Perceiving speech produced with face masks in competing talker environments

Faith Chiu[1], Laura Bartoševičiūtė[2], Albert Lee[3] and Yujia Yao[2]

[1]*University of Glasgow (UK),* [2]*University of Essex (UK),* [3]*Education University of Hong Kong (Hong Kong SAR)*
faith.chiu@glasgow.ac.uk, laurabartoseviciute@gmail.com, albertlee@eduhk.hk, vera.yujia.yao@gmail.com

Introduction. The COVID-19 pandemic has reshaped speech communication. Face-to-face communication often includes one or both parties sporting a face mask. The listener's comprehension effort now involves adapting to mask-imposed distortions to the acoustic speech signal [1]. Even native speakers [2] struggle with understanding speech produced with a face mask when presented in noise. Everyday speech communication can take place in a noisy background with a competing talker. It is also not uncommon these days to converse in one's second or additional language. Using two experiments, this study aims to understand the difficulty imposed by speech produced with face masks in a multi-talker environment. Target sentences produced with and without a face mask were presented to listeners in the presence of a competing talker. The competing speech either matched or differed in language from target sentences. Participants' linguistic background determined the intelligibility of the competing talker.

Stimuli. The auditory stimuli consisted of target sentences in English and in Lithuanian, and competing speech in English and in Lithuanian. English target sentences were based on the British English version of the International Matrix sentence test [3] using a 50-word base matrix (10 names, 10 verbs, 10 numerals, 10 adjectives, and 10 nouns). Sentences were generated using a random combination of one word of each category in a fixed syntactic structure ('Alan bought two big beds'). Lithuanian target sentences follow the same format and were constructed as original stimuli. Target sentences were recorded by a native female speaker of each language. Individual words were produced, with and without a cotton fabric face mask, then combined acoutically. The competing speech was semantically meaningful sentences in either English or Lithuanian produced by a male speaker without a face mask. The target sentences were presented at a challenging level (−10dB Signal-to-Noise ratio). Male voices were chosen for competing speech and female for target sentences so participants can utilise speaker sex as a segregation cue.

Participants. 24 native Lithuanian listeners (13 female and 11 male, age range: 18–37) took part in *Experiment 1*. Participants for *Experiment 2* were 22 monolingual British English speakers (16 female and 6 male, age range: 18–34) and 22 second language speakers of English with Mandarin Chinese as first language (19 female and 3 male, age range: 20–31).

Procedure. *Experiment 1* was conducted online. Participants were instructed that they would hear target sentences by a female talker in the presence of a male competing talker and that they were to listen only to the female voice and ignore the male. They then had to type what they heard after each sentence. Participants heard a total of 160 trials: from 2 TARGET LANGUAGES (English/Lithuanian) × 2 MASK conditions (YES/NO) × 2 COMPETING SPEECH LANGUAGES (English/Lithuanian) with 20 sentences each. *Experiment 2* features an identical procedure to *Experiment 1* except that participants heard only English target sentences produced with and without mask, in the presence of either competing English or Lithuanian speech. Responses were scored based on the number of words accurately reported in each sentence.

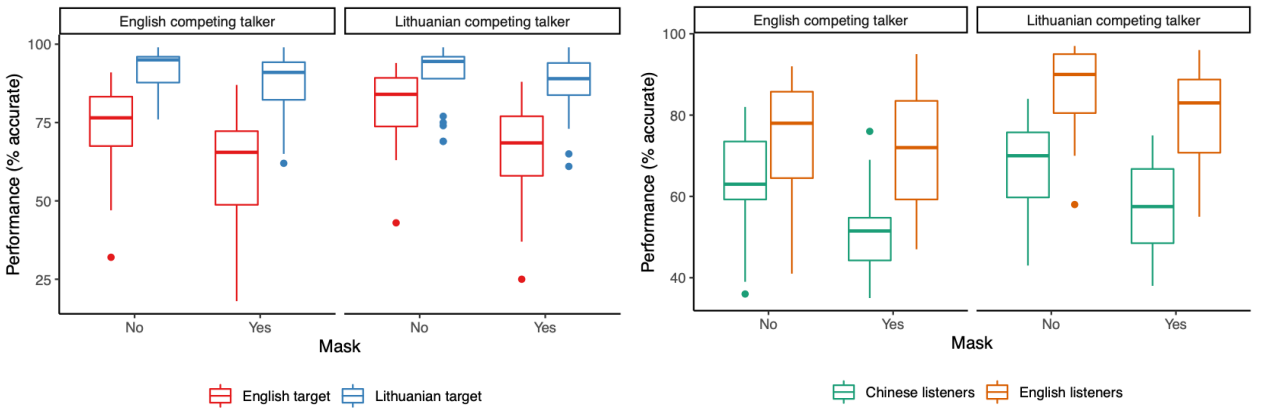
Results. The left figure reveals Lithuanian listeners' performance (*Experiment 1*). The right figure shows the performance of English and Chinese listeners (*Experiment 2*).

Experiment 1. A 2 × 2 × 2 analysis of variance (ANOVA) was conducted on the percentage of accurately reported keywords as a function of language of TARGET (English versus Lithuanian), MASK (with or without a face mask), and language of COMPETING SPEECH (English vs. Lithuanian). The results revealed two significant two way-interactions: TARGET × MASK ($F(1, 23) = 26.001$, $p < .001$, $\eta p^2 = .531$) and TARGET × COMPETING SPEECH ($F(1, 23) = 25.123$, $p < .001$, $\eta p^2 = .522$). Individual 2 × 2 ANOVAs were performed for each target language as follow-up, comparing the %accuracy as a function of MASK and language of COMPETING SPEECH. There was a significant main effect of MASK ($F(1, 23) = 54.448$, $p < .001$, $\eta p^2 = .703$) for English target sentences. More keywords were accurately reported on sentences produced without a face mask, in both English and Lithuanian competing speech. There was also a main effect of COMPETING SPEECH ($F(1, 23) = 31.304$, $p < .001$, $\eta p^2 = .576$). Lithuanians listeners were less accurate when the competing speech was in a language which matches

the English target sentences; this was true both when the targets were produced with and without a mask. However, when listening to Lithuanian target sentences, there was only a main effect of MASK ($F(1, 23) = 15.544$, $p < .001$, $\eta p^2 = .403$). Lithuanian target sentences produced with a face mask were more poorly perceived, and this was true regardless of whether it was presented in both English and Lithuanian competing speech. Unlike in English target sentences, there was no effect of COMPETING SPEECH. Planned comparisons showed that Lithuanian listeners reported more accurate keywords when listening to Lithuanian than English targets.

Experiment 2. A mixed 2 × 2 × 2 ANOVA was conducted on %accuracy as a function of MASK (with or without a face mask), language of COMPETING SPEECH (English vs. Lithuanian), and GROUP (English vs. Chinese listeners). The results indicated a significant three-way interaction of MASK × COMPETING SPEECH × GROUP ($F(1, 42) = 6.497$, $p = .015$, $\eta p^2 = .134$). Planned comparisons showed English listeners outperforming Chinese listeners in all conditions. Individual 2 × 2 ANOVAs performed for each listener group revealed that for English listeners, there was a main effect of MASK ($F(1, 21) = 5.439$, $p = .030$, $\eta p^2 = .206$) as well as a main effect of COMPETING SPEECH ($F(1, 21) = 78.729$, $p. < .002$, $\eta p^2 = .789$). Chinese listeners were similar with both a main effect of MASK ($F(1, 21) = 21.960$, $p < .001$, $\eta p^2 = .511$) and COMPETING SPEECH ($F(1, 21) = 19.869$, $p < .001$, $\eta p^2 = .486$).

Discussion. In sum, masked speech is always more poorly perceived across all listener groups in all conditions. This finding echoes existing reports of decreased perception performance when listening to speech produced with a face mask and presented in noise [2]. This across-the-board effect could be due to attenuation of the acoustic signal from mask-wearing in the form of dampening. In particular, high frequency information is lost [1]. Additionally, perception accuracy was higher when listening in one's first language, echoing previous work showing that speech perception with a competing talker is more difficult in one's non-native language [4]. Finally, a competing talker in a language which matches the target sentences had more of a detrimental effect on perception accuracy compared to a mismatched one. This replicates findings of a benefit of linguistic mismatch between target and competing speech for non-native speakers [5]. Exceptionally in our study, when Lithuanian participants (with both English and Lithuanian knowledge) listened for Lithuanian targets there was there no added challenge from matching language of target and competing speech. We conclude that acoustic distortions from face masks present an across-the-board difficulty while linguistic knowledge can reduce distraction from competing speech.



References

[1] Corey, M. R., Jones, U., Singer, C. A. 2020. Acoustic effects of medical, cloth, and transparent face masks on speech signals. *J. Acoust. Soc. Am.* 148, 2371–2375.

[2] Yi, H., Pingsterhaus, A., Song, W. 2021. Effects of wearing face masks while using different speaking styles in noise on speech intelligibility during the COVID-19 pandemic. *Front. Psychol.* 12, 682677.

[3] Hall, S. J. 2006. *The development of a new English Sentence in Noise Test and an English Number Recognition Test.* MSc thesis. University of Southampton.

[4] Bradlow, A. R., Alexander, J. A. 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust. Soc. Am.* 121, 2239–2249.

[5] Brouwer, S., Van Engen, K.J., Calandruccio, L., Bradlow, A.R. 2012. Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *J. Acoust. Soc. Am.* 131, 1449–64.