SPEAR: A SPEECH DATABASE FOR THE ADVANCEMENT OF INTRA-AURAL WEARABLE TECHNOLOGY

Rachel E. Bouserhal[1,2], Jérémie Voix[1,2]
[1] Mechanical Engineering, École de technologie supérieure, Montréal, Québec, Canada
[2] Centre for Interdisciplinary Research in Music Media and Technology, Montréal, Québec, Canada

***Keywords*: Speech, Lombard Effect, Database, Wearables, In-ear Microphone**

Introduction: The use of in-ear microphones with intelligent intra-aural devices is growing. This is due to the rise in hearable technology and the advantages to communication offered by using an in-ear microphone. In noisy conditions, the in-ear microphone captures a speech signal with a relatively high signal-to-noise ratio since it is usually placed past the passive attenuation of the earplug. However, speech captured inside the occluded ear is limited in its frequency bandwidth and has an amplified low frequency content. In addition, occluding the ear canal with an intra-aural device affects speech production, which could have detrimental effects on speech processing algorithms. An in-ear microphone speech database in noise and in quiet is therefore essential to the advancement of intra-aural technologies utilizing an in-ear microphone. Yet, to the authors knowledge no such database exists. This work presents a Speech-in-EAR (SpEAR) database in various conditions of the audio-phonation loop.

Methods: Twenty four participants (11 in English and 13 in French) took part in this study. Speech was collected in an audiometric booth using a reference microphone placed in front of the mouth as well as an intra-aural device equipped with in-ear microphones, outer-ear microphones, and miniature loudspeakers. Hearing-in-noise test sentences in French and English were read in four conditions: (1) quiet open-ear, (2) quiet occluded, (3) occluded in noise (ambient), and (4) occluded in noise (regenerated inside the ear). The first and second condition serve as a baseline reference for comparison when understanding the effects on speech production caused by occluding the ear in quiet and in noise. For the third condition, factory noise is played at 95 dBA in the room while the participants read each sentence. This condition is a realistic condition meant to aid researchers and developers working on denoising algorithms for intra-aural wearable devices. In the fourth condition noise is regenerated inside the occluded ear, causing the Lombard effect to be triggered but leaving the outer-ear microphones and the microphone in front of the mouth free of noise. In this case, the changes invoked by the Lombard effect and their interaction with occluding the ear can be studied without the nuisance of noise.

Results: Precursory acoustical analysis showed that the presence of noise and occlusion have a significant ($p < 0.001$) effect on the speech level of participants. Once occluded and in quiet, on average participants raise their speech level by 2.6 dBA compared to the open-ear condition. At the introduction of 95 dBA of factory noise, the average speech level increases on average by 6.5 dB. Analysis showed that males and females do not speak at different average speech levels and that language spoken has no significant effect on speech level over all conditions.

Conclusions: SpEAR aims to aid researchers and developers working with intra-aural wearable technology to develop speech algorithms for adverse realistic conditions. It is meant to respond to a lack of in-ear speech database and can deepen the understanding of changes in speech production caused by noise and occluding the ear.

Acknowledgements: The authors would like to acknowledge the funding received from the Centre for Interdisciplinary Research in Music Media and Technology, the Fonds de recherche du Québec Nature et technologies, the Natural Sciences and Engineering Research Council of Canada, and the NSERC-EERS Industrial Research Chair in In-Ear Technologies.