Effects of vibrotactile stimuli on perception of voiced and unvoiced bilabial stop consonants in noise

Asuka Ono, Momoko Nakatani, Ai Nakane (NTT Service Evolution Laboratories), Junji Watanabe, and Sadao Hiroya (NTT Communication Science Laboratories)

It has recently been reported that applying vibrotactile stimulation, the waveforms of which are generated from speech, for replicating the laryngeal vibration to the fingertips improves discrimination between degraded voiced and voiceless consonants in consonant-vowel (CV) syllables [1,2]. This stimulation has been applied to an entire speech segment, but our question is whether only vibrotactile stimulation replicating voiced consonants given to unvoiced consonants changes the perception of consonants in CV syllables. The vibrotactile effect was also tested in a noisy situation [1], but the relationship between the effect and auditory efficacy (noise level) remain to be clarified. In this study, we investigated the identification of speech stimuli /ba/ and /pa/ with vibrotactile stimulation just in the C region or across the CV region at various noise levels.

Speech samples of /ba/ and /pa/ spoken by a native Japanese female speaker were presented with and without vibrotactile stimuli. The following three vibration conditions were analyzed: no vibration, vibration across the CV region, and vibration just in the C region. The speech samples were masked with pink noise in five speech-to-noise ratio (SNR) steps and presented to participants via headphones. The SNR range for the /pa/ sample was set higher than that for the /ba/ sample to match intelligibility. Each vibrotactile stimulus was transmitted to the backs of the participants' hands [3] from a pair of 45-mm-diameter circular vibrating actuators (Acouve, Vp210). The vibrotactile stimuli were sinusoidal wave vibrations and were approximately 15 ms for the C region and 200 ms for the CV region. The vibrotactile stimuli given to the C and CV regions were common to both /ba/ and /pa/, so the frequency and phase of speech and vibrotactile stimulation did not match, unlike in a previous study [1]. Twenty-one native Japanese speakers (age 26 to 58) with no prior training were asked to identify consonants for each heard speech in an open-ended task. Each vibration condition was repeated 10 times, and the correct response rates were compared.

Figure 1 shows the correct response rates. Two-way repeated measure analysis of variance showed that there was no interaction between vibration conditions and SNRs, and that there was a main effect on vibration conditions for both /pa/ and /ba/. Bonferroni post-hoc tests revealed that a significant difference was found between no vibration and vibration across the CV region, but not between no vibration and vibration just in the C region for /pa/ and /ba/. Analysis of the responses showed that even if vibration across the CV region was given to either the /pa/ or /ba/ sample, it was biased toward /ba/. In a previous study [1], no experiment was conducted on hearing the sound of /pa/ as /ba/ by applying vibration to the C region. Furthermore, the difference between the correct response rates of no vibration and vibration across the CV region regarding the SNR showed that when the SNR was lowered, the effect of vibrotactile stimulation across the CV region became maximum, and the effect was weakened thereafter (Fig. 2).

Our results indicate that vibrotactile stimulation across the CV region but not just in the C region biased the consonant perception toward voiced. Therefore, to alter perception for a C region, vibrotactile stimulation during a vowel region is assumed necessary. According to a study on vibrotactile stimulation delayed by 50 ms [2], vibrotactile stimulation during just a vowel region will possibly alter perception for a C region, but further investigation is needed. The non-linear effect of vibrotactile stimulation on the SNR suggests that there is an appropriate SNR range for vibrotactile and auditory integration. For future work, we intend to investigate the vibrotactile effect in more complex contexts.

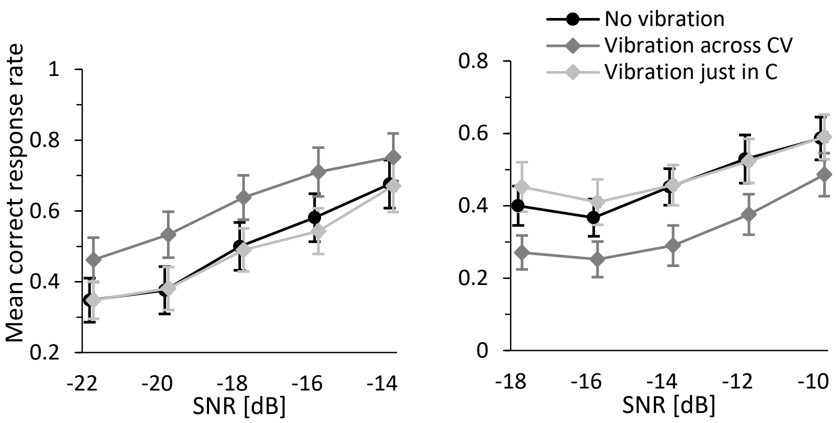

Fig. 1 Mean correct response rates for /ba/ (left) and /pa/ (right). Error bars indicate SEs.

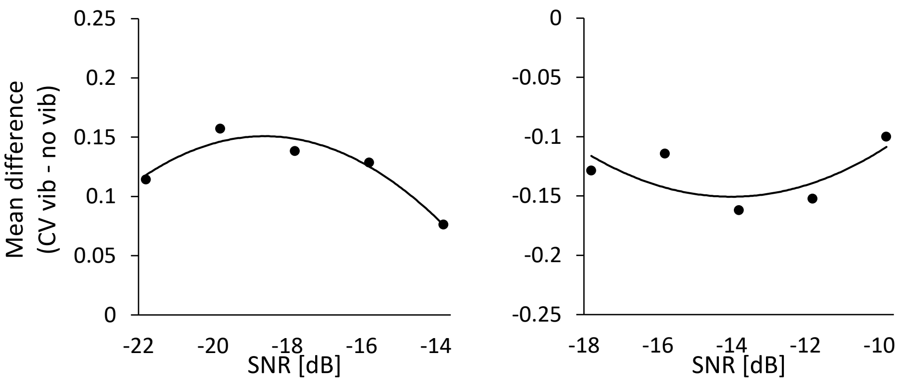

Fig. 2 Mean difference in correct response rate between vibration across CV and no vibration for /ba/ (left) and /pa/ (right) and 2nd order polynomial fit.

References

[1] D. G. Marino, H. Elbaggari, T. H. Chu, B. Gick, and K. MacLean, "Single-Channel Vibrotactile Feedback For Voicing Enhancement In Trained And Untrained Perceivers", Canadian Acoustics, vol. 46, no. 4, pp. 44–45, Feb. 2019.

[2] T. H. Chu, D. Marino, H. Elbaggari, B. Gick, and K. MacLean, "Testing Symmetry Of Temporal Window Of Integration In Vibrotactile And Auditory Speech Information On Phoneme Perception", Canadian Acoustics, vol. 46, no. 4, pp. 46–49, Feb. 2019.

[3] B. Gick and D. Derrick, "Aero-tactile integration in speech perception," Nature, vol. 462, no. 7272, pp. 502–504, 2009.