVA to analyze the NH data. The within-subject factor included nine levels: original, sine-wave and noise with 4, 8, 16, and 22 channels respectively. Greenhouse-Geisser correction was used for nonsphericity, and Bonferroni correction was used to correct for multiple comparisons.

For Exp.2, we compared observed counts of correct identification of the natural speech sample against the null hypothesis that, if a subject were to choose randomly, they would identify the correct sentence on half of the trials on average, and the number of correct choices observed would follow a binomial distribution with N = 20 and p = 1/2. The probability of observing 15 or more correct choices out of 20 by chance is then as small as 0.0386. Consequently, subjects can be said to significantly prefer one condition if their choices follow that condition in at least 15 of the 20 trials (75%).

3. Results

Results from both experiments are illustrated in Fig.3.

Word recognition results from Experiment 1 are shown in Fig.3(A).The 16 NH subjects got very high scores (91% to 100%) for both 16 and 8 channel conditions; their mean scores with 4 channels were much lower ($\sim$ 70%). The CI subjects' scores are much more variable, but most of them (7/11) also achieved high scores ($>$ 87.5%) for original speech and 8-to-22-ch vocoded speech; their mean scores with 4 channels were substantially lower ($\sim$ 49% - 20%). The results of 2 $\times$ 3 $\times$ 2 (group $\times$ channel count $\times$ carrier) repeated measures ANOVA showed that there was a significant main effect of group $[F(1, 25) = 15.22, p < .001, \eta_p^2 = .38]$. NHs showed significantly higher mean accuracy (91.2%) on word recognition than CIs (73.4%). The main effect of channel count was also significant $[F(2, 50) = 178.19, p < .001, \eta_p^2 = .88]$. The mean accuracy of word recognition was significantly lower for stimuli with 4 channels (64.4%) compared to stimuli with 8 (89.7%) or 16 channels (92.8%). There was no significant difference in the accuracy of word recognition between 8 and 16 channels. There was a significant interaction between group and channel $[F(2, 50) = 5.06, p < .05, \eta_p^2 = .17]$. No other effect was significant. The results of one-way repeated measures ANOVA on CIs' data showed that there was a significant main effect of the within-subject factor $[F(8, 80) = 32.98, p < .001, \eta_p^2 = .77]$. The mean accuracy of word recognition was significantly lower for stimuli with 4 channels (tone: 53.1%; noise: 51.8%) compared to stimuli under other conditions (all higher than 80%) among which there was no significant differ-

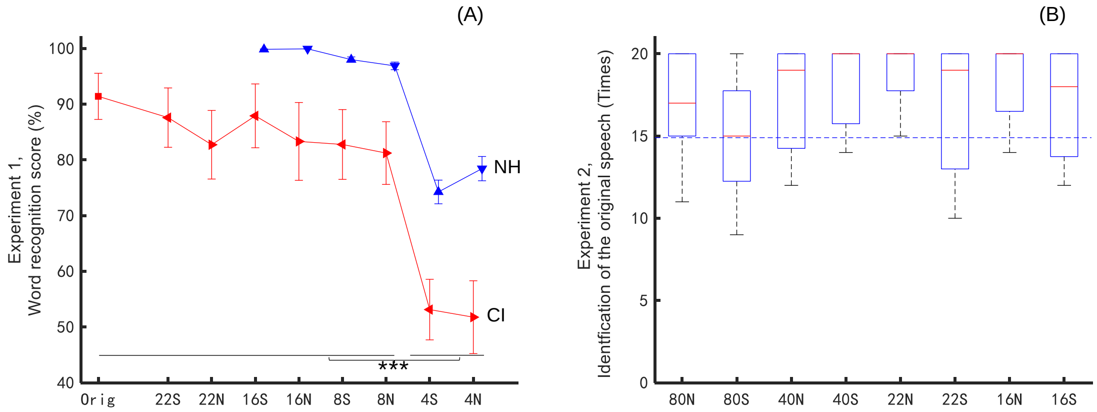

Figure 3: *Illustration of the mean intelligibility results with CI and NH subjects from Experiment 1 (A) and the boxplots of quality results with CI subjects from Experiment 2 (B). In the abscissas, the vocoder conditions are denoted by channel number + carrier type (S: sine-wave; N: noise). In (A), error bars show the* ± 1 *standard error of mean. Significant differences (*$p < 0.001$*) are marked by an asterisk.*

ence in recognition accuracy.

In Experiment 2, all NH subjects consistently identified the original sentences as more natural sounding in almost all (1119/1120) trials, irrespective of channel count and carrier. The identification results for the CI subjects are shown in the boxplots of Fig.3(B). According to the previously defined binomial distribution, under most conditions (except 80ch-sine-wave) most ($\geq 8/11$) CI subjects identified the original speech as more natural significantly more often than expected by chance. In the 80ch-sine-wave condition, 5/11 subjects failed to identify the original speech as more natural often enough to reach statistical significance. The variance among CI subjects was large. Five CI subjects identified the original speech as more natural in almost all trials, with no more than 5/160 mistakes, but other subjects found the task much more difficult.

4. Discussion

The present study tested the perception of CI simulated speech in CI listeners. Two experiments respectively tested speech recognition in quiet and naturalness preference in CI listeners and NH control subjects.

For speech recognition, both groups showed asymptotic performance with 8 channels for both sine-wave and noise vocoders. This is consistent with previous channel number studies of CI simulation in NH subjects [2, 34]. Even though large variances were found for CI subjects, half of the CI listeners showed NH-comparable performance. One perhaps interesting observation is that, with four channel vocoders, the best CI listeners in our cohort obtained much higher scores than the worst NH listeners in this acute test. It may be that CI listeners have so much practice in interpreting sounds based on the temporal envelopes of only a modest number of bandpassed channels that they can outperform naive NH subjects in conditions where only temporal-envelope based speech information is available. As for the carriers, even though noise vocoded and sine-wave vocoded speech show obviously different spectrogram and electrodogram patterns (see Fig.2), no significant difference was found for either subject group in Experiment 1. This observation is consistent with the earliest sine-wave vocoder study in NH subjects [2], and in line with the view that temporal envelope cues can be decoded by the auditory system for understanding speech in a quiet environment irrespective of the type of carrier that they are imposed on.

For naturalness judgments, most subjects easily distinguished the original speech from the vocoded counterpart, even in the 80-channel high resolution vocoded conditions. This indicates that word recognition scores on their own may underestimate the vocoded sound perception abilities in CI listeners, because our results indicate that CI users whose word recognition scores are similarly short of perfect for a set of natural and vocoded speech samples nevertheless typically have little difficulty in distinguishing which samples were natural speech and which were vocoded.

Admittedly there are technical limitations in this initial study. For example, the frequency allocations between the vocoders and CI processors were not matched, and cascading a vocoder and a CI speech processor could lead to additional signal degradation when that mismatch is considerable. However, such potential mismatch issues are not likely to play a big role in the cases where the number of vocoder channels is much larger than the likely number of effective CI channels, as would have been the case for our 40 and 80 channel vocoder signals. Nevertheless, most CI users had little difficulty in distinguishing normal from vocoded speech even at these very high numbers of vocoder channels, underscoring the somewhat "limited realism" of vocoded sounds as a simulation of cochlear implant hearing. Other aspects of vocoded speech perception in CI users which may be worth investigating in future experiments include, speech-in-noise perception, or the ability of CI users to distinguish whether vocoded signals use noise carriers or sine-wave carriers.

5. Acknowledgments

We thank all the subjects who participated in these experiments and Sun Jincangjian and Lin Chuxiang who helping with subject recruitment. This work was supported by the Guangdong Basic and Applied Basic Research Foundation Grant (2022A1515011361) and the Shenzhen Fundamental Research Program (20220809191805001). Corresponding authors: N. Zheng and Q. Meng.

6. References

[1] R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, no. 5234, pp. 303–304, 1995.

[2] M. F. Dorman, P. C. Loizou, and D. Rainey, "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," *J. Acoust. Soc. Am.*., vol. 102, no. 4, pp. 2403–2411, 1997.

[3] F. Kong, Y. Mo, H. Zhou, Q. Meng, and N. Zheng, "Channel-vocoder-centric modelling of cochlear implants: Strengths and limitations," in *Proceedings of the 9th Conference on Sound and Music Technology: Revised Selected Papers from CMST*. Springer, 2022, pp. 137–149.

[4] M. A. Stone, C. Füllgrabe, and B. C. Moore, "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," *J. Acoust. Soc. Am.*., vol. 124, no. 4, pp. 2272–2282, 2008.

[5] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.*, vol. 110, no. 2, pp. 1150–1163, 2001.

[6] A. H. Mehta, H. Lu, and A. J. Oxenham, "The perception of multiple simultaneous pitches as a function of number of spectral channels and spectral spread in a noise-excited envelope vocoder," *Journal of the Association for Research in Otolaryngology*, vol. 21, pp. 61–72, 2020.

[7] M. A. Stone, G. Prendergast, and S. Canavan, "Measuring access to high-modulation-rate envelope speech cues in clinically fitted auditory prostheses," *The Journal of the Acoustical Society of America*, vol. 147, no. 2, pp. 1284–1301, 2020.

[8] P. C. Loizou, M. Dorman, O. Poroy, and T. Spahr, "Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution," *J. Acoust. Soc. Am.*, vol. 108, no. 5, pp. 2377–2387, 2000.

[9] M. F. Dorman, P. C. Loizou, and D. Rainey, "Simulating the effect of cochlear-implant electrode insertion depth on speech understanding," *J. Acoust. Soc. Am.*, vol. 102, no. 5, pp. 2993–2996, 1997.

[10] H. Fathima, J. S. Bhat, and A. N. Pitchaimuthu, "Effect of interaural electrode insertion depth difference and independent band selection on sentence recognition in noise and spatial release from masking in simulated bilateral cochlear implant listening," *European Archives of Oto-Rhino-Laryngology*, pp. 1–9, 2023.

[11] A. Faulkner, S. Rosen, and C. Norman, "The right information may matter more than frequency-place alignment: simulations of frequency-aligned and upward shifting cochlear implant processors for a shallow electrode array insertion," *Ear Hear.*, vol. 27, no. 2, pp. 139–152, 2006.

[12] C. Karoui, C. James, P. Barone, D. Bakhos, M. Marx, and O. Macherey, "Searching for the sound of a cochlear implant: Evaluation of different vocoder parameters by cochlear implant users with single-sided deafness," *Trends in Hearing*, vol. 23, p. 2331216519866029, 2019.

[13] S. N. Garadat, R. Y. Litovsky, G. Yu, and F.-G. Zeng, "Role of binaural hearing in speech intelligibility and spatial release from masking using vocoded speech," *J. Acoust. Soc. Am.*, vol. 126, no. 5, pp. 2522–2535, 2009.

[14] S. A. Ausili, B. Backus, M. J. Agterberg, A. J. van Opstal, and M. M. van Wanrooij, "Sound localization in real-time vocoded cochlear-implant simulations with normal-hearing listeners," *Trends in Hearing*, vol. 23, p. 2331216519847332, 2019.

[15] A. Faulkner, S. Rosen, and C. Smith, "Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants," *J. Acoust. Soc. Am.*, vol. 108, no. 4, pp. 1877–1887, 2000.

[16] X. Luo and L. Hayes, "Vibrotactile stimulation based on the fundamental frequency can improve melodic contour identification of normal-hearing listeners with a 4-channel cochlear implant simulation," *Frontiers in neuroscience*, vol. 13, p. 1145, 2019.

[17] M. Thomas, S. Willis, J. J. Galvin III, and Q.-J. Fu, "Effects of tonotopic matching and spatial cues on segregation of competing speech in simulations of bilateral cochlear implants," *Plos one*, vol. 17, no. 7, p. e0270759, 2022.

[18] E. Gaudrain, N. Grimault, E. W. Healy, and J.-C. Béra, "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," *J. Acoust. Soc. Am.*, vol. 124, no. 5, pp. 3076–3087, 2008.

[19] M. Yuan, T. Lee, K. C. Yuen, S. D. Soli, C. A. van Hasselt, and M. C. Tong, "Cantonese tone recognition with enhanced temporal periodicity cues," *J. Acoust. Soc. Am.*, vol. 126, no. 1, pp. 327–337, 2009.

[20] Q. Meng, N. Zheng, and X. Li, "Loudness contour can influence mandarin tone recognition: Vocoder simulation and cochlear implants," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 641–649, June 2017.

[21] F. Chen, "Contributions of consonant-vowel transitions to mandarin tone identification in simulated electric-acoustic hearing." in *INTERSPEECH*, 2019, pp. 3138–3142.

[22] Q.-J. Fu and J. J. Galvin III, "Recognition of spectrally asynchronous speech by normal-hearing listeners and nucleus-22 cochlear implant users," *J. Acoust. Soc. Am.*, vol. 109, no. 3, pp. 1166–1172, 2001.

[23] S. F. Poissant, N. A. Whitmal III, and R. L. Freyman, "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," *J. Acoust. Soc. Am.*, vol. 119, no. 3, pp. 1606–1615, 2006.

[24] X. Li, K. Nie, N. S. Imennov, J. H. Won, W. R. Drennan, J. T. Rubinstein, and L. E. Atlas, "Improved perception of speech in noise and mandarin tones with acoustic simulations of harmonic coding for cochlear implants," *J. Acoust. Soc. Am.*, vol. 132, no. 5, pp. 3387–3398, 2012.

[25] Q. Meng, N. Zheng, and X. Li, "Mandarin speech-in-noise and tone recognition using vocoder simulations of the temporal limits encoder for cochlear implants," *J. Acoust. Soc. Am.*, vol. 139, no. 1, pp. 301–310, 2016.

[26] J. Schnupp, "What do cochlear implants sound like? - 1: speech," 2023. [Online]. Available: https://auditoryneuroscience.com/prosthetics/noise-vocoded-speech

[27] S. Sheldon, M. K. Pichora-Fuller, and B. A. Schneider, "Priming and sentence context support listening to noise-vocoded speech by younger and older adults," *J. Acoust. Soc. Am.*, vol. 123, no. 1, pp. 489–499, 2008.

[28] R. Newman and M. Chatterjee, "Toddlers' recognition of noise-vocoded speech," *J. Acoust. Soc. Am.*, vol. 133, no. 1, pp. 483–494, 2013.

[29] M. R. Bregman, A. D. Patel, and T. Q. Gentner, "Songbirds use spectral shape, not pitch, for sound pattern recognition," *Proc. Natl. Acad. Sci.*, vol. 113, no. 6, pp. 1666–1671, 2016.

[30] J. H. Won, C. Lorenzi, K. Nie, X. Li, E. M. Jameyson, W. R. Drennan, and J. T. Rubinstein, "The ability of cochlear implant users to use temporal envelope cues recovered from speech frequency modulation," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 1113–1119, 2012.

[31] D. D. Greenwood, "A cochlear frequency-position function for several species29 years later," *J. Acoust. Soc. Am.*, vol. 87, no. 6, pp. 2592–2605, 1990.

[32] Q.-J. Fu, M. Zhu, and X. Wang, "Development and validation of the mandarin speech perception test," *J. Acoust. Soc. Am.*, vol. 129, no. 6, pp. EL267–EL273, 2011.

[33] L. L. Wong, S. D. Soli, S. Liu, N. Han, and M.-W. Huang, "Development of the mandarin hearing in noise test (mhint)," *Ear Hear.*, vol. 28, no. 2, pp. 70S–74S, 2007.

[34] P. C. Loizou, M. Dorman, and Z. Tu, "On the number of channels needed to understand speech," *J. Acoust. Soc. Am.*., vol. 106, no. 4, pp. 2097–2103, 1999.