

Unmasking the Difficulty of Listening to Talkers With Masks: lessons from the COVID-19 pandemic

i-Perception

2021, Vol. 12(2), 1–11

© The Author(s) 2021

DOI: 10.1177/2041669521998393

journals.sagepub.com/home/ipe



Elena Giovanelli 

Center for Mind/Brain Sciences - CIMeC, University of Trento, Rovereto, Italy

Chiara Valzolgher

Center for Mind/Brain Sciences - CIMeC, University of Trento, Rovereto, Italy; Integrative, Multisensory, Perception, Action and Cognition Team - IMPACT, Centre de Recherche en Neurosciences de Lyon, Lyon, France

Elena Gessa

Center for Mind/Brain Sciences - CIMeC, University of Trento, Rovereto, Italy

Michela Todeschini

Dipartimento di Psicologia e Scienze Cognitive - DiPSCo, University of Trento, Rovereto, Italy

Francesco Pavani

Center for Mind/Brain Sciences - CIMeC, University of Trento, Rovereto, Italy; Integrative, Multisensory, Perception, Action and Cognition Team - IMPACT, Centre de Recherche en Neurosciences de Lyon, Lyon, France; Department of Psychology and Cognitive Sciences, University of Trento, Rovereto, Italy

The first two authors equally contributed to the work.

Corresponding author:

Francesco Pavani, Center for Mind/Brain Sciences, University of Trento, Corso Bettini 31, 38068 Rovereto, Italy.
Email: francesco.pavani@unitn.it



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Abstract

Interactions with talkers wearing face masks have become part of our daily routine since the beginning of the COVID-19 pandemic. Using an on-line experiment resembling a video conference, we examined the impact of face masks on speech comprehension. Typical-hearing listeners performed a speech-in-noise task while seeing talkers with visible lips, talkers wearing a surgical mask, or just the name of the talker displayed on screen. The target voice was masked by concurrent distracting talkers. We measured performance, confidence and listening effort scores, as well as meta-cognitive monitoring (the ability to adapt self-judgments to actual performance). Hiding the talkers behind a screen or concealing their lips via a face mask led to lower performance, lower confidence scores, and increased perceived effort. Moreover, meta-cognitive monitoring was worse when listening in these conditions compared with listening to an unmasked talker. These findings have implications on everyday communication for typical-hearing individuals and for hearing-impaired populations.

Keywords

speech processing, multisensory, speech in noise, facial masks, COVID-19

Date received: 21 September 2020; accepted: 8 February 2021

Introduction

From the onset of the COVID-19 pandemic, the use of face masks prevented many of us from seeing each other's lips in everyday interactions. Face masks cover about 60% to 70% of a face, hiding the lower part which is particularly useful for identifying emotional expressions (Blais et al., 2017; Carbon, 2020) and for supporting communication through lip reading (Rosenblum et al., 1996), especially in presence of background noise (Sumbly & Pollack, 1954). The few studies that have examined the impact of face masks on communication have reported a mask-induced attenuation of the voice between 2 and 12 dB (Atcherson et al., 2017; Goldin et al., 2020; Mendel et al., 2008) and showed a benefit of transparent masks in hearing-impaired individuals (Atcherson et al., 2017). In five normal hearing adults, Hampton et al. (2020) reported that word-identification performance drops by 30% to 35% when listening with background noise to a talker wearing a mask, compared with a control condition in which the talker's face was visible. In these previous studies, however, the impact of face masks on speech understanding was rarely examined in common everyday life conditions. Furthermore, the impact of concealing the lips (a visual feature) was not observed separately from the impact of voice distortions (an auditory feature) generated by the mask (i.e., transmission loss: Llamas et al., 2009).

In this study, we aimed to clarify the impact of face masks on speech understanding starting from one of the most common communication modalities imposed by social distancing, namely, video calls. Video conferences can impact on our possibility to exploit audio-visual cues in speech perception and can undermine the listening strategies normally involved in speech understanding (e.g., gazing at the talker), often making speech processing more challenging. In addition, a reduction of visual information about the talker is frequently present in multitalker video calls because talkers can decide to identify themselves through their name only. Finally, face masks can be adopted whenever the talker shares a physical space with others while participating in a video call.

In our study, we developed an on-line experiment that mimicked a real multitalker video call to measure the impact of these different visual conditions on speech comprehension in

typical hearing participants. Importantly, while the availability of visual cues changed across conditions (talkers fully visible, talker wearing surgical masks covering the lower part of the face, talkers hidden behind a black screen, and identified only by their names), the audio tracks were always recorded in the same natural conditions (i.e. they were not distorted by the fabric of the face mask, when this visual condition was used). In this way, we dissociated the visual impact of wearing a mask from the auditory impact of wearing a mask.

Materials and Methods

Participants

Thirty-six typical-hearing participants (mean age = 26.0 years, $SD = 4.64$, age range = [19–40], six males) took part in the study. All participants gave their informed consent before starting the experiment, which was conducted according to the criteria of the Declaration of Helsinki (1964, amended in 2013) and in accordance with the regulations at the University of Trento. All participants reported typical hearing by self-report (a necessary inclusion criterion).

Stimuli

We created an on-line listening-in-noise task that resembled a real video call with multiple talkers (see Figure 1). To prepare the experimental trials, a total number of six talkers (two males and four females) were video-recorded while uttering simple sentences. Each experimental trial was a video with sound (average duration 2.49 seconds, range [1.94–2.93]) that comprised one male talker (target) and one or three female talkers (distractors). Occasional trials with only the male talker were also included, to be sure that participants were paying attention to the task (catch trials). The two male talkers were equally represented in the resulting trials. Importantly, the talkers were either invisible and identified only by their names (Figure 1A), visible from the neck upward with the lower part of the faces covered by surgical masks (Figure 1B) or visible from the neck upward with their faces fully visible (Figure 1C).

All stimuli were prepared during the first Italian lock-down phase (April 2020), when access to the laboratories and gathering were forbidden. Each talker recorded the videos

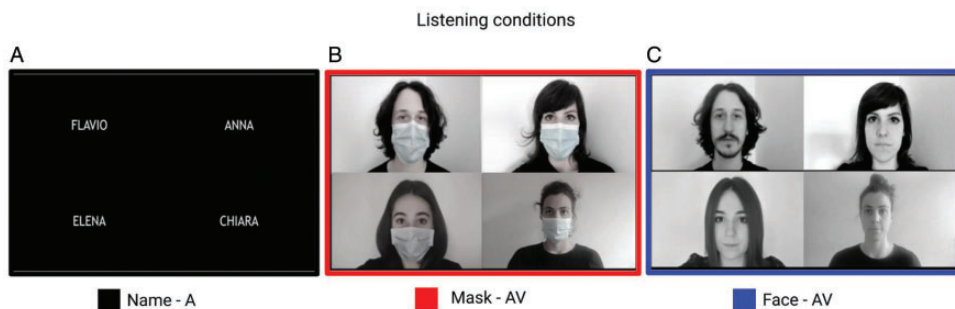


Figure 1. Listening conditions. A: Name condition, auditory only (A), in which talkers were identified only by their names. (B) Face mask condition, audio-visual (AV), in which the lower part of the face of each talker was covered with a surgical mask. (C) Face condition, AV, in which the entire face was visible. An example of the videos used in the study is available at: <https://osf.io/8dqbg/>

using their own Android smartphone, always in the same room and with similar illumination conditions, using empty white walls as backgrounds. The rooms in which the talkers recorded their videos were domestic furnished rooms (mean volume in m^3 : 38.22; *SD*: 5.38; as indicated by Díaz and Pedrero (2005), mean reverberation time for furnished rooms with volume between 30 and 40m^3 is 0.40 seconds; *SD*: 0.08). The distance between the talker and the phone camera was approximately 50 cm. Half of the videos were recorded while the talker wore a surgical mask, whereas the other half was recorded with the face fully visible. Each talker uttered a series of five-words phrases (e.g., “Anna compra dieci matite utili,” which is Italian for “Anna buys ten useful pencils”), adapted from the Italian version of the Matrix test (Puglisi et al., 2015). All phrases were syntactically and grammatically correct but semantically unpredictable to avoid effects of linguistic competence on listening.

Recorded audio and video tracks were treated separately using dedicated editing programs (Audacity 3.2.2 and DaVinci Resolve version 16.2.2.012, respectively). Depending on the smartphone, the audio sampling-rate ranged from 44.1 kHz to 48 kHz. First, each audio and video track was checked for quality (e.g., hesitation or pronunciation mistakes) and if necessary repeated. Next, all audio tracks were equalized in volume using the Contrast Analysis—WCAG 2 tool and saved in MP3 format (audio compression range: 170–210 kbps), before assembling them into the experimental trials. Each experimental trial was the result of two processes: the overlapping of audio tracks for all talkers in the trial and the assembly of the corresponding videos in a 2×2 matrix (when the trial comprised four talkers they were arranged as in Figure 1; when the trial comprised two talkers, they were arranged along one of the two diagonals; when the trials comprised a single talker it was placed in either one of the four squares of the matrix). Audio track equalization allowed measuring signal-to-noise-ratio (SNR) for each experimental condition: $\text{SNR} = 0$ dB for one distracting talker trials, $\text{SNR} = -4.78$ for three distracting talkers trials. A silent interval ranging between 850 milliseconds and 1000 milliseconds (50 milliseconds steps) was added to each video to avoid the abrupt onset of the auditory stimuli; when talkers were visible (face and mask conditions), the still image of each talker was presented during this phase.

Crucially, only the audio tracks recorded without a mask were used when assembling the experimental trials. As anticipated, the rationale for this choice was to avoid a confound between the effect of masks on the audio signal and the effect of masks on visual processing of speech. To this aim, each single audio track was paired with each of the three visual conditions. For the face visible condition, the actual video and the corresponding audio track were used; for the name-only condition, the audio track was paired with a black screen with the name of the talker; for the face mask condition, the audio track was paired with the video of the same sentence uttered with the mask, carefully synchronized for onset. When pairing the audio tracks with the corresponding video with face mask, we made sure that the onset and offset of the audio and video tracks matched. To this end, we first compared the audio track obtained in the masked condition with the audio track without mask to detect any difference in synchronization between the two audios; when mismatches were detected, we either rerecord the sentence to fit better the timing of the track without mask (in cases of substantial discrepancies) or adjusted the video speed using the editing tools available in DaVinci Resolve (in cases of minimal discrepancies). Note that the video adjustments were always in the range of few milliseconds and could not be detected by participants.

Ultimately, we obtained three different versions of each sentence, which were used in three different versions of the test to be administered across participants. In this way, the phrases that in the first version were uttered by masked talkers, in the second one, they were uttered by completely visible talkers, and in the third, they were heard while seeing only the talkers' names.

Procedure

In the invitation, mail sent to each participant we presented the study and the instructions through a short video. We asked participants to use a computer and to wear earphones. In the first part of the experiment, participants gave their consent to take part in the study and filled in a form reporting their age, sex, and self-assessed hearing abilities. Participants were clearly instructed to pay attention only to the male talker in the videos. They were then asked to write the exact words pronounced by the target talker and then to judge their confidence in what they heard (*Quanto sei sicuro della tua esperienza di ascolto?*, which translates in: *How confident were you about your listening experience?*) and their listening effort (*Quanto è stato difficile ascoltare la voce?*, which translates in: *How difficult was it listening to the [male] voice?*). Both answers were scored from 0 to 9, with 9 indicating *maximal confidence* and *maximal effort*, respectively.

Before starting the test trials, participants performed six sample trials, showing all possible trial types (i.e., the six combinations between number of distracting talkers and listening conditions) to become familiar with the task and to adjust their listening volume to an optimal level. During the task, six target-only videos (2 Male Target \times 3 Listening Conditions) were evenly interspersed with test trials to monitor the participants' performance in the absence of noise (catch trials).

Analysis

Performance scores were computed as the average count of correct words reported in each trial (from 0 to 5), separately for each experimental condition. Similarly, listening effort and confidence were computed as the average rating expressed in each trial (from 0 to 9, where 0 = *no effort/low confidence* and 9 = *great effort/high confidence*), separately for each experimental condition. Confidence rating refers to a subjective measure of awareness, that stands for how much people think they have chosen their answers by guessing or not (Dienes, 2008; Dienes & Seth, 2010; Norman et al., 2011; Thiede et al., 2003). Listening effort is the subjective assessment of the perceived difficulty in completing the task (Peelle, 2018). Finally, we extrapolated also a measure of meta-cognitive monitoring, that is, the ability to judge successfully one's own cognitive processes while engaged in a task (Son & Schwartz, 2002), using listening confidence and performance measures in each trial. It is worth noting that metacognitive monitoring was not a task, but a value obtained from the relation between perceived listening confidence and actual performance. Good metacognitive monitoring occurs when, irrespective of performance, the relative confidence judgment reflects its outcome (Dienes, 2008; Dienes & Seth, 2010). When measures are in the same direction (high performance – high confidence or low performance – low confidence), the person shows good monitoring, when the two measures are inconsistent (high performance – low confidence or low performance – high confidence) the participant has poor monitoring abilities. We assessed for each participant the percentage of trials in which the listening confidence judgments produced trial-by-trial matched actual performance (low score/low listening confidence + high score/high listening confidence trials/total number of trials), separately for each listening condition. High performance scores corresponded to three or more correctly reported words, while a low performance score was less than two correctly reported words. High listening confidence score corresponded to confidence ratings of five or more points on the Likert-type scale, while low listening confidence was scored for ratings from 0 to 4.

All data were analyzed implementing linear mixed effects (LME) models in R studio (R Core Team, 2013) using the packages lme4 (Bates et al., 2015), car (Fox & Weisberg, 2020; Weisberg, 2019), and lmerTest (Kuznetsova et al., 2017; see also Ruta et al., 2019).

Table 1. Results of the LME Analysis.

	χ^2	df	p	% Δ mR ² _{GLMM}
Performance				
Number of distracting talkers	1551.80	1	<.001	39.73
Visual condition	235.26	2	<.001	6.00
Number of Distracting Talkers \times Visual Condition	112.65	2	<.001	2.86
Confidence				
Number of distracting talkers	751.31	1	<.001	24.87
Visual condition	137.73	2	<.001	4.54
Listening effort				
Number of distracting talkers	1243.83	1	<.001	32.52
Visual condition	218.18	2	<.001	5.69
Metacognitive monitoring				
Number of distracting talkers	98.65	1	<.001	24.58
Visual condition	9.99	2	.007	2.36

Following previous works (Amenta et al., 2020; Günther & Marelli, 2019), we computed effect size as a percentage increase of marginal R^2 GLMM obtained by adding each parameter to the null model (i.e., the model containing only the random structure) one at a time (Nakagawa & Schielzeth, 2013). Marginal R^2 GLMM values were computed using the MuMIn R package (Barton, 2018). GLMM = generalized linear mixed model.

Visual condition and number of talkers were used as fixed categorical predictors, and the intercept of each participant as random effect. First, we examined the model with all main effects and interactions and then reduced the model if appropriate using Bayesian Information Criterion as model fitting criterion. In the results section, we only reported the most appropriate model.

Results

To study the effect of listening condition (name, mask, and face) and number of distracting talkers (1, 3) on performance, we entered the number of correctly reported words in an LME with listening condition and number of distracting talkers as categorical fixed effects and the intercepts of the participants as random effect. The main effects of distracting talkers' number and visual condition emerged (see Table 1). As shown in the Figure 2A, seeing the talker's faces led to better performance compared with seeing faces with masks ($t = 12.30$, $p < .001$) or seeing names ($t = 13.36$, $p < .001$), which instead were comparable ($t = 1.06$, $p = .53$). This listening advantage was especially evident for the noisiest condition.

Similar LME models, applied separately to listening confidence ratings and to listening effort revealed the main effects of distracting talkers' number and listening condition (see Table 1). Seeing the talker's face resulted in higher listening confidence, compared with both the mask ($t = 8.91$, $p < .001$) and name conditions ($t = 11.07$, $p < .001$), which instead were comparable ($t = 2.17$, $p = .08$; see Figure 2B). Similarly, seeing the talker's faces reduced listening effort compared with the mask ($t = 11.70$, $p < .001$) and name conditions ($t = 13.66$, $p < .001$), which were comparable ($t = 1.96$, $p = 0.12$; see Figure 2C).

Finally, when applying the same LME model to the computed metacognitive monitoring measure, both main effects of distracting talkers' number and listening condition were again found (see Table 1). Participants' monitoring of their own trial-by-trial performance was better when they saw the talker's face, compared with when they saw the same person wearing a mask ($t = 3.089$, $p = .007$) suggesting that seeing a talker wearing a mask can hinder meta-cognitive judgments (Figure 2D).

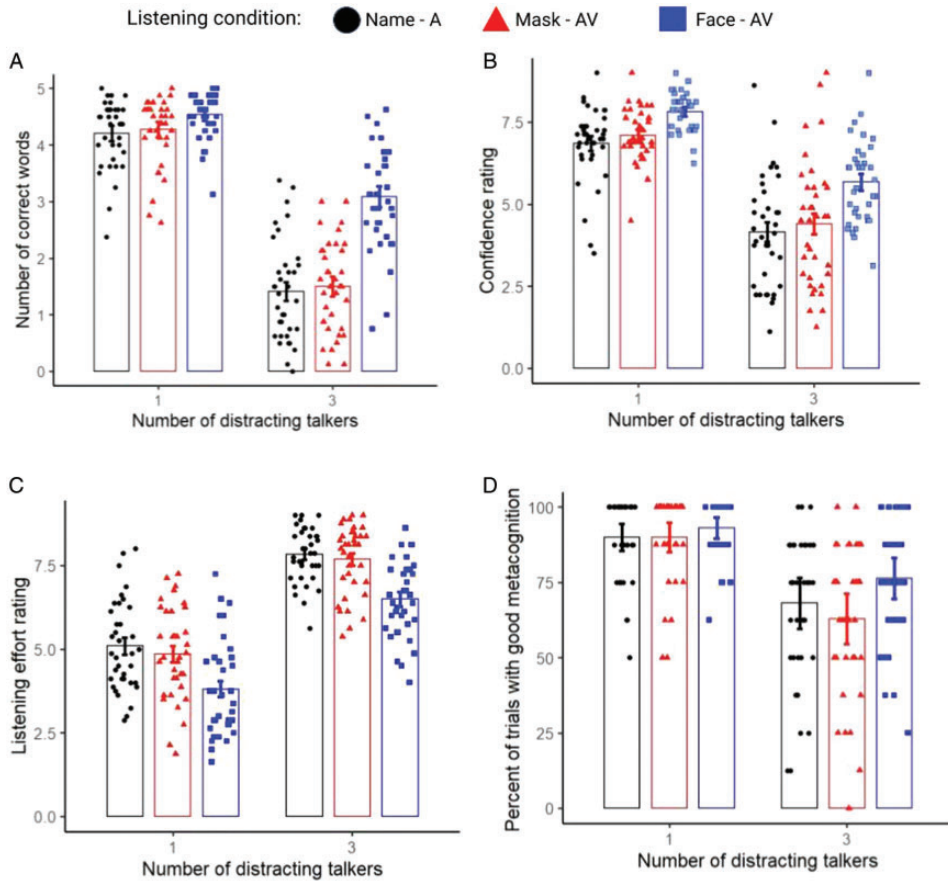


Figure 2. Effects of listening conditions and number of distracting talkers on (A) performance (number of correctly reported words); (B) perceived confidence in the heard words; (C) perceived listening effort; (D) meta-cognitive monitoring, as indexed by the percentage of trials in which confidence judgment and actual performance were concordant.

Discussion

In the present work, we mimicked a real multitalker video call to measure the impact of different visual conditions on speech comprehension in typical hearing participants. Results showed that hiding the talkers behind a black screen or concealing their lips via a face mask led to lower performance and lower listening confidence scores as well as increased listening effort. These differences between listening conditions suggest that the actual audio-visual benefit coming from vision relies on lip reading and demonstrate the impact of face masks on speech comprehension. Understanding a talker wearing a face mask in noise was, in our study, comparable to not seeing him or her at all. Importantly, these findings emerged in a context in which we disentangled the impact of visual information related to wearing a mask from the voice distortions generated by the mask. In this way, our results can be interpreted as the consequences of altering or removing visual information from lip movements in speech processing.

Our visual manipulation also impacted on the ability to successfully judge one's own cognitive processes while engaged in a task, namely, meta-cognitive monitoring. Face

masks reduced meta-cognitive monitoring abilities. In this condition, participants' listening confidence about their performance was less consistent with their objective performance (e.g., they could be confident about their performance, when in fact their speech comprehension was poor, or vice versa). This result is in line with previous work concerning the effect of face masks on confidence in reading emotions (Carbon, 2020), which found lower confidence and accuracy scores in recognizing expressions displayed by faces wearing surgical masks. This result supports the idea that hiding the lower part of a face undermines the efficacy of a conversation not only linguistically but also from a nonverbal point of view. While this result merits further investigation, it may suggest that when interacting with people wearing a mask, we not only feel less confident about our listening experience overall, but we are also less capable of monitoring whether we understood the message correctly or not. In addition, the confusion they generate on emotional reading of face expressions could further contribute to lowering the efficacy of our everyday life communications, preventing us from reconstructing the emotional tone of a conversation, which could partially contribute to better speech comprehension. This novel result is particularly interesting because compensatory strategies (e.g., asking our conversational partner to speak slower or in a louder voice) are typically triggered by adequate meta-cognitive monitoring of the success of the communication exchange (Boldt & Gilbert, 2019).

In June 2020, the World Health Organization warned about the potential risks and harms of face masks on daily communications. As evidenced by this study, when a talker wears a face mask the listening effort increases, while performance and confidence in what we listen decrease (see also Coniam, 2005; Llamas et al., 2009; Saunders et al., 2020). This could potentially result in stress and misunderstandings during communications, and even lead to risky behaviors, such as pulling down face masks or reducing social distancing while trying to understand each other better. In this study, we intentionally focused on a population of young adults, native speakers of Italian (the language used in the experiment), who reported no hearing difficulties. This is because we reasoned that any effect observed in this sample could only be exacerbated in populations that experience difficulties with language and communication. These populations include hearing children developing their L1, for whom the observation of adults' mouths can play a key role in an educational context (Spitzer, 2020); hearing children and adults learning a new language (L2); adults and aging people with normal hearing but sensitive to noisy contexts (Tremblay et al., 2015); and obviously all the populations with hearing loss or profound deafness. We believe it is a social priority to extend research on the effects of face masks on communication as well as other aspects of interpersonal perception (such as emotional processing or personal identity identification: Carbon, 2020) to all these populations.

The question arises then of how we can combine safe behavior and effective communication. One approach is to consider the introduction of transparent masks on a large scale. At the moment, they are only used in few medical settings (e.g., in the United Kingdom; Action on Hearing Loss, 2020), but they are gaining increasing attention among the hearing-impaired community (Taylor-Coleman, 2020). Even though this solution may seem the best way to reinstate lip reading into verbal communication, the current generation of transparent masks have several limitations. On the one hand, their materials impact greatly on the high frequencies of the human voice (Corey et al., 2020) affecting consonant perception (Divenyi et al., 2005; Roth et al., 2011). On the other hand, transparent masks are difficult to find because there is only a limited number of producers (Chodosh et al., 2020). Finally, in many countries, these devices are not approved by health authorities.

To conclude, our findings provide a clear example of the audio-visual nature of speech processing, and they emphasize the perceptual and meta-cognitive limitations that result from

occluding the face of our conversational partner. From the methodological point of view, our study represents a successful attempt to investigate audio-visual communication using an on-line task and simulating an ordinary listening context, such as the video call with a limited number of talkers. Clearly, when conducting hearing research online, a number of criteria need to be relaxed. It would be important to replicate and extend these observations running similar experimental protocols in a more controlled laboratory context in which individual hearing thresholds are also measured (unlike here). Moreover, it would also be important to increase the number of trials per participant (that said, our linear mixed-effect model approach to the analysis implies that we worked on a dataset of 1728 measures overall). Future experiments should also consider using audio tracks recorded both with and without masks, in order to objectively estimate the actual transmission loss produced by the masks and directly compare the effects of those distortions on speech comprehension. It is clear that such a comparison should necessarily exploit professional audio tools and accurate measures, only obtainable in a laboratory context. Nonetheless, our results agree with a vast literature on the multisensory contributions to speech perception and already provide support to recent petitions that pressured the main video conferencing platforms to offer real-time speech-to-text captioning (Chodosh et al., 2020). Most importantly, our findings indicate that audio-visual communication should be pursued even in the case of the health constraints imposed by a world pandemic. This is necessary for everyone, but especially for those individuals for whom face masks could become a severe obstacle to social inclusion.

Acknowledgements

The authors are grateful to Flavio Pizzorno and Samuele Conti who volunteered to serve as male talkers for our stimuli. We thank Catherine Riley and Nicholas Pedrazzoli for improving English in a previous version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: C. V. was supported by a grant of the Università Italo-Francese/ Université Franco-Italienne, the Zegna Founder's Scholarship and Associazione Amici di Claudio Demattè. F. P. was supported by a grant of the Agence Nationale de la Recherche (ANR-16-CE17-0016, VIRTUALHEARING3D, France), by a prize of the Foundation Medisite (France), by the Neurodis Foundation (France) and by a grant from the Italian Ministry for Research and University (MUR, PRIN 20177894ZH).

ORCID iD

Elena Giovanelli  <https://orcid.org/0000-0001-7753-632X>

References

- Action on Hearing Loss. (2020). *Coronavirus position statement*. <https://actiononhearingloss.org.uk/about-us/research-and-policy/coronavirus-policy-statement/>
- Amenta, S., Artesini, L., Musola, D., Frau, G. N., Vespignani, F., & Pavani, F. (2020). Probing language processing in cochlear implant users with visual word recognition: Effects of lexical and orthographic word properties. *Language, Cognition and Neuroscience*, 36(2), 1–12. <https://doi.org/10.1080/23273798.2020.1804600>

- Atcherson, S. R., Mendel, L. L., Baltimore, W. J., Patro, C., Lee, S., Pousson, M., & Spann, M. J. (2017). The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss. *Journal of American Academy of Audiology*, 28(1), 058–067. <https://doi.org/10.3766/jaaa.15151>
- Barton, K. (2018). *MuMIn: Multi-model inference* [Computer software manual]. <https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion*, 17(7), 1107–91119. <https://doi.org/10.1037/emo0000283>
- Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research*, 4, 45. <https://doi.org/10.1186/s41235-019-0195-y>
- Carbon, C. C. (2020). Wearing face masks strongly confuses counterparts in reading emotions. *Frontiers in Psychology*, 11, 2526. <https://doi.org/10.3389/fpsyg.2020.566886>
- Chodosh, J., Freedman, M. L., Weinstein, B. E., & Blustein, J. (2020). Face masks can be devastating for people with hearing loss. *British Medical Journal*, 370, m2683. <https://doi.org/10.1136/bmj.m2683>
- Coniam, D. (2005). The impact of wearing a face mask in a high-stakes oral examination: An exploratory post-SARS study in Hong Kong. *Language Assessment Quarterly*, 2(4), 235–261. https://doi.org/10.1207/s15434311laq0204_1
- Corey, R. M., Jones, U., & Singer, A. C. (2020). Acoustic effects of medical, cloth, and transparent face masks on speech signals. *The Journal of the Acoustical Society of America*, 148(4), 2371–2375. <https://doi.org/10.1121/10.0002279>
- Déaz, C. & Pedrero, A. (2005). The reverberation time of furnished rooms in dwellings. *Applied Acoustics* 66(8), 945–956. <https://doi.org/10.1016/j.apacoust.2004.12.002>
- Dienes, Z. (2008). Subjective measures of unconscious knowledge. *Progress in Brain Research*, 168, 49–269. [https://doi.org/10.1016/s0079-6123\(07\)68005-4](https://doi.org/10.1016/s0079-6123(07)68005-4)
- Dienes, Z., & Seth, A. (2010). Gambling on the unconscious: A comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Consciousness and Cognition*, 19(2), 674–681. <https://doi.org/10.1016/j.concog.2009.09.009>
- Divenyi, P. L., Stark, P. B., & Haupt, K. M. (2005). Decline of speech understanding and auditory thresholds in the elderly. *Journal of the Acoustic Society of America*, 118, 1089–1100. <https://doi.org/10.1121/1.1953207>
- Fox, J., & Weisberg, S. (2020). *Using car and effects functions in other functions*. <http://brieger.esalq.usp.br/CRAN/web/packages/car/vignettes/embedding.pdf>
- Goldin, A., Weinstein, B., & Shiman, N. (2020). Speech blocked by surgical masks becomes a more important issue in the era of COVID-19. *Hearing Review*, 27(5), 8–9. <https://www.hearingreview.com>
- Günther, F., & Marelli, M. (2019). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(10), 1872. <https://doi.org/10.1037/xlm0000677>
- Hampton, T., Crunkhorn, R., Lowe, N., Bhat, J., Hogg, E., Afifi, W., De, S., Street, I., Krishnan, M., Clarke, R., Dasgupta, S., Ratnayake, S. & Sharma, S. (2020). The negative impact of wearing personal protective equipment on communication during coronavirus disease 2019. *The Journal of Laryngology & Otology*, 134(7), 577–581. <https://doi.org/10.1017/S0022215120001437>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Llamas, C., Harrison, P., Donnelly, D., & Watt, D. (2009). Effects of different types of face coverings on speech acoustics and intelligibility. *York Papers in Linguistic Series*, 2(9), 80–104. <https://pure.york.ac.uk/portal/en/>
- Mendel, L. L., Gardino, J. A., & Atcherson, S. R. (2008). Speech understanding using surgical masks: A problem in health care? *Journal of the American Academy of Audiology*, 19, 686–695. <https://doi.org/10.3766/jaaa.19.9.4>

- Nakagawa, S., & Schielzeth, H. (2013) A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Norman, E., Price, M. C., & Jones, E. (2011). Measuring strategic control in artificial grammar learning. *Consciousness and Cognition*, 20(4), 1920–1929. <https://doi.org/10.1016/j.concog.2011.07.008>
- Peelle, J. E. (2018). Listening effort. *Ear and Hearing*, 39(2), 204–214. <https://doi.org/10.1097/aud.0000000000000494>
- Puglisi, G. E., Warzybok, A., Hochmuth, S., Visentin, C., Astolfi, A., Prodi, N., & Kollmeier, B. (2015) An Italian matrix sentence test for the evaluation of speech intelligibility in noise. *International Journal of Audiology*, 54, 44–50. <https://doi.org/10.3109/14992027.2015.1061709>
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Rosenblum, L. D., Johnson, J. A., & Saldaña, H. M. (1996). Point-light facial displays enhance comprehension of speech in noise. *Journal of Speech, Language, and Hearing Research*, 39(6), 1159–1170. <https://doi.org/10.1044/jshr.3906.1159>
- Roth, T. N., Hanebuth, D., & Probst, R. (2011). Prevalence of age-related hearing loss in Europe: A review. *European Archives of Oto-Rhino-Laryngology*, 268(8), 1101–1107. <https://doi.org/10.1007/s00405-011-1597-8>
- Ruta, N., Burleigh, A., & Pepperell, R. (2019). *Does visual space correspond to imagined space?* https://repository.cardiffmet.ac.uk/bitstream/handle/10369/10782/ECVP_2019NR11.pdf?sequence=1&isAllowed=y
- Saunders, G. H., Jackson, I. R., & Visram, A. S. (2020). Impacts of face coverings on communication: An indirect impact of COVID-19. *International Journal of Audiology*, 1–12. <https://doi.org/10.1080/14992027.2020.1851401>
- Son, L. K., & Schwartz, B. L. (2002). The relation between metacognitive monitoring and control. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 15–38). Cambridge University Press. <https://doi.org/10.1017/CBO9780511489976.003>
- Spitzer, M. (2020). Masked education? The benefits and burdens of wearing face masks in schools during the current Corona pandemic. *Trends in Neuroscience and Education*, 20, 100138. <https://doi.org/10.1016/j.tine.2020.100138>
- Sumbly, W. H., & Pollack I. (1954) Visual contribution to speech intelligibility in noise. *The Journal of the Acoustic Society of America*, 26, 212. <https://doi.org/10.1121/1.1907309>
- Taylor-Coleman, J. (2020). Coronavirus: Call for clear face masks to be “the norm.” *BBC News*. <https://www.bbc.com/news/world-52764355>
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66–73. <https://doi.org/10.1037/0022-0663.95.1.66>
- Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E. K., Klein, R., Levy, S., Tweed, T. S., & Cruickshanks, K. J. (2015). Self-reported hearing difficulties among adults with normal audiograms. *Ear and Hearing*, 36(6), e290–e299. <https://doi.org/10.1097/aud.0000000000000195>
- Weisberg, F. J. S. (2019). *An R companion to applied regression* (3rd ed.). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- World Health Organization. (2020). *Advice on the use of masks in the context of COVID-19: Interim guidance*. [https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-\(2019-ncov\)-outbreak](https://www.who.int/publications/i/item/advice-on-the-use-of-masks-in-the-community-during-home-care-and-in-healthcare-settings-in-the-context-of-the-novel-coronavirus-(2019-ncov)-outbreak)

How to cite this article

Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., & Pavani, F. (2021). Unmasking the difficulty of listening to talkers with masks. *i-Perception*, 12(2), 1–11. <https://doi.org/10.1177/2041669521998393>