THE INFLUENCE OF FACE MASKS ON NATIVE AND NON-NATIVE MEMORY RECALL OF SPOKEN SENTENCES

Thanh Lan Truong & Andrea Weber

English Department, University of Tübingen, Germany
thanh-lan.truong@uni-tuebingen.de, andrea.weber@uni-tuebingen.de

ABSTRACT

The impact of face masks on listeners‘ recall of spoken sentences was examined. Native (= L1) and non-native (= L2) listeners of German watched video clips of a native talker producing German sentences with and without a face mask, followed by a cued-recall task. Results showed that face masks significantly reduced L1 listeners' recall performance. By contrast, no significant effect was found for L2 listeners. While the mask effect attests to the importance of visual speech cues for L1 listeners, the absence of a mask-effect for L2 listeners likely suggests that only listeners with a higher language proficiency in the second language can benefit from an audiovisual context for memory encoding.

Keywords: face mask, cued-recall, memory, audiovisual

1. INTRODUCTION

In times of the COVID-19 pandemic, face-to-face communication has often become more challenging, because face masks are frequently used and can degrade the acoustic speech signal such that higher frequencies are particularly affected, similar to a low-pass filter [1, 2]. In addition, face masks cover the mouth region of a talker's face, thus restricting access to visual articulatory cues that can be helpful in language comprehension, particularly when speech is degraded [3, 4].

Visual cues, such as lip and jaw movements, can facilitate comprehension (e.g., [5, 6]) as they convey important phonological information about speech sounds [3, 4]. For example, while closed lips indicate a bilabial place of articulation (e.g., /p/ and /b/), an open jaw indicates vowel height (e.g., more open jaw for the vowel /a/ and less open jaw for /i/). Thus, visual cues can supplement and complement information about speech sounds that is not present in the auditory signal itself [6, 7].

Recent findings suggest that the difficulties listeners encounter when listening to speech produced with a face mask are likely to stem from both the acoustic degradation of the speech signal and the lack of visual cues of the talker's mouth movements. For example, face masks increase listening effort and reduce L1 adults' correct identification of words and sentences in noisy conditions [1, 8, 9]. In addition, they affect L1 adults' memory recall in quiet conditions such that L1 listeners remember fewer words when talkers are wearing a mask compared to when they are not wearing one [10]. This effect has also been replicated for L2 speech but only in noisy listening conditions [11].

All of these findings are in line with the Framework for Understanding Effortful Listening [12], the Effortfulness Hypothesis [13], and the Ease of Language Understanding [14], postulating that more cognitive resources are utilized for speech comprehension when listening conditions are adverse (i.e., speech produced with a face mask), leaving fewer resources available for memory encoding [15], which is fueled by limited cognitive resources [16]. While findings for L1 listeners imply that face masks negatively affect memory encoding, the question remains whether face masks also affect L2 listeners' memory for spoken language. Previous studies on L2 listening found that the presence of visual articulatory cues can enhance L2 participants' perception of French [17], Korean [18], Irish and Spanish sounds [19]. Therefore, visual cues can facilitate L2 listeners' perception [20], and L2 listeners might even pay more attention to the information that is conveyed by visual articulatory movements to make up for their poorer comprehension skills in the L2 language [21]. We, therefore, further investigated if sufficient cognitive resources are left for visual cues to affect memory encoding in L2.

The goal of the present study was to compare L1 and L2 participants' memory performance of German sentences produced by an adult talker when the talker was wearing a face mask or was wearing no mask using a cross-modal cued-recall task.

2. METHOD

2.1. Participants

For the experiment, we recruited two groups of participants: Forty L1 listeners of German via social media and university email[1] between 19 and 37 years of age (mean = 23.5; SD = 4.0; 30 females) and forty L2 listeners of German, with English as their native language, between 18 and 58 years of age (mean = 31.7; SD = 12.7; 21 females) via Prolific. Four participants had to be excluded from further analyses because they did not follow the instructions or did not complete the experiment. For the L2 participants, we used the pre-screeing function of Prolific, which ensured that only participants who registered themselves with an intermediate or advanced level of German were allowed to participate. Participants' average current daily use of German ranged from 0-90%; 17 reported having lived in Germany before (min: six months, max: seven years). Additionally, the L2 group's German proficiency was measured via self-report on a scale ranging from 1 (very poor) to 7 (very good) for all four modalities (i.e., writing, listening, speaking, and reading). The average proficiency score was 4.85 (SD = 1.45). Particularly, average score for listening was 4.65 (SD = 1.48).

2.2. Material

The stimuli consisted of 48 meaningful German sentences that were low in predictability and had been modelled after the Oldenburger Satztest [22]. Since sentence context did not semantically restrict lexical options, the possibility of context rectifying comprehension failures was fairly low and ensured a more careful processing of the input [23]. The syntactic structure was the same for all 48 sentences, and each content word occurred only once in the complete set of sentences. The sentences all began with a determiner and a noun, followed by a verb, an adverb, an adjective, and a noun (e.g., *Die Köchin hilft montags armen Kindern*, "the cook helps on Mondays poor children").

A 22-year-old female native talker of German was recorded on video in a sound-attenuated room, and produced all sentences with and without a face mask. The face mask consisted of two fabric layers: The inner layer was a thin fleece, and the outer layer was cotton. When necessary, the talker repeated a sentence until it was produced without any errors or hesitations. The talker was instructed to produce all sentences at a normal speaking rate and to avoid enunciating more loudly or clearly when wearing the mask. The videos were recorded by using a Sony (Tokyo, Japan) DSC-Hx90 camera with video resolution parameters set to FULL HD 1920x1080. Audio was recorded at a sampling rate of 48 kHz with a high-quality microphone placed in front of the talker. The average F0 value of the talker was 235.5 Hz. Durations for sentences produced without a mask were on average 3255 ms, and with a mask they were 3178 ms (t = 1.35 p = 0.18). Spectral analysis [root mean square (rms) power] of the talker revealed no difference between sentences with (56.6 dB) and without a face mask (56.7 dB) (t = 0.28, p = 0.77).

2.3. Procedure

The experiment was conducted online.[2] Prior to the experiment, participants electronically gave informed consent.[3] The experiment started with two practice trials followed by eight blocks with six sentences each. The self-paced cued-recall task followed immediately after each block. For this task, sentences of the preceding block were presented up to the adverb orthographically on the screen (e.g., *Die Köchin hilft montags*, "the cook helps on Mondays"), and participants were asked to enter the missing two final words (e.g., *armen Kindern*, "poor children") on their keyboard. The recall cues for all six sentences of a block were presented at the same time in the order of block presentation, and participants could fill in their responses in any order. At the end of the experiment, participants filled out a brief language background questionnaire and were asked about technical problems of which none were reported.

3. RESULTS

Each correctly recalled word received the score correct (1)[4] and each erroneously recalled or missing word received the score incorrect (0) (see Fig. 1). There were two keywords for each of the 48 sentences, making a total of 96 keywords to be recalled per participant. L1 participants recalled 57.3% of the words correctly and 42.7% incorrectly. In contrast, L2 participants recall performance was extremely low such that only 16.9% of the words were recalled correctly and 81.3% incorrectly. Subsequent descriptive analyses of the incorrectly recalled words indicated that the majority of incorrectly recalled words had been full omissions of a keyword (68% for L1 listeners; 50% for L2 listeners). The remaining incorrect responses consisted of a variety of error types. Some responses were unrelated in form and in meaning

to the intended words (e.g., *schwarze Schuhe*, "black shoes," for *staubige Kissen*, "dusty pillows"), fewer responses were closely semantically related (e.g., *Dackel*, "Dachshund," for *Hunde*, "dogs") and an even smaller number of responses consisted of potential typos (e.g., the nonword Lmpen for *Lampen*, "lamps").

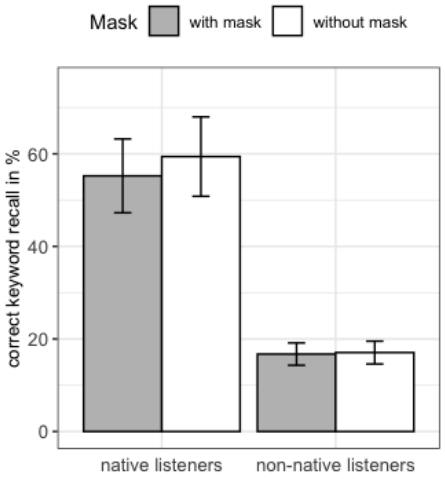

Figure 1: Average percentage of keywords recalled correctly for sentences with and without a face mask. The vertical bar represents standard errors.

A logistic mixed-effects regression model with the lme4 package in R (version: version 4.0.5) [24] was used to analyze the effect of face masks on participants' correctly recalled keywords [25]. Keyword recall (success vs. failure) was the dichotomous dependent variable, and native language (L1 vs. L2), face mask (mask vs. no mask), and block (8 blocks) were the independent variables; face mask x native language x block was included as an interaction term. To test linear and quadratic effects of block, orthogonal polynomials were used [26]. Additionally, to account for extra variation, fixed factors of sentence duration, rms power, and language proficiency were also included in the model. The German proficiency factor consisted of the sum of self-reported ratings (i.e., from 1 the lowest to 7 the highest) in the category of writing, listening, speaking, and reading. We found a main effect of native language (b = 4.49, SE = 0.75, p < .001) with higher recall rates for the L1 group than the L2 group, an interaction between linear effects of block and native language (b = 1.0, SE = 0.42, p = .02), suggesting better recall performance as the experiment progressed for participants who match the first language with the talker (i.e., L1 German as opposed to L2 German) and an interaction between face mask and native language (b = -0.51, SE = 0.17, p = .003), suggesting different recall performance patterns based on participants' native language.

We then grouped the data based on the participants' native language (i.e., L1 vs. L2). The separate analysis for L1 participants showed a significant effect of face mask, with L1 listeners recalling fewer words when the talker was wearing a mask compared to when the talker was not wearing a mask (b = -0.28, SE = 0.11, p = .01). Recall was furthermore better for shorter sentence durations than for longer ones as the main effect for sentence duration showed (b = -0.47, SE = 0.23, p = .04).

Also, a main effect of rms (b = -83.6, SE = 36.6, p = .02) was found, indicating that sentence recordings with less rms power were recalled better than sentences with higher rms power. By contrast, L2 listeners showed no effect for face mask (b = 0.09, SE = 0.14, p = .52), rms (b = 50.2, SE = 57.1, p = .38), sentence duration (b = -0.19, SE = 0.42, p = .65) and language proficiency (b = 0.11, SE = 0.13, p = .37).

The results suggest that, for L1 listeners only, processing was easier when visual cues were available than when they were not, and this availability left more cognitive resources for successful memory encoding. In short, the findings for L2 listeners suggest that the covering of visual articulatory cues with a face mask did not negatively influence recall performance.

4. CONCLUSIONS

In a cued-recall experiment, we investigated the impact of face masks on L1 and L2 memory for spoken language. We found an effect of mask for L1 listeners but not for L2 listeners. That is, L1 listeners significantly recalled fewer words correctly when the talker produced the sentences in quiet with a mask compared to when the talker was not wearing one, suggesting that processing speech with face masks leaves fewer cognitive resources available for memory encoding [14, 13, 12], as face masks both cover visual speech cues and degrade the acoustic signal [1, 2]. At first glance, our findings for L1 listeners contradict Smiljanic et al. [11] who only found a negative effect of face masks for L1 listeners when additional noise was added to the speech signal. One possible reason for the difference in findings could be attributed to the material of our study. To avoid a facilitatory influence of context on comprehension and recall, we used dissociated sentences with low predictability whereas Smiljanic

et al. [11] used a coherent text, which made the listening environment more naturalistic but also possibly allowed correctly guessing of individual words, especially when comprehension was more difficult, for example due to a face mask. Since cohesive information is easier to encode than dissociated information (e.g., [27]), this alleviation through cohesion might have been responsible for the lack of a mask effect when listening to the stimuli in quiet in Smiljanic et al. [11]. The present study was furthermore the first to test recall memory for L2 listeners, and while the results cannot be directly compared to previous studies, a number of reasons could be responsible for the lack of a mask effect for L2 listeners. First of all, it could be the case that L2 listeners' encoding was indeed unaffected by the mask because these listeners did not use visual speech information for perception in the first place. While previous studies mostly confirmed that L2 listeners can use visual cues [17, 19, 18] for perception, Wang et al. [28] found that the ability for audiovisual speech perception in L2 varies significantly across native languages of the participants. Thus, participants with different native language backgrounds might still show an effect of face mask on recall. Alternatively, the lack of contextual cues could have made the task of comprehension of the speech signal so challenging that no attention was paid to the visual speech cues by L2 participants. In that case, an effect of face masks should emerge for L2 participants when cohesive texts are used as, for example, in Smiljanic et al.. However, the most likely explanation has to do with the generally low performance of our L2 participants. While the L1 listeners performed with an accuracy of 57.3% on average, the L2 listeners only reached an average of 16.9%. Although our L2 participants reported intermediate to advanced proficiency in German and our sentences comprised of common German words with a high lexical frequency, the language skills of our L2 participants might have been too low for an effect of masks to be observed. In other words, the L2 listeners' performance could be a floor effect. Higher proficiency in L2 might thus be necessary to optimally make use of the enhancement provided by visual speech cues. Indeed, advanced linguistic expertise can improve working memory performance, leading to greater automatic processing and thus to smaller processing cost in comprehension of the stimuli. This in turn leaves a larger amount of cognitive resources available that can be employed for memory encoding. As such, L2 comprehension and working memory capacity are mediated by language proficiency. Given that working memory operates on limited cognitive resources [16], we propose that the following happened for the L2 listeners: The low language competence in German in combination with the difficulty of the task required extra processing demands for comprehension, leaving no resources for retaining information in the working memory [29, 30].

In conclusion, we present evidence for a negative impact of face masks on memory recall for L1 listeners. We interpret the absence of a mask effect for L2 listeners as being modulated by the task's plausibility and the language proficiency of L2 listeners. Although the results showed no impact of face mask on memory encoding for L2 listeners, the results of the present study reinforce previous findings that state that L2 language proficiency plays a crucial role in extracting visual cues from the lips of the talker [28, 31].

5. REFERENCES

[1] P. Bottalico, S. Murgia, G. E. Puglisi, A. Astolfi, and K. I. Kirk, "Effects of masks on speech intelligibility in auralized classrooms," *The Journal of the Acoustical Society of America*, vol. 148, no. 5, pp. 2878–2884, 2020.

[2] R. M. Corey, U. Jones, and A. C. Singer, "Acoustic effects of medical, cloth, and transparent face masks on speech signals," *The Journal of the Acoustical Society of America*, vol. 148, no. 4, p. 2371–2375, 2020.

[3] R. Campbell, "The processing of audio-visual speech: empirical and neural bases," *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 363, no. 1493, pp. 1001–1010, 2008.

[4] Q. Summerfield, "Lipreading and audio-visual speech perception," *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 335, no. 1273, pp. 71–78, 1992.

[5] J. Navarra and S. Soto-Faraco, "Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds," *Psychological Research*, vol. 71, no. 2, pp. 4–12, 2007.

[6] W. H. Sumby and I. Pollack, "Visual contribution to speech intelligibility in noise," *The Journal of the Acoustical Society of America*, vol. 26, p. 212, 1954.

[7] D. W. Massaro, *Categorical perception: The groundwork of cognition*. Cambridge University Press, 1987, ch. Categorical partition: A fuzzy-logical model of categorization behavior., pp. 254–283.

[8] V. A. Brown, K. J. Van Engen, and J. E. Peelle, "Face mask type affects audiovisual speech intelligibility and subjective listening effort in

young and older adults," *Cognitive Research: Principles and implications*, vol. 6, no. 1, p. 49, 2021.
[9] M. Randazzo, L. L. Koenig, and R. Priefer, "The effect of face masks on the intelligibility of unpredictable sentences," *Proceedings of Meetings on Acoustics*, vol. 42, no. 1, p. 032001, 2020.
[10] T. L. Truong and A. Weber, "Intelligibility and recall of sentences spoken by adult and child talkers wearing face masks," *The Journal of the Acoustical Society of America*, vol. 150, no. 3, pp. 1674–1681, 2021.
[11] R. Smiljanic, S. Keerstock, K. Meerman, and S. M. Ransom, "Face masks and speaking style affect audio-visual word recognition and memory of native and non-native speech," *The Journal of Acoustical Society of America*, vol. 149, no. 6, p. 4013, 2021.
[12] M. K. Pichora-Fuller, S. E. Kramer, M. A. Eckert, B. Edwards, B. W. Y. Hornsby, L. E. Humes, U. Lemke, T. Lunner, M. Matthen, C. L. Mackersie, G. Naylor, N. A. Phillips, M. Richter, M. Rudner, M. S. Sommers, K. L. Tremblay, and A. Wingfield, "Hearing impairment and cognitive energy: The framework for understanding effortful listening (fuel)," *Ear and Hearing*, vol. 37, pp. 5S–27S, 2016.
[13] S. L. McCoy, P. A. Tun, L. Clarke Cox, M. Colangelo, and J. A. Stewart, "Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech," *Quarterly Journal of Expoerimental Psychology*, vol. 58, no. 1, pp. 22–33, 2005.
[14] J. Rönnberg, T. Lunner, A. Zekveld, P. Sörqvist, H. Danielsson, B. Lyxell, Dahlström, C. Signoret, S. Stenfelt, M. K. Pichora-Fuller, and M. Rudner, "The ease of language understanding (elu) model: Theoretical, empirical, and clinical advances," *Frontiers in Systems Neuroscience*, vol. 7, p. 31, 2013.
[15] J. E. Peelle, "Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior," *Ear and Hearing*, vol. 39, no. 2, pp. 204–214, 2018.
[16] M. A. Just and P. A. Carpenter, "A capacity theory of comprehension: Individual differences in working memory," *Psychological Review*, vol. 99, no. 1, pp. 122–149, 1992.
[17] D. Reisberg, J. McLean, and A. Goldfield, *Hearing by Eye: The Psychology of Lip-Reading*. Erlbaum, Hillsdale, NJ, 1987, ch. Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli, pp. 97–113.
[18] C. Davis and J. Kim, "Audio–visual interactions with intact clearly audible speech," *The Quarterly Journal of Experimental Psychology Section A*, vol. 57, no. 6, pp. 1103–1121, 2004.
[19] V. D. Erdener and D. Burnham, "The role of audiovisual speech and orthographic information in nonnative speech production," *Language Learning*, vol. 55, pp. 191–228, 2005.
[20] D. W. Massaro, *Perceivig talkig faces: From speech perceptio to a behavioral principle*. The MIT Press, 1998.
[21] L. Drijvers and A. Özyürek, "Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension," *Language and Speech*, vol. 63, no. 2, pp. 209–220, 2020.
[22] *Oldenburger Satztest: Handbuch und Hintergrundwissen*. Oldenburg, Germany: Hörtech GmbH, 2000.
[23] J. Rommers and K. D. Federmeier, "Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects," *Cortex*, vol. 101, pp. 16–30, 2018.
[24] R Core Team, "R: A language and environment for statistical computing," Vienna, Austria, 2021, versions 4.0.5. [Online]. Available: https://www.r-project.org
[25] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, "Parsimonious mixed models," 2015.
[26] D. Mirman, *Growth curve analysis and visualization using R*. London: Taylor Francis, 2017.
[27] J. B. Black and H. Bern, "Causal coherence and memory for events in narratives," *Journal of Verbal Learning and Verbal Behavior*, vol. 20, no. 3, pp. 267–275, 1981.
[28] Y. Wang, D. M. Behne, and H. Jiang, "Influence of native language phonetic system on audio-visual speech perception," *Journal of Phonetics*, vol. 37, no. 3, pp. 344–356, 2009.
[29] E. Service, M. Simola, O. Metsänheimo, and S. Maury, "Bilingual working memory span is affected by language skill," *European Journal of Cognitive Psychology*, vol. 14, no. 3, pp. 383–408, 2010.
[30] M. Noort, P. Bosch, and K. Hugdahl, "Foreign language proficiency and working memory capacity," *European Psychologist*, vol. 11, pp. 289–296, 2006.
[31] Z. Xie, H. Yi, and B. Chandrasekaran, "Nonnative audiovisual speech perception in noise: Dissociable effects of the speaker and listener," *PloS one*, vol. 9, p. e114439, 2014.
[32] A. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. Evershed, "Gorilla in our midst: An online behavioral experiment builder," *bioRxiv*, 2019.

[1] The L1 memory data of these participants have been previously reported in [10].

[2] L1 participants were tested on *Alchemer* and L2 listeners on *Gorilla Experiment Builder* (www.gorilla.sc) [32].

[3] The headphone screening task was a special feature in Gorilla Experiment Builder which was implemented prior to our actual experiment. This task required participants to do the experiment with headphones. Those who passed the test were immediately directed to

the experiment.
[4] Umlauts with two letters were accepted as correct (e.g., Veogel, “birds,” for “Vögel/ Voegel”).