Chapter 0

Acoustic simulations of cochlear implants in human and machine hearing research

Cong-Thanh Do

Idiap Research Institute, CH-1920 Martigny

Switzerland

In this chapter, we review conventional uses of acoustic simulations in cochlear implants research and introduce novel applications of acoustic simulations of cochlear implants, both in cochlear implants research and in other domains, such as automatic speech recognition (ASR) research. In this respect, the first part of the chapter is dedicated in reviewing the conventional applications of acoustic simulations in cochlear implants research. Actually, acoustic simulations are used in two typical applications in the research of cochlear implants. In the first one, acoustic simulations of cochlear implants are used to define how many independent channels are needed in order to achieve high levels of speech understanding? (Loizou et al., 1999). In fact, it is difficult to answer this question basing on the results of subjective tests performed by cochlear implant users, since their performance might be affected by many cofounding factors (e.g. number of surviving ganglion cells). For example, if a cochlear implant user obtains poor auditory performance using 4 channels of stimulation, the rationale might be that 4 stimulating channels are not enough but it could probably be blamed for the lack of surviving ganglion cells near the stimulating electrodes (Loizou et al., 1999). Using acoustic simulations can help in separating the rationale coming from the lack of surviving ganglion cells. The second application of acoustic simulations in cochlear implants research is for determining the effect of electrode insertion depth on speech understanding. In cochlear implantation, electrode arrays are inserted only partially to the cochlea, typically 22-30 mm, depending on the state of the cochlea. Acoustic simulations of cochlear implants could be used to simulate the effect of depth of electrode insertion on identification accuracy. In this respect, normal hearing listeners performed identification tasks with an acoustic simulation of cochlear implant whose electrodes are separated by 4 mm (Dorman et al., 1997). Insertion depth was simulated by outputting sine waves from each channel of the processor at a frequency determined by the cochlear place of electrode inserted 22-25 mm into the cochlea, through the Greenwood's frequency-to-place equation (Greenwood, 1990). The results indicated that simulated insertion depth had a significant effect on speech identification performance (Dorman et al., 1997).

In the second part of the chapter, we introduce our own experiences in extending the use of acoustic simulations of cochlear implants beyond conventional applications in cochlear

implant research. To this end, we investigate the automatic recognition of cochlear implant-like spectrally reduced speech (SRS) which are, essentially, acoustic simulations of cochlear implants (Shannon et al., 1995). Actually, state-of-the-art ASR systems rely on relevant spectral information, extracted from original speech signals, to recognize input speech in a statistical pattern recognition framework. We show that from certain SRS spectral resolution, it is possible to achieve (automatic) recognition performance as good as that attained with the original clean speech even though the cochlear implant-like SRS is synthesized only from subband temporal envelopes of the original clean speech (Do et al., 2010b). Basing on this result, we introduce a novel framework for noise robust ASR using cochlear implant-like SRS. In this novel framework, cochlear implant-like SRS is used in both the training and testing conditions. Experiments show that the (automatic) recognition results are significantly improved, compared to the baseline system which does not employ the SRS (Do et al., 2011). On the other hand, we present quantitative analyses on the fundamental frequency (F0) of the cochlear implant-like SRS. These analyses support the report of Zeng et al. (2005), which was based on subjective tests, about the difficulty of cochlear implant users in identifying speakers. Following the results of our analyses, the F0 distortion in state-of-the-art cochlear implants is great when the SRS, which is acoustic simulation of cochlear implant, is synthesized only from subband temporal envelopes (Do et al., 2010a). The analyses revealed also a significant reduction of F0 distortion when the frequency modulation is integrated in cochlear implant, as proposed by (Nie et al., 2005). Consequently, the results of such quantitative analyses, performed on relevant acoustic traits, could be exploited to conduct subjective studies in cochlear implant research.

1. References

Do, C.-T., Pastor, D. & Goalic, A. (2010a). On normalized MSE analysis of speech fundamental frequency in the cochlear implant-like spectrally reduced speech, *IEEE Trans. on Biomedical Engineering* **Vol. 57**(No. 3): 572–577.

Do, C.-T., Pastor, D. & Goalic, A. (2010b). On the recognition of cochlear implant-like spectrally reduced speech with MFCC and HMM-based ASR, *IEEE Trans. on Audio, Speech and Language Processing* **Vol. 18**(No. 5): 2993–2996.

Do, C.-T., Pastor, D. & Goalic, A. (2011). A novel framework for noise robust ASR using cochlear implant-like spectrally reduced speech, *Speech Communication* . (in revision).

Dorman, M., Loizou, P. C. & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding, *J. Acoust. Soc. Am.* **Vol. 102**(No. 5): 2993–2996.

Greenwood, D. D. (1990). A cochlear frequency-position function for several species - 29 years later, *J. Acoust. Soc. Am.* **Vol. 87**(No. 6): 2592–2605.

Loizou, P. C., Dorman, M. & Tu, Z. (1999). On the number of channels needed to understand speech, *J. Acoust. Soc. Am.* **Vol. 106**(No. 4): 2097–2103.

Nie, K., Stickney, G. & Zeng, F.-G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise, *IEEE Trans. on Biomedical Engineering* **Vol. 52**(No. 1): 64–73.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. (1995). Speech recognition with primarily temporal cues, *Science* **Vol. 270**(No. 5234): 303–304.

Zeng, F.-G., Nie, K., Stickney, G., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C. & Cao, K. (2005). Speech recognition with amplitude and frequency modulations, *PNAS* **Vol. 102**(No. 7): 2293–2298.