Unintelligible Low-Frequency Sound Enhances Simulated Cochlear-Implant Speech Recognition in Noise

Janice E. Chang, John Y. Bai, and Fan-Gang Zeng*, *Member, IEEE*

Abstract—**Speech can be recognized by multiple acoustic cues in both frequency and time domains. These acoustic cues are often thought to be redundant. One example is the low-frequency sound component below 300 Hz, which is not even transmitted by the majority of communication devices including telephones. Here, we showed that this low-frequency sound component, although unintelligible when presented alone, could improve the functional signal-to-noise ratio (SNR) by 10–15 dB for speech recognition in noise when presented in combination with a cochlear-implant simulation. A similar low-frequency enhancement effect could be obtained by presenting the low-frequency sound component to one ear and the cochlear-implant simulation to the other ear. However, a high-frequency sound could not produce a similar speech enhancement in noise. We argue that this low-frequency enhancement effect cannot be due to linear addition of intelligibility between low- and high-frequency components or an increase in the physical SNR. We suggest a brain-based mechanism that uses the voice pitch cue in the low-frequency sound to first segregate the target voice from the competing voice and then to group appropriate temporal envelope cues in the target voice for robust speech recognition under realistic listening situations.**

Index Terms—**Cochlear implant, electro-acoustic stimulation, low-frequency sound, speech recognition, voice pitch.**

I. INTRODUCTION

ALTHOUGH human voice pitch is usually below 300 Hz, essentially all communication systems, including telephones, do not carry any information below this frequency [1]. Speech perception is not affected in normal-hearing listeners because these low-frequency sounds alone provide negligible intelligibility [2] and voice pitch can be recovered from high-frequency harmonics, a phenomenon known as the "missing fundamental" [3]. Unfortunately, neither the low-frequency sounds nor the harmonics are accessible to the majority of the 100 000 cochlear-implant users who rely primarily on the temporal waveform envelope cue for speech recognition [4]–[6]. As a result, they enjoy a high level of speech recognition in quiet listening environments including telephone use, but have difficulty in noisy situations [7], [8].

Recent studies have shown low-frequency sounds to enhance cochlear implant performance, particularly for speech recognition in noise [9]–[13]. However, the utility and mechanisms of this low-frequency sound enhancement effect remain unclear. One problem in these previous studies was that the degree and range of hearing loss at low frequencies varied greatly in a typical clinical population, making it difficult to infer the sources of improvement. The other problem was the difficulty in establishing well-controlled conditions under which the enhancement effect could be observed. For example, does the observed benefit require physical interactions between acoustic and electric stimulations in the inner ear [14]? Can the benefit be simply the sum of intelligibilities between low- and high-frequency channels [2]? Will the high-frequency acoustic sound component induce a similar benefit?

Using a cochlear-implant simulation, we conducted two experiments to estimate the size of the low-frequency sound enhancement effect and to probe the underlying mechanisms. To overcome the variability in the degree and range of hearing loss at low frequencies, we low-passed the signal at 250, 500, and 1000 Hz in normal-hearing subjects to separate the contributions of fundamental frequency and formants to the observed benefit (Exp. 1). To probe the underlying mechanisms, we paired the acoustic sound and the cochlear-implant simulation in several different ways to establish the necessity and sufficiency of the low-frequency sound component in the observed benefit (Exp. 2).

II. METHODS

A. Subjects

A total of 16 young adults with normal hearing participated in this study. Eight of them participated in Exp. 1 and the remaining 8 participated in Exp. 2.

B. Stimuli

Hearing-in-noise-test (HINT) sentences produced by a male talker were used as the test materials [15]. The competing voice was a single sentence: "A large size in stockings is hard to sell," produced by a female talker. The competing sentence was longer than all of the HINT sentences. The male talker's voice pitch varied between 90 and 130 Hz, while the female talker's voice pitch varied between 250 and 450 Hz. The target sentences were fixed at 65 dBA, while the competing voice level was varied to produce signal-to-noise ratios (SNRs) from

---

Manuscript received May 3, 2006; revised July 5, 2006. This work was supported in part by the National Institutes of Health (NIH) under Grant 2RO1DC002267. *Asterisk indicates corresponding author.*

J. E. Chang was with the Bioengineering Dept., University of California, Berkeley, CA 94720 USA. She is now with the Hearing and Speech Research Laboratory, University of California, Irvine, CA 92697 USA.

J. Y. Bai is with the Hearing and Speech Research Laboratory, University of California, Irvine, CA 92697 USA

*F.-G. Zeng is with the Hearing and Speech Research Laboratory, University of California, Irvine, CA 92697 USA (e-mail: fgzeng@uci.edu).

Digital Object Identifier 10.1109/TBME.2006.883793

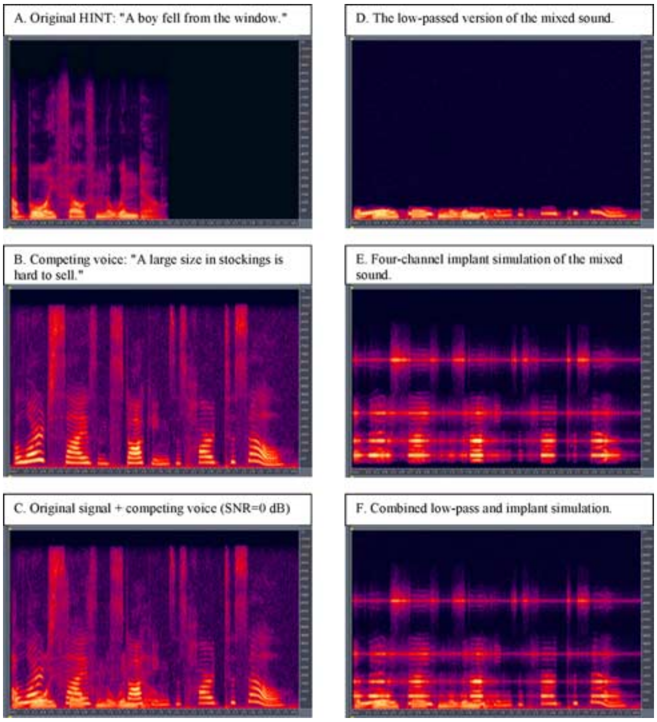

Fig. 1. Spectrograms of an original target sentence [(A) "A boy fell from the window,"] the original competing voice [(B) "A large size in stockings is hard to sell,"], the mixed sentence at 0-dB SNR (panel C), the $<$500-Hz, low-frequency version of the mixed sentence (panel D), the 4-channel cochlear-implant simulation of the mixed sentence (panel E), and the combined low-frequency sound and the 4-channel implant simulation (panel F). Audio demonstrations are available online (http://www.ucihs.uci.edu/hesp). Y-axis is frequency from 0 to 11 000 Hz, X-axis is time from 0 to 3.45 s, and color scale represents intensity.

-20 to $+20$ dB. Stimuli were presented to the subjects through Sennheiser headphones in a double-walled, sound-proof room.

We produced a 4-channel cochlear-implant simulation [4], [16] by band-passing the original signal through four bandpass filters with their cutoff frequencies designed to have equal distances along the cochlea [17]. The temporal envelope from each band was extracted by full-wave rectification, followed by a 400-Hz low-pass filter (second-order Butterworth). The envelope was then used to amplitude modulate a sinusoidal carrier with its frequency being the same as the band's center frequency. Summing the modulated sub-band signals formed a 4-channel cochlear implant simulation. The simulation captures the speech processing part in a cochlear implant but unlikely the electrode-nerve interaction and central processing deficits in an actual cochlear implant listener [4], [16].

Exp. 1 generated three low-frequency stimuli by low-passing the original signal at 250, 500, and 1,000 Hz (a 15th-order Butterworth filter). The 250-Hz low-pass stimulus contained essentially fundamental frequency information from the target male talker, whereas the 500- and 1000-Hz stimuli contained fundamental frequency and lower formant information from both the target and the masker. The corresponding cochlear-implant simulation had a contiguous, but non-overlapping, bandwidth (i.e., from 250, 500, or 1000 Hz to 8800 Hz, respectively). Fig. 1 shows spectrograms of sample stimuli used in the present study.

Exp. 2 used the 500-Hz low-passed acoustic component and the 500–8,800 Hz, 4-channel, cochlear-implant simulation as the baseline condition. In addition, we produced 4 experimental conditions, including a "5-Channel" implant simulation, diotic implant mode, dichotic mode, and combining a 4-channel simulation with a high-frequency sound component. The "5-channel" simulation was produced by extracting the envelope from the 500-Hz low-passed sound and then amplitude-modulating a fixed sinusoidal carrier at 290 Hz (the center of the 80–500 Hz band). This "5-channel" condition tested whether similar performance can be achieved by the temporal envelope cue from the low-frequency sound or the temporal fine structure in the original low-frequency is required. The "diotic implant" condition presented identical 4-channel simulations to both ears, testing whether the observed low-frequency effect is related to binaural hearing. The "dichotic mode" presented the 4-channel simulation to one ear and the low-frequency sound to the other, testing whether the observed effect requires direct interactions in the same ear between low- and high-frequency sounds. Finally, the high-frequency sound condition paired a 4-channel simulation with a 4000-Hz high-passed acoustic component (a fourth-order Elliptic filter) to test whether the high-frequency sound can produce a similar beneficial effect.

C. *Procedure*

A 1-up, 1-down adaptive procedure was used to estimate the SNR at which 50% sentences were correctly recognized [18]. This SNR is referred to SNR50% hereinafter. To estimate the SNR50% value, the initial SNR was set at -8 dB. The SNR was increased by 2 dB until the subject correctly identified all keywords in the first sentence. After that, the SNR was increased by 1 dB if the sentence was incorrectly identified, or decreased by 1 dB if the sentence was correctly identified (i.e., all keywords must be correct). The SNR50% was estimated by averaging the SNRs, at which the subject went from correct response to incorrect response or vice versa, using the final 1-dB step. A two-way ANOVA with a within-subjects design was used to analyze the obtain data in Exp. 1. A Student's t-test with two-tailed distribution and two-sample unequal variance was used to analyze the obtained data in Exp. 2.

III. RESULTS

A. *Exp. 1. Effect of Low-Cutoff Frequencies*

With the 250-, 500-, and 1000-Hz low-frequency sound alone, the subjects on average correctly identified 5, 51, and 75% keywords, corresponding to a sentence recognition score of 0%, 10%, and 11%, respectively. Fig. 2 shows the SNR50% value for the 4-channel cochlear-implant simulations in the presence and absence of the low-frequency sound. All 4-channel simulations produced a 10-dB SNR50% $[\mathrm{F}(2,18) = 0.02,\ \mathrm{p} = 0.98]$, whereas the additional low-frequency sound produced an SNR50% of 0, -1, and -5 dB for low-frequency components below 250, 500, and 1000 Hz, respectively $[\mathrm{F}(2,18) = 5.13,\ \mathrm{p} = 0.017]$. The corresponding 10-, 11-, and 15-dB improvement in SNR50% was significant $[\mathrm{F}(1,9) = 98.79,\ \mathrm{p} < 0.001]$, indicating that the presence of

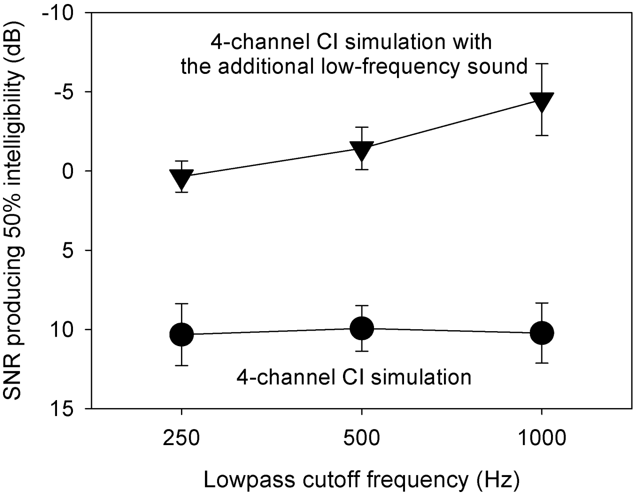

Fig. 2. SNR that produced 50% speech intelligibility (SNR50%) is plotted as a function of the low-pass cutoff frequency. Circles represent data from the 4-channel cochlear-implant simulation and the triangles represent the data from combining the 4-channel simulation with the low-frequency sound component. Error bars represent s.e.m.

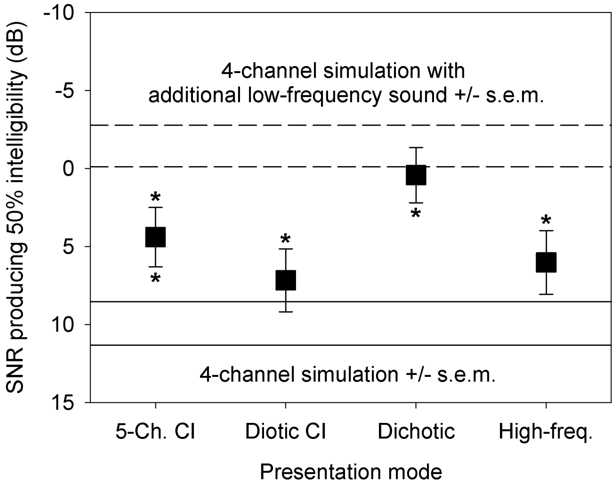

Fig. 3. SNR50% with a 5-channel cochlear-implant simulation ("5-ch. CI"), diotic presentation of two 4-channel simulations ("Diotic CI"), dichotic presentation of the 4-channel simulation to one ear and the low-frequency sound to the other ("Dichotic"), and a combination of the 4-channel simulation and a high-frequency sound with 4000-Hz high-pass cutoff frequency ("High-freq."). The two solid lines represent the 4-channel simulation performance and the two dashed lines represent the performance with the additional low-frequency sound (mean ± s.e.m.) from Exp. 1. The asterisk represents a statistically significant difference ($p < 0.05$).

the relatively unintelligible low-frequency sound enhanced the cochlear-implant speech recognition in noise.

B. *Exp. 2. Effect of Stimulation Modes*

Fig. 3 re-plots the performance of the 4-channel cochlear-implant simulation (represented by the two solid lines on the bottom) and of the 4-channel simulation with the additional low-frequency sound (represented by the two dashed lines on the top). Squares represent the performance from the 4 experimental conditions. First, the 5-channel cochlear-implant simulation produced a 4-dB SNR50%, which was significantly better than the 4-channel simulation but worse than the 4-channel simulation plus the low-frequency sound. This result suggests that the envelope cue from the low-frequency sound cannot entirely account for the observed low-frequency enhancement effect (~4 out of 11 dB). Second, diotic presentation of the 4-channel cochlear-implant simulation produced similarly poor performance (8 dB SNR50%) as the monaural presentation, suggesting disassociation between the low-frequency sound and binaural hearing effects. Third, the dichotic condition presenting the low-frequency sound in the opposite ear to the 4-channel simulation produced the same performance (1 dB SNR50%) as presenting both to the same ear, suggesting independence on physical interactions between low-frequency and high-frequency sounds in the same ear. Finally, the high-frequency sound alone produced 18% intelligibility in keyword recognition, but it did not significantly improve the performance (7 dB SNR50%), suggesting that the low-frequency sound is necessary for the observed enhancement effect.

IV. DISCUSSION AND CONCLUSION

The overwhelming majority of previous speech perception studies used stimuli consisting of a single, contiguous frequency band. The few exceptions that used a low-frequency band and a non-adjacent high-frequency band all found a surprising enhancement effect of the low-frequency sound on consonant recognition, even when the low-frequency band was as low as 55–110 Hz [19]–[23]. Because these studies typically tested phoneme recognition in quiet, they did not receive as much attention as they deserved. This letter shows a much more significant effect of the low-frequency sound on sentence recognition in noise than consonant recognition in quiet. We consider the following possibilities to account for the observed low-frequency sound enhancement effect.

First, the low intelligibility scores associated with the low-frequency sound suggest that the 10-dB difference in functional SNR cannot be due to a linear addition of intelligibilities between the low- and high-frequency sounds. If the addition of intelligibility hypothesis holds, then we would predict that adding the high-pass acoustic sound, which had 18% keyword intelligibility compared with only 5% for the 250-Hz low-pass sound, should similarly enhance the cochlear-implant simulation performance. Second, the lack of overlap in bandwidth between the low-frequency sound and the high-frequency cochlear-implant simulation suggests that the improvement cannot be due to physically increased signal to noise ratios at any frequencies. Finally, the similar enhancement effect between the dichotically-combined and the monaurally-combined low-frequency and high-frequency sounds suggest a brain-based mechanism, rather than an ear-based mechanism.

We hypothesize that the presence of low-frequency sounds helps to segregate information-bearing high-frequency sounds into different streams, improving performance at difficult SNRs. Better performance with higher low-pass cutoff frequencies likely reflects increased contribution from speech formants, particularly the first formant. Because the 250-Hz low-pass condition contained essentially only information regarding the target talker's voice pitch, we further hypothesize that fundamental frequency plays a major role in the observed low-frequency sound effect. We also note that this centrally-based fundamental frequency mechanism is consistent with similar observations in multi-modal integration studies, showing that unintelligible voice pitch can enhance lipreading [24] and speech recognition via tactile stimulation [25]. Physiologically, a low-frequency sound (e.g., 100 Hz tone presented at a comfortable loudness level) produces much broader cochlear and cortical excitation patterns than a high-frequency sound (e.g., a 10 000-Hz tone presented at the same level) [26], [27]. It is possible that the brain uses the low-frequency cue to help analyze the complex auditory scene [28], [29].

Practically, the present result challenges the wisdom of bilateral cochlear implantation in persons with residual acoustic low-frequency hearing because better performance may be achieved with combined cochlear implant and hearing aid stimulation than bilateral cochlear implants [12], [13], [30]. The present result likely represents the best condition in actual cochlear implant users with residual acoustic hearing. Actual performance of the combined acoustic and electric stimulation still needs to be customized and optimized. Because most modern communication devices, including telephone and auditory prostheses, do not transmit low-frequency sounds below 300 Hz, the present result suggests that explicit encoding of the low-frequency sounds may enhance their performance in noisy environments.

ACKNOWLEDGMENT

The authors would like to thank M. Marsala for data collection and A. Copeland, T. Lu, and E. Rubel for their comments.

REFERENCES

[1] H. Fletcher, *Speech and Hearing in Communication*, 2nd ed. New York: Van Nostrand, 1953.

[2] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.*, vol. 19, pp. 90–119, 1947.

[3] H. Fletcher, "Newer concepts of the pitch, the loudness, and the timbre of musical tones," *J. Franklin Inst.*, vol. 220, pp. 405–429, 1935.

[4] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, vol. 270, pp. 303–304, 1995.

[5] F. G. Zeng, "Trends in cochlear implants," *Trends Amplif.*, vol. 8, pp. 1–34, 2004.

[6] B. S. Wilson, C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz, "Better speech recognition with cochlear implants," *Nature*, vol. 352, pp. 236–238, 1991.

[7] K. Nie, G. Stickney, and F. G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 1, pp. 64–73, Jan. 2005.

[8] 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, pp. 1150–1163, 2001.

[9] C. W. Turner, B. J. Gantz, C. Vidal, A. Behrens, and B. A. Henry, "Speech recognition in noise for cochlear implant listeners: benefits of residual acoustic hearing," *J. Acoust. Soc. Am.*, vol. 115, pp. 1729–1735, 2004.

[10] C. von Ilberg, J. Kiefer, J. Tillein, T. Pfenningdorff, R. Hartmann, E. Sturzebecher, and R. Klinke, "Electric-acoustic stimulation of the auditory system. New technology for severe hearing loss," *J. Otorhinolaryngol. Related Spec.*, vol. 61, pp. 334–340, 1999.

[11] M. F. Dorman, A. J. Spahr, P. C. Loizou, C. J. Dana, and J. S. Schmidt, "Acoustic simulations of combined electric and acoustic hearing (EAS)," *Ear Hear.*, vol. 26, pp. 371–380, 2005.

[12] T. Y. Ching, P. Incerti, and M. Hill, "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," *Ear Hear.*, vol. 25, pp. 9–21, 2004.

[13] Y. Y. Kong, G. S. Stickney, and F. G. Zeng, "Speech and melody recognition in binaurally combined acoustic and electric hearing," *J. Acoust. Soc. Am.*, vol. 117, pp. 1351–1361, 2005.

[14] K. V. Nourski, P. J. Abbas, C. A. Miller, B. K. Robinson, and F. C. Jeng, "Effects of acoustic noise on the auditory nerve compound action potentials evoked by electric pulse trains," *Hear Res.*, vol. 202, pp. 141–153, 2005.

[15] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95, pp. 1085–1099, 1994.

[16] 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, pp. 2403–2411, 1997.

[17] D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.*, vol. 87, pp. 2592–2605, 1990.

[18] H. Levitt, "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.*, vol. 49, pp. 467–477, 1971.

[19] B. Franklin, "The effect on consonant discrimination of combining a low-frequency passband in one ear and a high-frequency band passband in the other ear," *J. Auditory Res.*, vol. 9, pp. 365–378, 1969.

[20] ——, "The effect of combining low- and high-frequency passbands on consonant recognition in the hearing impaired," *J. Speech Hear. Res.*, vol. 18, pp. 719–727, 1975.

[21] ——, "A comparison of the effect on consonant discrimination of combining low- and high-frequency passbands in normal, congenital, and adventitious hearing-impaired subjects," *J. Am. Aud. Soc.*, vol. 5, pp. 168–176, 1979.

[22] R. P. Lippmann, "Accurate consonant perception without mid-frequency speech energy," *IEEE Trans. Speech Audio Process.*, vol. 4, no. 1, pp. 66–69, Jan. 1996.

[23] R. D. Rosenthal, J. K. Lang, and H. Levitt, "Speech reception with low-frequency speech energy," *J. Acoust. Soc. Am.*, vol. 57, pp. 949–955, 1975.

[24] S. M. Rosen, A. J. Fourcin, and B. C. Moore, "Voice pitch as an aid to lipreading," *Nature*, vol. 291, pp. 150–152, 1981.

[25] R. Drullman and A. W. Bronkhorst, "Speech perception and talker segregation: effects of level, pitch, and tactile support with multiple simultaneous talkers," *J. Acoust. Soc. Am.*, vol. 116, pp. 3090–3098, 2004.

[26] J. G. Arenberg, S. Furukawa, and J. C. Middlebrooks, "Auditory cortical images of tones and noise bands," *J. Assoc. Res. Otolaryngol.*, vol. 1, pp. 183–194, 2000.

[27] M. A. Ruggero, "Responses to sound of the basilar membrane of the mammalian cochlea," *Curr. Opin. Neurobiol.*, vol. 2, pp. 449–456, 1992.

[28] A. L. Giraud, C. Lorenzi, J. Ashburner, J. Wable, I. Johnsrude, R. Frackowiak, and A. Kleinschmidt, "Representation of the temporal envelope of sounds in the human brain," *J. Neurophysiol.*, vol. 84, pp. 1588–1598, 2000.

[29] T. D. Griffiths, J. D. Warren, S. K. Scott, I. Nelken, and A. J. King, "Cortical processing of complex sound: a way forward?," *Trends Neurosci.*, vol. 27, pp. 181–185, 2004.

[30] R. J. van Hoesel and R. S. Tyler, "Speech perception, localization, and lateralization with bilateral cochlear implants," *J. Acoust. Soc. Am.*, vol. 113, pp. 1617–1630, 2003.