PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Reverberant speech recognition with actual cochlear implants: verifying a pulsatile vocoder simulation method

Fanhui KONG[1]; Xianren WANG[2]; Xiangbin TENG[3]; Nengheng ZHENG[1]; Guangzheng YU[4]; Qinglin MENG[4]

[1] The Guangdong Key Laboratory of Intelligent Information Processing, College of Information Engineering, Shenzhen University, Shenzhen, China

[2] Department of Otorhinolaryngology, the First Affiliated Hospital, Sun Yat-Sen University and Institute of Otorhinolaryngology, Sun Yat-Sen University, Guangzhou, China

[3] Department of Neuroscience, Max-Planck-Institute for Empirical Aesthetics, Frankfurt, Germany

[4]Acoustics Lab of School of Physics and Optoelectronics and State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou, China

ABSTRACT

Classic channel vocoders have been widely used for simulating cochlear implants (CIs). However, some important features of state-of-the-art CI strategies were usually ignored, such as pulsatile current, broad current spread, and low quantization resolution of intensity. To solve this problem, a novel vocoder was proposed, which is combined from Gaussian-shaped acoustic pulses each corresponding to an electric pulse and can directly transformed CI-electrodograms to sounds. Compared with the classic channel vocoders on a reverberant speech recognition task, this new vocoder was considerably more sensitive to the reverberant condition (published in APSIPA-2018). Moreover, its performance is comparable to actual CI users in previous CI-reverberation studies. Here, the same paradigm as in the APSIPA-2018 paper was tested on actual CI users. We found that under reverberant conditions the CI users got comparable scores to the simulated performance from the new vocoder simulation in the NH cohort, while the classic vocoder predicted significantly much higher performance This study demonstrated that the classic vocoders overestimated the task performance and further lent great support to the usage of the novel vocoder for the simulation of reverberant speech recognition in adult post-lingually deafened CI users.

Keywords: Cochlear implant, Reverberation, Vocoder

1. INTRODUCTION

1.1 Standard Cochlear Implant Strategy: Continuous Interleaved Sampling

Modern cochlear implants (CIs) can help deaf people to regain hearing by delivering electric pulses sequentially through the implanted electrodes. The electric signal is converted by a CI speech processor from the incoming sound. Even nature is generally analog at least within the psychophysical range of human feeling, we can easily find that in the very early CI system design studies biphasic pulses were considered as one default choice of electric signal waveforms (1). Historically, both analog and pulsatile waveforms were used in 1980s and 1990s (2, 3), but the analog strategies faded out then. Currently, the strategy framework of most CI products is the continuous interleaved sampling (CIS) strategy (4). Its idea is to present “brief pulses to each electrode in a nonoverlapping sequence”, which surprisingly defeated the performance of a contemporary analog strategy of compressed analogue (CA) strategy which “presented analogue waveforms simultaneously to all electrodes”. What’s more, in CIS the brief pulses were only modulated by the temporal envelope from each band and the temporal fine structure, which was then

[4] mengqinglin@scut.edu.cn

well-acknowledged to be important for perception of speech-in-noise, pitch, and sound localization were mostly discarded. One advantage of non-simultaneous pulsatile stimulation is reducing current interaction between electrodes. This is an example of "Less is more". The sparser pattern carried by the pulses indeed dominates the CI products now. For more posterior explanations about the success of CIS, please refer to the article (5) published in 2015 from Blake S. Wilson who was awarded the 2013 Lasker DeBakey Clinical Medical Research Award mainly because of the CIS strategy development.

1.2 Vocoder Simulation: Temporal Envelope × Noise or Sine-wave Carrier

In the academic field of CIs, there is an important branch of studies which use speech synthesis method to assemble some key intermediate signals from the CI processing procedure into a sound and then play synthesized sound to normal hearing (NH) subjects in order to simulate some aspects of CI processing (6). The speech synthesis algorithms are called vocoder, a word came from the voice encoder in machine speech synthesizer or telecommunication. The most widely used CI simulation vocoders are noise and sine-wave carrier channel vocoders, which use temporal envelopes from each band to modulate the continuous noise or sine-wave carrier, since Shannon et al. (7). However, CIS's pulsatile feature, which is critical for actual CIs, does not exist in the simulated ones at all. This may be one of the reasons for some observation, like in (8), that behavioral results with the classical vocoders overestimated actual CI users' performance.

1.3 New Pulsatile Vocoders: Mapping Electric Pulses to Acoustic Pulses

Therefore, a pulsatile vocoder was proposed recently in Meng et al. (9) and Lin and Meng (10). The idea was pretty straightforward that each electric pulse $e(t)$ within a CI electrodogram (i.e., a Time × Electrode × Intensity graph) was replaced by an acoustic pulse $a(t)$ which should be carefully controlled on spectral-temporal domain. The acoustic pulses can replicate the timing feature of electric pulse, but it should be emphasized that the electric pulse in CI and the acoustic pulse in the vocoder has different physical properties. According to the time and frequency scaling property of Fourier transform, we know if $A(j\omega)$ is the Fourier transform of the acoustic pulse $a(t)$, then a time scaled acoustic pulse $a(\lambda t)$ has a Fourier transform of $\frac{1}{|\lambda|}X\left(\frac{j\omega}{\lambda}\right)$. This means there is an inverse relationship between the time and frequency domains, which is also known as the time-frequency uncertainty principle. However, for all electrodes of a CI usually use the same electric pulse is used, i.e., $e(t)$'s standard waveform is a biphasic pulse with tens-of-microseconds duration, microseconds of inter-phasic gap, and envelope-controlled current intensity. Corresponding to each electric pulse, the auditory spectral range, which may be determined by neurons being stimulated by the pulse, is influenced by many factors including electrode place, physiological status of the auditory nerves, conditions of tissues between the electrode and the neurons, and so on.

However, the acoustic models of vocoder simulation have never been expected to replicate all features of CIs. Most of the studies using the noise and sine-wave carrier vocoders just wanted to predict a trend of CI behavioral data versus some strategy parameters, e.g., channel number, envelope cutoff frequency, and current spread.

The pulsatile vocoder idea can be found in Lu et al., (11) (an abstract in 2007 Conference on Implantable Auditory Protheses), where Gaussian-enveloped tones (GET) was proposed to control place, rate, and electric field spread of stimulation. However, several subsequent publications in formal journals based on the GET vocoder only reported some basic psychophysical experiments. No related speech recognition result has been reported. Meng et al. (9) carried out a preliminary study in this direction. Reverberant speech recognition with a Gaussian-enveloped-pulse vocoder was tested in NH subjects. Compared with previous CI data, the results with the new vocoder showed better simulation of actual CIs than the classic vocoders. In this paper, we add a CI experiment using the almost the same paradigm as Meng et al. (9) and provide supportive evidence.

2. METHODS

2.1 Brief Summary of Previous Experiment Methods

Binaural room impulse responses (BRIRs) were recorded in two classrooms (No.107 and 212). Their reverberation times (T30) were 1.5 and 0.5 s respectively. The signal flowchart of the experiment in Meng et al., (9) is shown in Fig.1(A). Convolution output of Mandarin sentence

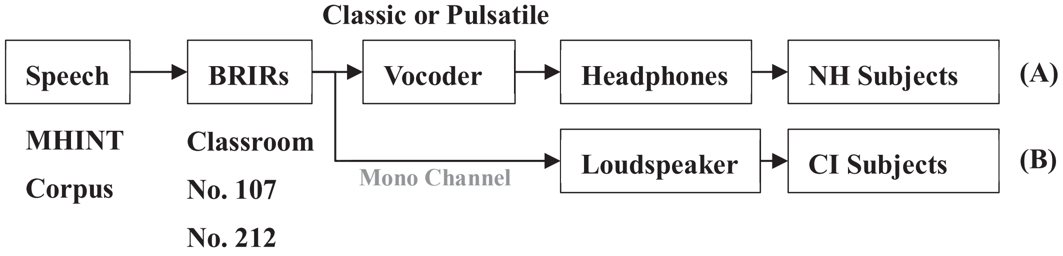

Figure 1 – Signal flowchart of the experiments. (A) Meng et al. (9) (B) Current Paper

signals from MHINT corpus and the BRIRs were further processed by a vocoder of a classic noise-carrier vocoder or a new pulsatile noise-carrier vocoder. Then, the vocoded stimuli were presented through headphones to NH subjects. The classic vocoder used temporal envelopes from 22 or 16 bands to modulate bandlimited noises. The pulsatile vocoder converted electrodograms from a 22 or 16 channel Advanced Combination Encoder strategy (ACE; a default strategy for Cochlear company's product) into sound stimuli directly. The spectrograms of the vocoded reverberant speech were shown in Fig.3 of (9).

2.2 Methods for Current Experiment

Convolution output of Mandarin Sentences from MHINT and RIRs at the left ear were presented to CI subjects through a high-quality soundcard (TASCAM US-366) and a high-quality loudspeaker (Yamaha HS5I) in a sound-proof room. The loudspeaker was located about 0.5 m in-front-of the subject's implanted-side ear. The sound pressure level is about 70 dB A. The experiment paradigm for measuring speech intelligibility was the same as previous study (9).

Four CI subjects participated in this experiment and their information is shown in Table 1. Participation was compensated and all subjects gave informed consent in accordance with the Shenzhen University's review board.

Table 1 – CI user demographic information, hearing history, and device information

Subject	Gender	Age(yr)	CI Experience (yr)	CI Processor	Etiology
C20	M	10	8	Right: Cochlear Freedom	Congenital
C21	F	34	7	Right: Cochlear CP900	Drug-induced
C28	F	40	11	Right: Cochlear N6	Ototoxicity
C30	F	23	1	Right: Cochlear Freedom	Unknown

3. RESULTS AND DISCUSSION

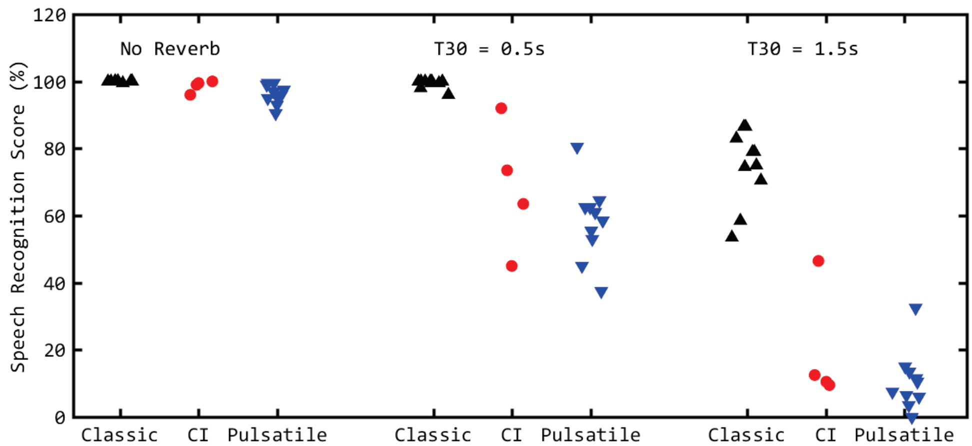

Figure 2 – Speech Recognition Scores with 4 CI listeners and classic and pulsatile vocoders 10 NH listeners under three reverberation conditions (i.e., no reverberation, T30 = 0.5 s, and T30 = 1.5 s)

The speech recognition scores of CI results and the NH results with 22 channel vocoders are shown in Fig.2. For original speech, i.e., under no reverberation condition, all three groups show very high scores approaching 100 %. For reverberation conditions, vocoder simulation with the classic vocoders overestimated the CI performance. The new pulsatile vocoder simulation showed comparable results with the CI subjects. Both CI cohort and pulsatile vocoder cohort were very sensitive to the reverberation. Stronger reverberation derived worse scores for them. Especially for the RIRs in Classroom 212 with T30=1.5, most CI listeners (3/4) and most NH listeners with the pulsatile vocoder (9/10) got scores lower than 20%, while most NH listeners with the classic vocoder (8/10) still got scores higher than 60%.

Actual CI listeners were tested in this study to verify a previous study on a new pulsatile vocoder (9, 10). The CI results combined with the vocoder simulation results from that study showed supportive evidences for the new pulsatile vocoder.

The pulsatile vocoder here uses Gaussian-shaped pulses. We know that Gaussian function's Fourier transform is also a Gaussian function. The time duration and frequency bandwidth have an inverse relationship. In the future, more quantitative analysis (on vocoded stimuli) and perceptual experiments (comparing CI and NH with various parameter settings) should be carried out to fully explore the potentials of the pulsatile vocoder on simulating the electric hearing.

4. CONCLUSION

Pulsatile vocoders directly convert CI electrodograms into sounds which are consisting of acoustic pulses with precise timing information and controllable frequency range to some extent. Reverberant speech perception with actual CI listeners were overestimated by classic vocoders and showed comparable performance as simulation results using the recently proposed pulsatile vocoder. This supportive evidence encourages us to further explore its potentials in the future.

ACKNOWLEDGEMENT

We thank all the subjects who participated in these experiments. This work is jointly supported by NSF of China [Grant No. 11704129 and 61771320], the Fundamental Research Funds for the Central Universities (SCUT), State Key Laboratory of Subtropical Building Science [SCUT, Grant No. 2018ZB23], and Shenzhen Science and Innovation Funds [Grant No. JCYJ 20170302145906843]. F. Kong and X. Wang made equal contribution to this research.

REFERENCES

1. Clark GM, Tong YC, Black R, Forster IC, Patrick JF, Dewhurst DJ. A multiple electrode cochlear implant. The Journal of Laryngology & Otology. 1977; 91(11):935-45.
2. Loizou P. Introduction to cochlear implants. IEEE Engineering in Medicine and Biology Magazine. 1999;18(1):32-42.
3. Moore JA, Teagle HFB. An Introduction to Cochlear Implant Technology, Activation, and Programming. Language, Speech, and Hearing Services in Schools. 2002;33(3):153-61.
4. Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. Nature. 1991; 352(6332):236-8.
5. Wilson BS. Getting a decent (but sparse) signal to the brain for users of cochlear implants. Hearing Research. 2015; 322:24-38.
6. Goupell MJ. Pushing the envelope of auditory research with cochlear implants. Acoustics Today. 2015; 11:26-33.
7. Shannon RV, Zeng F, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270(5234):303-4.
8. Svirsky MA, Ding N, Sagi E, Tan C, Fitzgerald M, Glassman EK, et al. Validation of acoustic models of auditory neural prostheses. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013. May 26-31: p. 8623-8633.
9. Meng Q, Yu G, Wan Y, Kong F, Wang X, Zheng N. Effects of Vocoder Processing on Speech Perception in Reverberant Classrooms. 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC); 2018. Nov. 12-15: 761-765.
10. Lin C, Meng Q. A cochlear implant acoustic model directly mapping electric pulses to acoustic pulses. 2018 Annual Conference of Acoustical Society of China; 2018; Beijing. 2018 Nov. 10-11.
11. Lu T, Carroll J, Zeng FG. On acoustic simulations of cochlear implants. Conference on Implantable Auditory Prostheses (abstract); Lake Tahoe, CA. July 15-20, 2007.