DOI: 10.61782/fa.2023.0293

forum acusticum 2023

PSYCHOACOUSTIC STUDY OF THE IMPACT OF SANITARY DEVICES ON SPEECH INTELLIGIBILITY

Anne BOUVET[1*] **Laurianne DELCOR**[1]

[1] Capgemini Engineering, Research
Innovation Direction, 4 Avenue Didier Daurat, 31700 Blagnac, France

ABSTRACT

It has often been reported, during the COVID-19 pandemic, that wearing a face mask impacts speech understanding during a conversation. Different research [1–3] has been realized with speech record analysis or listening tests to understand the phenomenon, but none tried to do both simultaneously. The study focuses on the subjective and objective analysis of the influence of sanitary devices on speech listening intelligibility. The first goal is to understand if there is a real perception of the loss of intelligibility with a listening test, and at the same time, analyse the same signals to attest if and where this feeling comes from. The results show that the subjective intelligibility is only slightly degraded with face masks and a bit more with the protective barrier. However, the signal processing shows that none of the sound level nor the signal-to-noise ratio decrease with the devices. However, a significant loss of frequency is observed beyond 1 kHz.This phenomenon can explain why it is more difficult to understand or distinguish between some phonemes. Beyond this study, the loss of intelligibility due to sanitary devices must also be affected by the loss of the visual component of speech.

Keywords: *psychoacoustic, speech, intelligibility, sanitary devices.*

*__Corresponding author__: *anne.bouvet@capgemini.com.*

1. INTRODUCTION

The COVID-19 pandemic has imposed new practices to protect people from contamination. In particular, by containing the propagation of droplets, cause by breathing and speaking, using to masks and transparent panels.
These protections prevent the propagation of the air flow produced, especially during the production of speech. This can create problems of understanding between people. Objective studies [1–3], have shown that masks, regardless of their design (fabric, surgical, FFP2, transparent...) attenuate the sound signal over frequencies greater than 1kHz. A review of the literature [4] studied the effects of masks on children's attention and found a study that objectively measured a sound attenuation of 3 to 12 dB depending on the mask between 2 kHz and 7 kHz. Another large study [5], involving 460 respondents, to an online questionnaire showed that masks had a negative impact on hearing and comprehension.
Thus, the current literature mostly proposes objective studies, based on sound levels, or subjective questionnaires, without listening to speech signals, to describe the degradation of intelligibility related to masks. What about the understanding of a sound message from a perceptual point of view, by listening to the signals? In addition, besides masks, do glass or plexiglass panels degrade intelligibility? This study aims at answering these questions by using the experimental protocol and the resulting voice recordings and listening.
For this purpose the experiment is conducted in four phases. 1) Recording speech signals with different sanitary devices (masks, glass panel), 2) extract objective parameters of temporal and frequency intelligibility, 3) to carry out a second experiment this time of listening and

evaluation of the recorded speech signals by a panel of participants and finally 4) to analyze the obtained results and compare them with the objective results.

2. PROTOCOL

The objective of the study is to compare all the sanitary protection devices encountered in normal practice. Therefore, 3 types of masks are used for the experiment: surgical (Ergum medical), FFP2 (ANNEW) and fabric (unbranded, cotton, single layer). As the main difficulties of comprehension are encountered at the counters, the influence of a glass panel (ForMat - width 75 cm, height 50 cm, thickness 0.4 cm) is also studied.

2.1 Recording experiment

Recordings are performed with microphone Blue Yeti (sensibility measured 4.5 mV/Pa, in cardioid mode), placed at 1m from the participant as shown in Figure 1.
A total of 8 people were recorded (3 women and 5 men).

Figure 1. Recording experiment installation.

Each person uttered 7 french sentences with the 8 sanitary protection configurations:

- WM: No mask, no panel
- SM: Surgical mask, without panel
- FFM: FFP2 mask, without panel
- FM: Fabric mask, without panel
- PWM: Without mask, with glass panel
- PSM: Surgical mask, with glass panel
- PFFM: FFP2 mask, with glass panel
- PFM: Fabric mask, with glass panel

Among the recorded sentences, 3 were selected :

- Bonjour, comment vas-tu ?
  Hello, how are you?
- Je suis allée voir un film au cinéma.
  I went to see a movie at the cinema.
- Take-off est à quatre cents mètres de Synapse.
  Take-off is four hundred meters from Synapse.

The selection of these 3 sentences was made by principal component analysis (PCA) using 70 calculated level and spectral parameters.

2.2 Listening experiment

For the listening experiment shown in Figure 2, the sentences were played with the headphones Bose QuietComfort II (maximum noise reduction) with a constant sound level for all participants. The study is based on the absolute evaluation with reference of 168 stimuli (3 sentences $\times$ 8 voices $\times$ 7 configurations). It consists in comparing the intelligibility of a voice equipped with a sanitary protection device, with a reference signal of the same voice without device. The 36 tested participants (21 men and 15 women), scored the intelligibility degradation of 165 signal pairs on a scale of 0 to 100. 0 corresponds to no degradation at all and 100 to extreme degradation.

Figure 2. Listening experiment installation.

3. RESULTS

The subjective and objective analysis of the intelligibility degradation will allow the understanding of which sanitary device configurations penalize the perceived speech intelligibility.

3.1 Subjective results

3.1.1 Normality of the data

The Shapiro test [6] gives a normal distribution for 52% of the stimuli (comparison). This is explained by the fact that the rating scale of 0 corresponds to "not at all degraded", which gives a semi-normal distribution. By performing the test with random multiplication of each value by +1 or -1, the evaluations are distributed around 0. The Shapiro tests then indicate a normal distribution.

3.1.2 Variance in intelligibility degradation ratings

A three-factor repeated measures ANOVA [6] was performed on a model linking the variance of the intelligibility degradation evaluations to the parameters of the voice recordings of the stimuli (mask configuration, with or without glass panel), the person recorded and the sentence pronounced. The results show that each of these three parameters significantly influences the evaluations of intelligibility degradation, as their p-values are less than 0.05 (see Table 1). The F-values show that the variations in intelligibility degradation ratings are larger as a function of the different configurations than as a function of the voice or sentence. This implies that the configuration factor has a stronger impact on the intelligibility degradation evaluations than the voice and sentence factors.

Table 1. ANOVA results for the full model.

Model : eval. config. + voice + sentence, random = 1 |participant		
Facteur	F-Value	p-value
Config	**1530**	< 0.001
Voice	32	< 0.001
Sentence	43	< 0.001

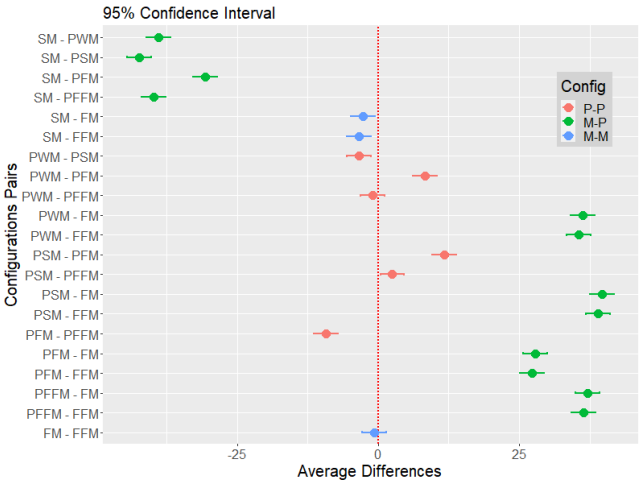

Figure 3. Differences in average intelligibility degradation ratings as a function of the sanitary device configurations used for the stimulus recordings. *WM: Without mask, SM: Surgical mask, FFM: FFP2 mask,FM: Fabric mask, P: Glass panel*

3.1.3 ANOVA model with configuration only

The study now seeks to highlight the impact of different configurations (different sanitary devices) on the intelligibility degradation evaluations.
A one-factor repeated measures ANOVA is performed, and the results [F-value = 1457, p-value <0.001] confirm those from the full model: configuration has a significant influence on intelligibility impairment ratings . This time, the F-value indicates that the differences in intelligibility impairment ratings are high. A post-hoc Tukey test is performed to understand more precisely which configurations impact the intelligibility degradation evaluations. 7 configurations were tested, thus it is possible to compare 21 pairs of configurations two by two. The results of the analysis show that only 2 pairs of configurations have non-significantly different intelligibility degradation ratings (PFM - PSM and FM - FFM). Figure 3 shows these results. The majority of the confidence intervals of the differences in averages are far enough from 0. For the signals with significant differences in averages, the associated values vary from 10 to 40 points in absolute value, with some pairs of configurations for which the differences are close to 0 while still having significant differences in averages. The colorization of Figure 3 highlights the pairs of configurations in which both config-

urations were recorded with the glass panel in red (P-P: Panel - Panel), the pairs for which one configuration was recorded with panel and one with only a mask in green (M-P: Mask - Panel) and the pairs for which both configurations were recorded with only the masks in blue (M-M: Mask - Mask). Thus, it can be seen that the pairs of configurations with the most different average intelligibility degradation ratings are those containing stimuli recorded with a panel and recorded with only any mask (green). This implies that the perception of intelligibility degradation varies greatly depending on the presence or absence of the glass panel between the speaker and the listener.
As a conclusion, the presence of the glass panel seems to strongly degrade intelligibility compared to speech recordings performed without the panel. Intelligibility is also degraded, to a smaller extent, when the person speaks behind a panel and with a mask compared to a configuration behind a panel without a mask. Finally, in this study, the type of mask did not seem to significantly influence the degradation of intelligibility.

3.2 Objective results

In this study, different parameters are calculated and plotted on the 3 sentences of each voice recorded:

- Bonjour, comment vas-tu ?
  Hello, how are you?
- Je suis allée voir un film au cinéma.
  I went to see a movie at the cinema.
- Take-off est à quatre cents mètres de Synapse.
  Take-off is four hundred meters from Synapse.

3.2.1 Sound level parameters

The objective analysis is performed on the global sound level (Lg), the loudness and the signal to noise ratio (SNR) of each sentences.
The results show no signification difference between only masks configurations but present a loss of level with the glass panel as shown in Table 2

Table 2. Results of level, loudness and SNR loss with panel configurations.
WM: Without mask, SM: Surgical mask, FFM: FFP2 mask,FM: Fabric mask, P: Glass panel

	PWM	PSM	PFFM	PFM
Lg [dB]	-1.9	-1.8	-1.6	-2.6
Loudness [-]	-2.4	-.6	-2.5	-3.3
SNR [dB]	-2.4	-1.9	-1.8	-2.9

4. POWER SPECTRAL DENSITY

The analysis of the spectral density (PSD) provides a very fine vision of the frequency distribution of signals. The results are presented for the third sentence and for two representative voices of a man (Figure 4) and a woman (Figure 5) For all signals, the loss is clearly visible at frequencies above 1 kHz. The comparison of the general average of the difference, between the spectrum of the signals with sanitary device and without, gives a more marked impact with : SM = - 1.9 dB, FFM = - 2.4 dB, FM = - 2.6 dB, PWM = - 6.7 dB, PSM = - 8.5 dB, PFFM = - 8.9 dB, PFM = - 8.7 dB.
In conclusion, this loss of level in the high frequencies can significantly influence intelligibility. Indeed, speech and especially consonants which are mainly discerned by ears thanks to the first two formants (resonance frequency of the vocal tract). It can be suppose that, due to the 2^{nd} formant is located, for the majority of the vowels above 1 kHz. Thus vowels having the same 1^{st} formant will be difficult to differentiate if their 2^{nd} is attenuated.

5. CONCLUSION & DISCUSSION

The aim of the research was to analyze and quantify subjectively and objectively the causes of intelligibility degradation.
Regarding the perceived degradation the order is from least to most degraded:
Surgical mask $<$ **FFP2 mask** $<$ **Fabric mask** $\ll$ **Panel without mask** $\ll$ **Panel with surgical mask** $<$ **Panel with fabric mask** $<$ **Panel with FFP2 mask**.
We conclude that the glass panel is the device with the greatest impact on intelligibility impairment. However, all the listeners in the subjective study often reported that the message remained clearly understandable for all the samples presented.
The objective analysis of the signals confirms these results. Indeed, the signal-to-noise ratio, the noise level and the loudness do not show a significant degradation, especially concerning the masks, which confirms that the speech message remains clearly understandable. However, the spectral analysis shows that it is the high frequencies (above 1kHz) that are mainly impacted and diminished.

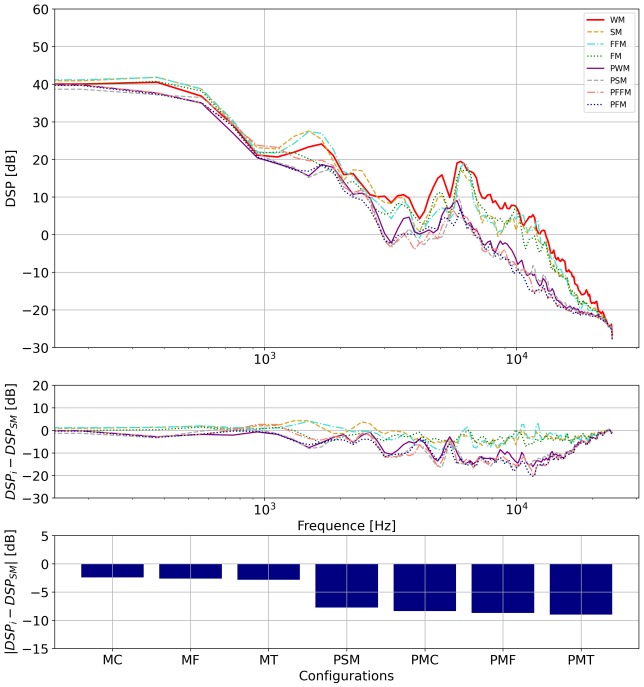

Figure 4. PSD, differences between configurations with and without sanitary device and average differences for a man's voice.
WM: Without mask, SM: Surgical mask, FFM: FFP2 mask,FM: Fabric mask, P: Glass panel

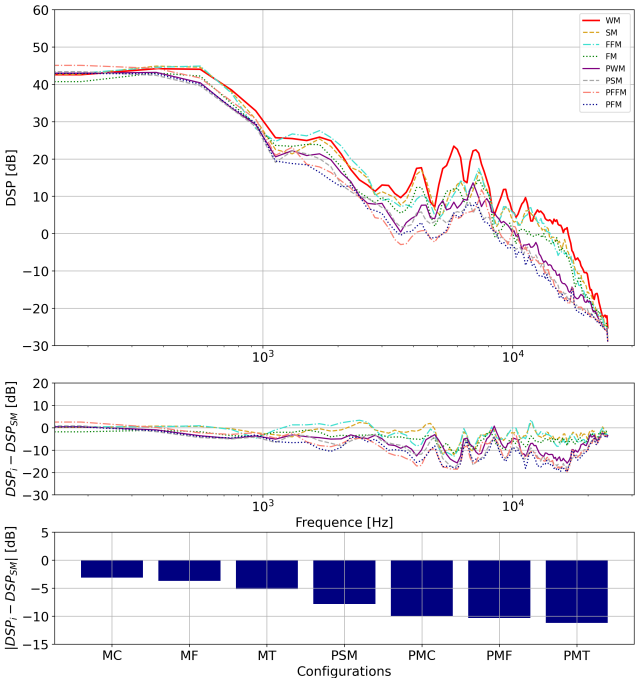

Figure 5. PSD, differences between configurations with and without sanitary device and average differences for a woman's voice.
WM: Without mask, SM: Surgical mask, FFM: FFP2 mask,FM: Fabric mask, P: Glass panel

The observed results are in phase with the litterature. The sound signal attenuation from 1 kHz was also observed with masks in [1–3]. Further experiments are currently performed to refine the results and the understanding of intelligibility degradation as a function of sanitary devices.

The approach in [1] aims to minimize measurement bias associated with speaker reproducibility. To achieve this, the sentences were recited by speakers wearing sanitary devices. This ensures the reproducibility of the voice, so that any observed acoustic variations are solely attributed to the sanitary devices. However, it is important to note that the approach does not consider potential pronunciation difficulties arising from restricted airflow caused by the masks. The study in [7] confirmed that without background noise, masks have no real impact on intelligibility, however, adding noise impacted the intelligibility as well as the speech production. Finally, it might be interesting to measure intelligibility without using a reference signal in order to not bias the participants' responses by giving them the content of the vocal message. A final approach explored in [8] combines both acoustic and visual context of speech. The study reveals that visual degradation primarily affects children's ability to understand speech.

6. ACKNOWLEDGMENTS

This study was conducted within the project of Research and Innovation OPTIMIND, and could succeed thanks to the participation of Capgemini Engineering TEC employees.

7. REFERENCES

[1] S. R. Atcherson, B. R. McDowell, and M. P. Howard, "Acoustic effects of non-transparent and transparent face coverings," *The Journal of the Acoustical Society of America*, vol. 149, pp. 2249–2254, apr 2021.

[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, pp. 2371–2375, oct 2020.

[3] J. C. Toscano and C. M. Toscano, "Effects of face masks on speech recognition in multi-talker babble noise," *PLOS ONE*, vol. 16, p. e0246842, feb 2021.

[4] M. Nobrega, R. Opice, M. M. Lauletta, and C. A. Nobrega, "How face masks can affect school performance," *International Journal of Pediatric Otorhinolaryngology*, vol. 138, p. 110328, nov 2020.

[5] G. H. Saunders, I. R. Jackson, and A. S. Visram, "Impacts of face coverings on communication: an indirect impact of COVID-19," *International Journal of Audiology*, vol. 60, pp. 495–506, nov 2020.

[6] Z. F. Andy Field, Jeremy Miles, *Discovering Statistics Using R*, vol. 81. International Statistical Review, apr 2014.

[7] V. A. Brown, K. J. V. 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, jul 2021.

[8] J. Schwarz, K. K. Li, J. H. Sim, Y. Zhang, E. Buchanan-Worster, B. Post, J. L. Gibson, and K. McDougall, "Semantic cues modulate children's and adults' processing of audio-visual face mask speech," *Frontiers in Psychology*, vol. 13, jul 2022.