e-ISSN 2499-1562

Annali di Ca' Foscari. Serie occidentale
Vol. 55 – Settembre 2021

Effect of Different Face Masks on Speech and Singing: Self-Perception and Acoustic Analysis

Claire Pillot-Loiseau
Université Sorbonne Nouvelle, France

Bernard Harmegnies
University of Mons, Belgium

Abstract The aim of this preliminary study is to better understand the effects of transparent, surgical, cloth, KN95 (FFP2), and singer's face masks on speech and singing in French. A survey gathered self-perception, and a local and global acoustic analysis of conversational, loud spoken and sung productions by the same individual were conducted. According to the 303 subjects surveyed, plosive consonants seem to be produced with the greatest difficulty. Consonants requiring lip involvement seem to be the most affected. The transparent and KN95 (FFP2) masks attenuate the intensity of all the consonants and spoken utterances as a whole, unlike the singer's mask.

Keywords Phonetics. Face masks. Consonants. Speech. Singing. Self-perception.

Edizioni Ca'Foscari

Peer review

Submitted 2021-08-01
Accepted 2021-09-07
Published 2021-10-11

Open access

Citation Pillot-Loiseau, C.; Harmegnies, B. (2021). "Effect of Different Face Masks on Speech and Singing: Self-Perception and Acoustic Analysis". *Annali di Ca' Foscari. Serie occidentale*, 55, 9-28.

DOI 10.30687/AnnOc/2499-1562/2021/09/012

1 Introduction

The current global pandemic requires all individuals to protect themselves from COVID-19 by wearing a face mask (Howard et al. 2021); its effects on everyday communication have become a major and lasting societal concern, especially at a time when new variants appear to decrease the protective effect of vaccines.

However, although there is a broad consensus on its health effectiveness, it is often criticised for various discomforts it is supposed to generate. Apart from general user complaints conveyed by mainstream media regarding the undesirable effects of these devices,[1] there is little scientific literature on its effects on speech and singing as expressed by subjects. According to the study by Ribeiro et al. (2020), face masks increase users' perception of vocal effort, difficulty in speech intelligibility, auditory feedback and difficulty in coordinating speech and breathing. These perceived discomforts appear to be more cumbersome for people who wear face masks for professional and essential activities.

More objectively, wearing a mask primarily means preventing visual perception of the articulatory gestures that accompany speech; these gestures play an essential role, especially in less ideal communication contexts, for example, in noise (Erber 1969), or when hearing loss is treated with cochlear implants (Stevenson et al. 2017). The need to visualise the individual's mouth while wearing a mask has thus led to the creation of masks with transparent windows (among others: Atcherson et al. 2017). Other devices allowing for better freedom of the mandible and lips, particularly important for singers, have emerged, such as the large mask or the singer's mask.[2] These types of masks are complemented by more frequently worn equipment such as surgical masks, fabric cloth masks of various materials, and KN95 (KN95 (FFP2)) masks.

It has been shown that the speech perception of an individual wearing a face mask is impaired, mainly in terms of intelligibility (Mendel, Gardino, Atcherson 2008; Palmiero et al. 2016; Atcherson et al. 2017; Bottalico et al. 2020), the type of speech (informal, clear or emotional: Cohn, Pycha, Zellou 2021), but also in terms of consonant identification in English (Fecher, Watt 2013).

At the acoustic level, some works have shown that masks filter the transmission of entire speech utterances (Corey, Jones, Singer

1 E.g., https://www.sciencesetavenir.fr/sante/port-du-masque-comment-gerer-les-effets-indesirables_144282.

2 See for example: https://www.cbc.ca/radio/asithappens/as-it-happens-friday-edition-1.5695035/choir-director-invents-performer-s-face-mask-for-safe-singing-1.5695376 or https://www.mymusicfolders.com/product/resonance-singers-mask-with-disposable-biofilters/ or https://www.legrandmasque.fr/.

e-ISSN 2499-1562

2020; Magee et al. 2020; Nguyen et al. 2021; Fiorella et al. 2021) by attenuating frequencies above 1kHz. This attenuation occurs in different proportions depending on the type of mask, to the detriment of transparent masks and in favour of surgical masks (among others: Corey, Jones, Singer 2020; Magee et al. 2020). In singing, only one study mentions the acoustic and perceptual effects of a gaiter, a disposable surgical mask and an N95 mask, as well as an acoustic foam (Oren et al. 2021). This study was based on a recorded excerpt of the *Star-Spangled Banner* produced by a soprano and played through a miniature speaker placed in the mouth of a masked mannequin, and not on an actual recorded production by a singer wearing the mask. Oren et al. (2021) showed that the acoustic energy around the singer's formant was reduced when using surgical masks and N95 masks. On the segmental level, Fecher (2014) demonstrated that in English, several facial covering devices such as motorbike helmets, bonnets, mouth plasters, niqabs, scarves or rubber masks, could alter the acoustic composition of unvoiced plosive and fricative consonants.

The global (whole utterance) and local (segments, especially fricative and plosive consonants) acoustic effects of wearing anti-COVID-19 face masks has therefore been studied mainly in conversational speech and in English. To the best of our knowledge, no previous research has shown this effect 1° for voiced plosives and fricatives; 2° in loud speech and in classical soloist singing in French for the current facial devices. Other masks more specifically adapted to singers have yet to be employed. In addition, the discomfort experienced by mask users was not studied with regard to segments that were felt to be difficult to produce with the mask depending on the device worn.

The aim of this exploratory research is to better define the components of the nuisance caused by the mask from a subjective point of view (difficulties experienced by mask users) as well as from the objective point of view (acoustic distortions caused by the mask), both locally (plosive and fricative consonants) and globally (whole utterances with and without the mask), as a function of the task produced in French, the speaker, session and the type of mask. This case study is in line with the dynamics of Perturbation Theory (Sock 2001): the presence of a mask constitutes an external source of perturbation to the speech production system. It is assumed that this external disturbance leads to local and global acoustic changes, which vary according to the type of mask, the corpus, the session, and the speech task produced.

e-ISSN 2499-1562

2 Feelings of Participants Wearing a Mask

2.1 Method

A survey in French gathering the opinions of French-speaking mask users was distributed via Google Forms between November, 20, 2020 and February, 14, 2021. The survey consisted of 15 questions designed to collect information on the profiles of the participants (gender, nationality and languages spoken, profession; discipline, type of school and environment, average age of pupils, class size, teacher's schedule; extra-professional vocal activity) and on the types of masks they wear. Eleven other questions were then asked in order to find out the nature of the discomfort they might feel in connection with wearing the mask (from never to always: misunderstanding with interlocutors, absence of visual cues, constant repetition requested by interlocutors, reduction of contact with others, incomprehension of speech by interlocutors, incomprehension of interlocutors' speech, difficulty speaking, breathing; nature of the segments that are difficult to produce with the mask, other discomforts, solutions envisaged by them to resolve these difficulties).[3] This study focuses on the phonetic and acoustic repercussions of wearing a mask, we, therefore, limit ourselves to the analysis of the participants' responses regarding the effect of the mask on the segments produced, and other discomforts regarding the quality of the sound resulting from wearing these masks. The question about the production of different classes of segments with the mask was phrased as follows: "Which consonants and/or vowels do you have difficulty articulating due to wearing the mask?".

Because the literature mostly discusses the effect of the mask on consonant production and perception, we presented in the multiple-choice responses the consonants first, naming their class and orthographic character. Thus, participants had a choice from the following responses: Occlusive consonants ("p t k b d g"); Nasal vowels ("an, in on");[4] Nasal consonants ("m, n, gn");[5] Other: … (complete).

3 The complete questionnaire items can be seen at the following link: `https://docs.google.com/forms/d/e/1FAIpQLSdNQoroaySNmefW9F12lx0-36FEBztDqyxUFFDaMAf-hZrixbQ/viewform`.

4 I.e., respectively /ɑ̃/, /ɛ̃/, and /ɔ̃/.

5 I.e., respectively /m/, /n/, and /ɲ/.

e-ISSN 2499-1562

2.2 Results

303 French participants (262 women and 41 men) responded to the survey, including 216 teachers, 50 professional soloists and small ensemble singers (80% classical singers), 82 amateur singers with choral activity, and 171 all-purpose participants. Of these, 101 wore cloth masks, 175 wore surgical masks, 7 wore KN95 (FFP2) masks, 2 wore transparent masks and 18 alternated between the types of masks (most often cloth and surgical).

Figure 1 illustrates the participants' answers concerning difficulties in producing segments according to the type of mask worn. It shows that plosive consonants were the primary source of difficulty for 39-42% of the subjects wearing cloth and surgical masks. The two transparent mask wearers (not shown in figure 1) experienced difficulties in producing plosive and nasal consonants. Only 7 participants wore KN95 (FFP2) masks, so we must interpret the results for these individuals with caution. Cloth and surgical mask wearers also expressed difficulties articulating nasal consonants (24% and 25.5% respectively), as well as nasal vowels (18.5% and 10.5% respectively) and fricative consonants (1.5% and 2.9% respectively).

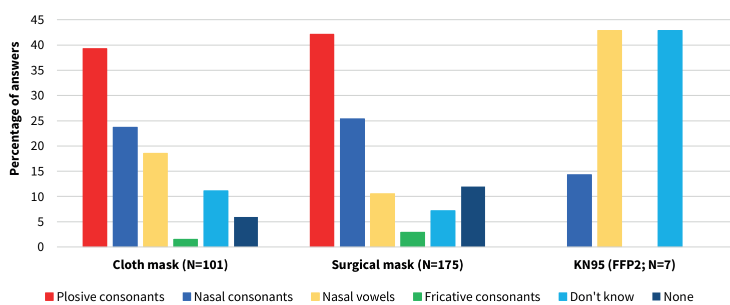

Figure 1 Segments judged by 303 users as difficult to produce with the mask, depending on the device worn

In addition, an analysis of the other complaints spontaneously written by the participants wearing the mask reveals a lack of vocal carrying power, a feeling that the sound of the voice 'cut-off' (especially in the case of the participants wearing the transparent mask) leading to constant repetition requested by their interlocutors; this according to 55% of the participants' answers.

As the segments that are difficult to produce are mainly plosives according to the answers of our participants, and as the acoustic effects of masks affect the high frequencies present in the production of fricatives, we first chose to determine the nature of local acoustic distortions in voiced and unvoiced French fricatives and plosives in a non-word production task in speech by the female speaker, as presented in the next section.

3 Production Experiment: Local and Global Acoustic Distortions

3.1 Method

3.1.1 Recordings and Corpus (One Female Speaker and One Male Speaker)

At a sampling rate of 44100Hz, and using an AKGC520 headset microphone placed 3 cm from the mouth, an Edirol UA-25 USB Audio Capture sound card and Audacity software, a 50-year-old French-native female speaker recorded in a quiet room 3 repetitions of the syllables /aCa/, where C= /p, t, k, b, d, g, f, s, ʃ, v, z, ʒ/, in order to study the local acoustic distortions of the masks in voiced and unvoiced French plosive and fricative consonants.

In the same recording session, the same person, a trained classical singer, also produced a spoken and a loud reading of the French version of the text *The North Wind and the Sun* (*La Bise et le Soleil*), as well as the French sung melody *Clair de Lune* by Fauré.

The female speaker recorded the whole corpus during a second session one month after the first, under the same conditions. For these two sessions, the productions were recorded without and with the following face masks: surgical, fabric (cloth), KN95 (FFP2), with transparent window, and two masks designed for singers (called 'large mask' and 'singer's mask').

Using a Sennheiser e840S microphone placed 35 cm from the mouth, an A-D Focusrite Scarlett 2i4 USB sound card and Praat software (Boersma, Weenink 1992-2021), a 64-year-old French-native male speaker recorded in a quiet room in a single session a spoken reading of *La Bise et le Soleil* without and with surgical, cloth, KN95 (FFP2), and transparent window masks.

3.1.2 Local and Global Acoustic Analysis Procedures

The consonants of the female speaker's productions from the /aCa/ syllables were then annotated on Praat. Using a Praat script, and mid-consonant, the spectral centres of gravity[6] (of friction for fricatives and of explosion for plosives, Fecher 2014), standard deviations,[7] skewness coefficients[8] (indicator of the (a)symmetry (overall slant) of the energy distribution relative to a Gaussian distribution where skewness = 0) and kurtosis[9] were extracted and calculated. Their long-term averaged spectra (which represents the logarithmic power spectral density as a function of frequency) were also calculated using Praat (number of frequency band: 100; width of each band: 100Hz; first band centred at 50Hz) in order to qualitatively visualise the nature of the acoustic distortions produced by each mask type.

The productions of whole utterances produced in two recording sessions (conversational, loud read text, and sung melody) by the female speaker, and the conversational text read in one session (for the male speaker) were then qualitatively processed by generating their long-term averaged spectra for each type of production, speaker and session. The SDDD indices (Standard Deviation of the Differences Distribution, Harmegnies 1988) and the Bravais-Pearson correlation coefficients between each long-term averaged spectrum of a production without a mask and that of the same production with a mask type, were then calculated using Praat. SDDD is a dissimilarity index: the higher it is, the more different the spectra are. The Bravais-Pearson correlation coefficient is an index of similarity: the higher it is, the more similar the spectra are.

Finally, a matched statistical comparison (paired two-tail t-test) was carried out between the latter two long-term averaged spectra, using the following procedure: in each long-term averaged spectrum with val-

6 "The center of gravity (CG) is the first spectral moment of the spectral distribution. It expresses the frequency at which the spectral energy is predominantly concentrated, and is thus related but not equal to the peak" (Fecher 2014, 87).

7 "The standard deviation is a measure of how distributed the energy is along the frequency axis. In other words, the standard deviation (SD) specifies the bandwidth of energy on either side of the mean (Jongman, Wayland, Wong 2000; Stuart-Smith, Timmins, Wrench 2003; Harrington 2010)" (quoted in Fecher 2014, 89).

8 "Skewness values (dimensionless) are positive when the intensity energy is primarily concentrated in low frequency bands (negative spectral tilt), and negative when the energy is predominantly found in higher frequencies (positive spectral tilt). A value of zero denotes a normal (Gaussian) distribution, i.e., no difference in energy around the CG (Harrington 2010)" (quoted in Fecher 2014, 91-2).

9 "An indicator of the 'peakedness' of the distribution, i.e. it expresses to what extent the spectral energy is concentrated in a peak relative to low and high frequencies (Jongman, Wayland, Wong 2000; Stuart-Smith, Timmins, Wrench 2003; Harrington 2010)" (quoted in Fecher 2014, 94).

ues over a frequency range of 0 to 10,000Hz, 100 points of spectral amplitude values every 100Hz were extracted using Praat software. Since the intensity gains were identical for the same individual regardless of the session during recordings, the 100 points of the long-term averaged spectra of the productions without mask were compared in parallel to the 100 points of the long-term averaged spectra of the productions with each type of mask, to see if the attenuation caused by each mask was significant compared to the spectral composition without mask. In addition, identical productions were compared between the first and second sessions in the female speaker. VassarStats[10] software was used for this statistical analysis ("t-test for correlated samples").

3.2 Results 1: Local Acoustic Distortions (Plosive and Fricative Consonants)

Figures 2 and 3 show the averaged spectra of the French plosives and fricatives /p t k b d g/ and /f v ʃ ʒ s z/. They show that the most important acoustic attenuations are caused by the KN95 (FFP2) and transparent masks: these energy attenuations are increasing as the frequencies rise for /p/ (KN95 (FFP2) mask) and /f/ (transparent mask). For the transparent mask, attenuations result in the presence of antiformants around 4000Hz, 8000Hz, and, to a lesser degree, 6000Hz for the consonants /f ʃ ʒ t s z k g/. For other types of masks, the consonants /p b f v/ show more acoustic distortions compared to the unmasked condition than other consonants, from 500Hz onwards.

In order to quantify these acoustic distortions, measurements of the spectral centre of gravity, standard deviation, asymmetry and kurtosis coefficients were carried out as shown in figure 4. Results show little change in parameters regardless of the type of mask for /s z/ and for /ʃ ʒ/ (except for the kurtosis coefficient for the latter consonant pair). /p b d f v/ show the most variable acoustic changes as a function of mask type for the spectral centre of gravity and its standard deviation. The transparent mask shows the most significant acoustic distortions compared to the unmasked condition, notably in the form of a uniformity of centre of gravity values (higher for /p b/, and lower for /t d k g f v/: the trend is reversed for these consonants concerning the asymmetry coefficient, less positive (equivalent to an energy concentration in slightly higher frequency bands) for bilabials, and more positive for /d k g f v/ (i.e. an increase in the energy concentration in lower frequencies, compared to the productions without mask). The same trends can be observed for the standard deviation and the kurtosis coefficient respectively (lower, i.e. less ener-

10 http://vassarstats.net/ by © Richard Lowry 1998-2021.

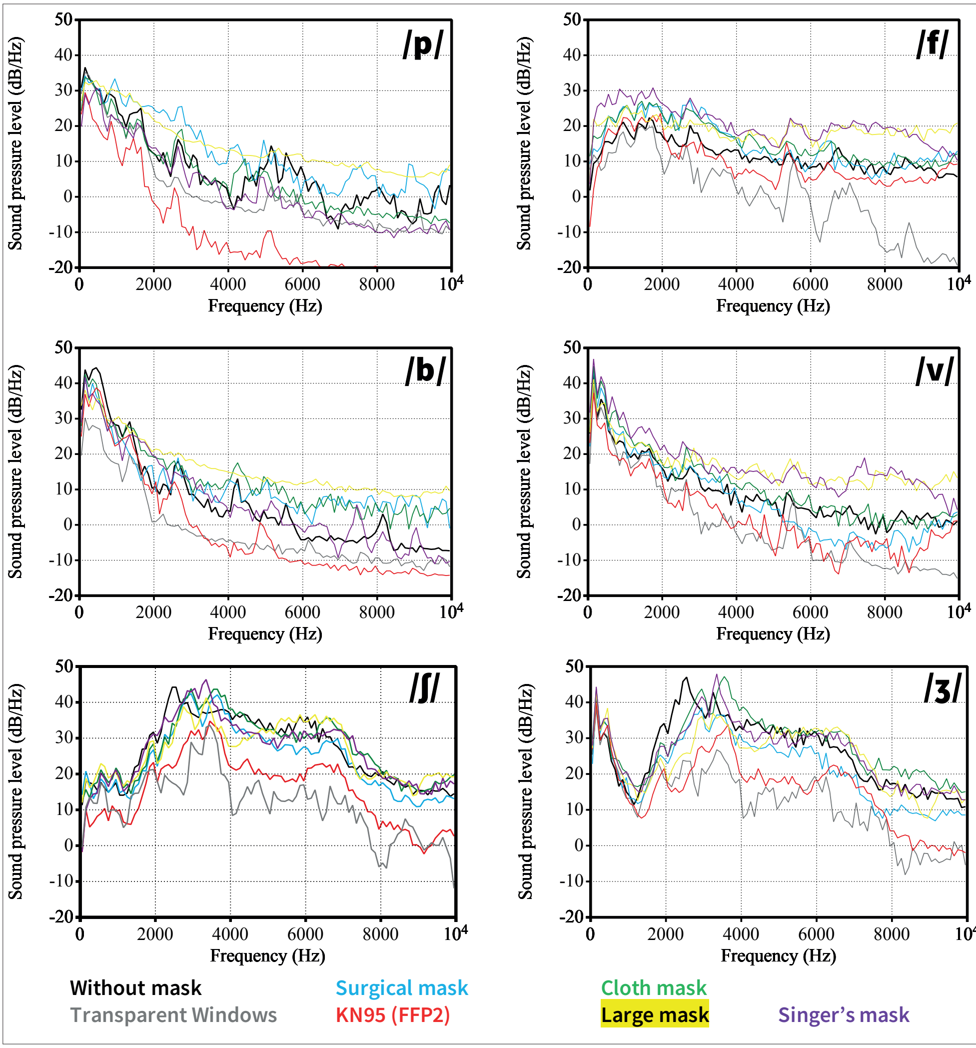

Figure 2 Averaged spectra (3 repetitions) of the French bilabial (/p b/), labio-dental (/f v/) and post-alveolar (/ʃ ʒ/) consonants produced by the female speaker. Each averaged spectrum is plotted against the type of mask worn by the speaker during the first recording session

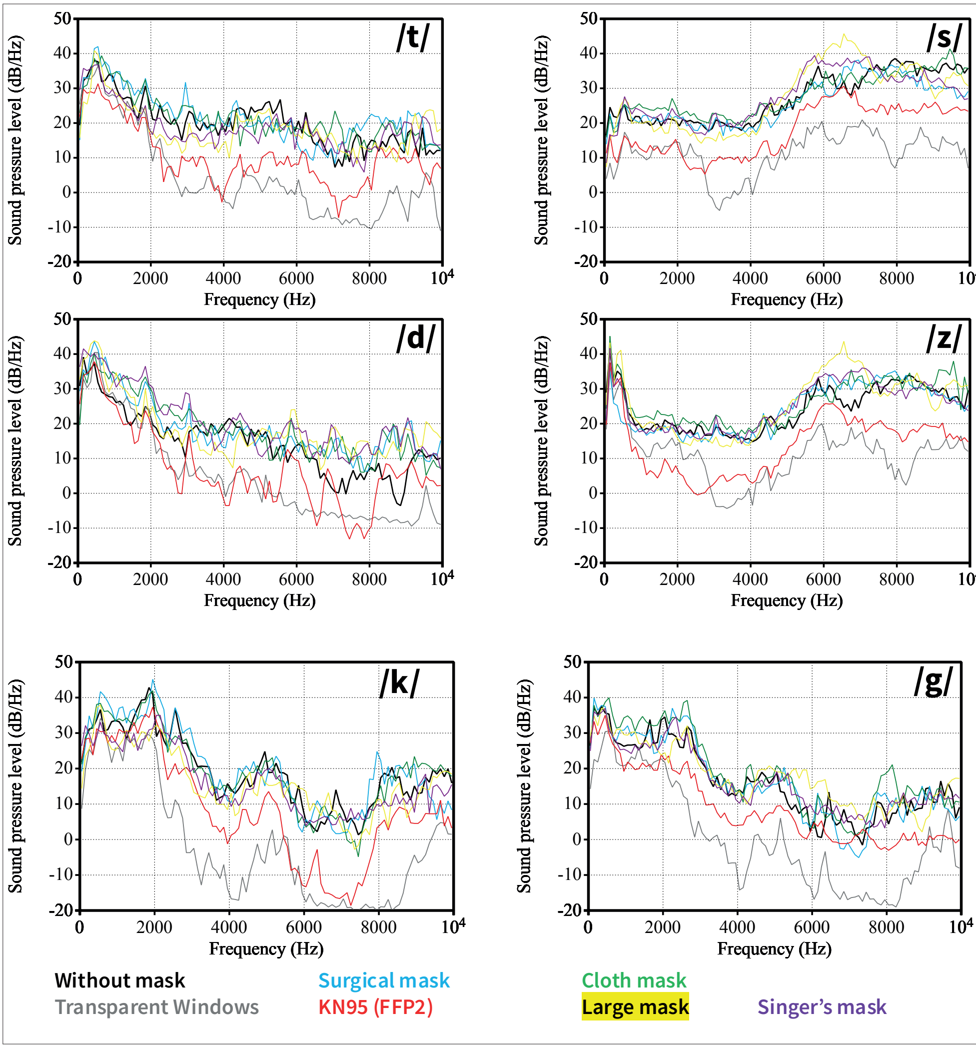

Figure 3 Averaged spectra (3 repetitions) of the French dentals (/t d s z/) and velars (/k g/) produced by the female speaker. Each averaged spectrum is plotted against the type of mask worn by the speaker during the first recording session

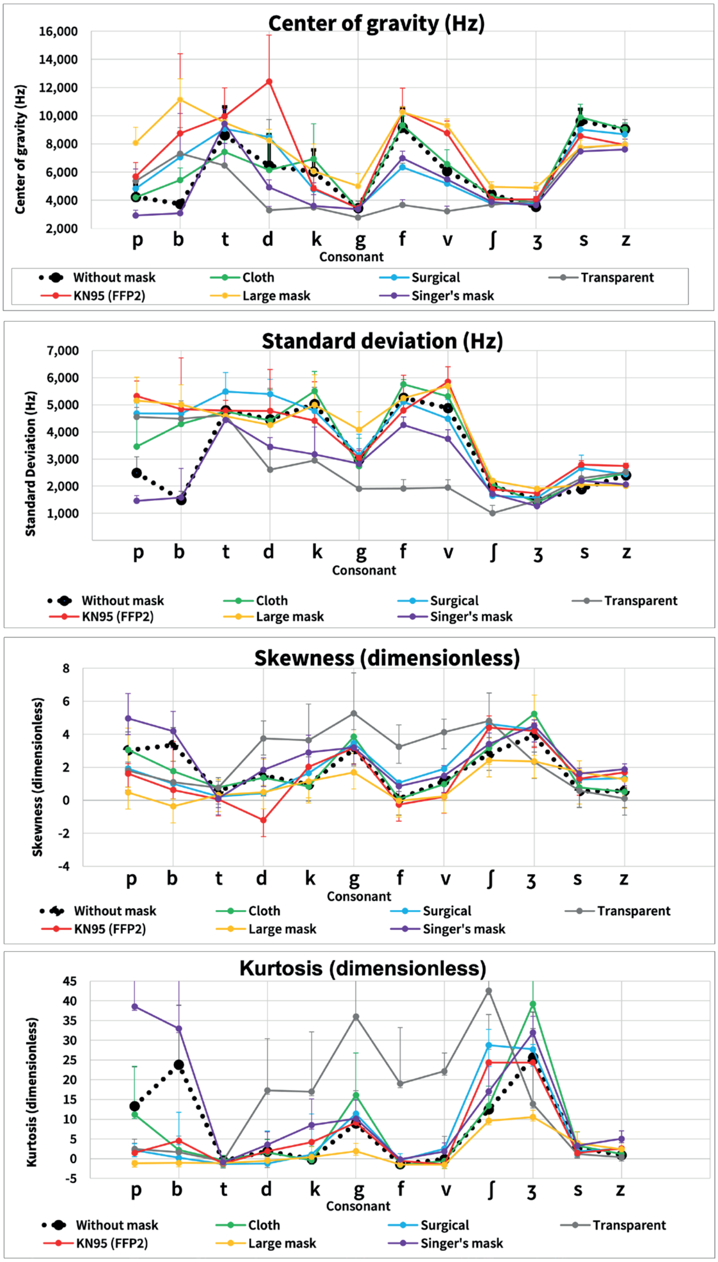

Figure 4 Spectral centre of gravity, standard deviation, coefficient of asymmetry (skewness) and kurtosis as a function of the type of mask worn and the consonant produced by the female speaker: mean and standard deviation of the 3 repetitions

gy concentration at given frequencies for /p b/, in contrast to /d k g f v ʃ/. The use of cloth masks, then surgical and singer's masks, does not cause major acoustic changes in these parameters.

3.3 Results 2: Global Acoustic Distortions in Conversational, Loud Speech and Singing

Since the subjects surveyed mentioned in their free comments a distortion by the masks on utterances as a whole: sound 'cut-off', lack of carrying power of the voice and consequent difficulty of comprehension by their interlocutors asking them to repeat their words, we chose, consequently, to determine what were the global acoustic distortions on conversational, loud spoken and sung utterances by a male and a female speaker as a whole.

Figure 5 shows the long-term averaged spectra of the set of texts read and sung by the male speaker (one session, spoken task) and the female speaker (two sessions, conversational, loud spoken, and sung). This illustrates that the filtering caused by the masks starts from 1kHz, and occurs mainly in the higher frequencies, from 2kHz.

In addition, the KN95 (FFP2) masks and especially the transparent masks are the most filtering, whatever the task and the speaker, except for the first session of sung productions by the female speaker: the transparent mask shows antiformants around 3000Hz for the male speaker and around 3800Hz for the female speaker, especially in loud voice. We also notice that the singer's mask is the least filtering, and that the fabric (cloth) mask worn by the male speaker is more filtering than the one worn by the female speaker. In session 2 of these singing productions, the frequencies corresponding to the singer's formant, between 2000Hz and 4000Hz, are attenuated by the large mask and the transparent mask, then by the KN95 (FFP2) mask and, to a lesser degree, the singer's mask.

Moreover, there are visible differences between the two sessions produced by the female speaker, such as a greater attenuation of the energy from 2300Hz onwards by the transparent mask in the spoken productions of the second session, or greater difference between the averaged spectra of the different masks in the second sung session than in the first.

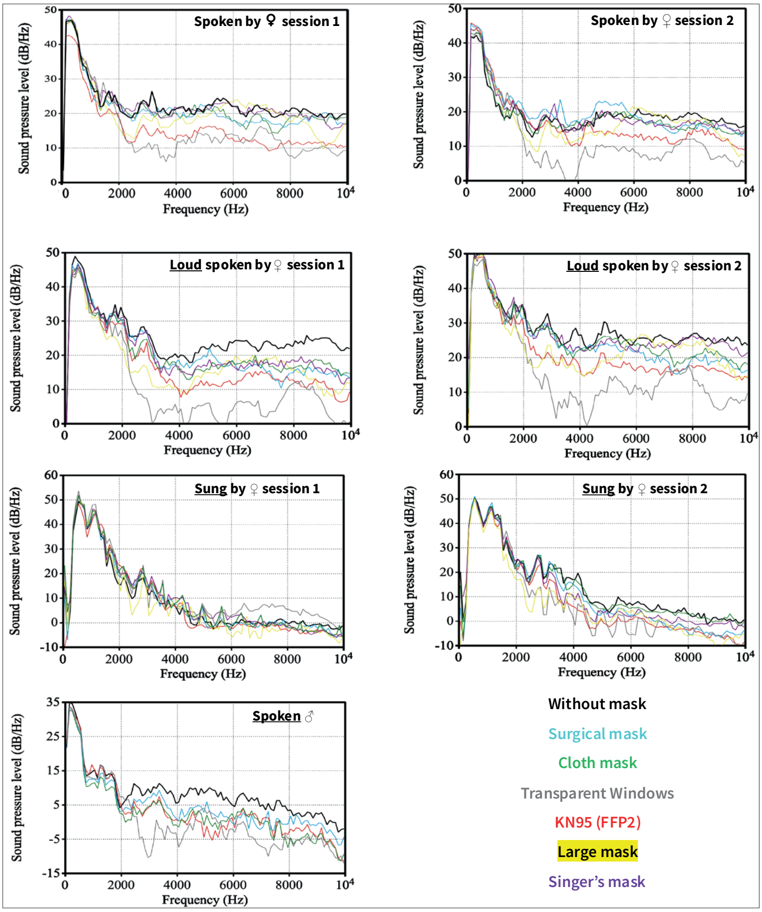

Figure 5 Long-term averaged spectra of conversational spoken (female speaker, two sessions: top; male speaker, one session: bottom), loud spoken (female speaker, two sessions) and sung (female speaker, two sessions) texts as a function of mask type

e-ISSN 2499-1562

Paired two-tailed t-test comparisons between the long-term averaged spectra of the productions without mask and those with each type of device worn show significant results [tab. 1], except for the singer's mask in the female speaker's spoken (2 sessions) and loud spoken (session 2) productions, and concerning the surgical and KN95 (FFP2) masks for the first session of sung productions. In the same tasks except for the last one, the KN95 (FFP2) and transparent masks show maximum t-values, and the surgical and singer's masks show almost minimum values. These t-values are minimal in spoken task by the male speaker.

Table 1 Paired comparisons (paired two-tailed t-test) between each of the 100 sound pressure level values (from 0 to 10,000Hz) of the long-term averaged spectra of productions without mask and those with each type of wearable device: difference in means (M diff), t-values and significance. The colours refer to the type of mask and are identical to those in figures 2 to 5. The results are listed in ascending order of t absolute values, except for the right column. 99: degree of freedom. Right column: paired comparison of sessions 1 and 2 for the differences in spectral amplitude every 100Hz between not wearing a mask and wearing each mask type

Spoken by ♀ session 1				Spoken by ♀ session 2				Session effect	
M diff	Mask	t(99)	Significance	M diff	Mask	t(99)	Significance	t(99)	Sign.
0,33	Singer's mask	1,91	**0,059**	0,25	Singer'smask	1,21	**0,233**	-3,07	0,002
2,47	Large mask	8,03	<.0001	-0,92	Surgical	-3,01	0,003	-3,06	0,003
1,76	Surgical	8,69	<.0001	0,74	Cloth	3,99	0,0001	1,42	**0,159**
1,58	Cloth	9,19	<.0001	1,85	Large mask	5,72	<.0001	1	**0,319**
7,7	Transparent	15,62	<.0001	2,74	KN95 (FFP2)	6,73	<.0001	11,13	<.0001
7,34	KN95 (FFP2)	30,54	<.0001	7,99	Transparent	13,99	<.0001	-2,2	0,03

Spoken ♂			
M diff	Mask	t(99)	Significance
5,12	KN95 (FFP2)	14,59	<.0001
7,65	Transparent	15,27	<.0001
3,73	Surgical	19,49	<.0001
5,84	Cloth	25,42	<.0001

Loud spoken by ♀ session 1				Loud spoken by ♀ session 2				Session effect	
M diff	Mask	t(99)	Significance	M diff	Mask	t(9)	Significance	t(99)	Sign.
4,5	Surgical	11,98	<.0001	1,09	Singer's mask	1,11	**0,268**	7,46	<.0001
4,43	Singer's mask	15,89	<.0001	2,64	Cloth	9,82	<.0001	7,35	<.0001
8,3	KN95 (FFP2)	19,26	<.0001	3,73	Surgical	11,77	<.0001	3,34	0,001
4,7	Cloth	19,7	<.0001	4,96	Large mask	12,06	<.0001	8,37	<.0001
7,83	Large m.	21,35	<.0001	12,86	Transparent	20,82	<.0001	4,36	<.0001
14,8	Transparent	21,83	<.0001	7,08	KN95 (FFP2)	21,45	<.0001	4,43	<.0001

Sung by ♀ session 1				Sung by ♀ session 2				Session effect	
M diff	Mask	t(99)	Significance	M diff	Mask	t(99)	Significance	t(99)	Sign.
-0,16	Surgical	-0,78	**0,437**	0,82	Cloth	4,87	<.0001	5,61	<.0001
0,66	KN95 (FFP2)	1,6	**0,113**	2,93	Singer's mask	9,05	<.0001	-2,57	0,01
-1,05	Cloth	-3,52	0,0006	3,32	Surgical	10,76	<.0001	-6,94	<.0001
1,29	Large mask	3,97	0,0001	6,95	Transparent	14,56	<.0001	-6,82	<.0001
-1,4	Singer's mask	-4,62	<.0001	7,15	Large mask	15,33	<.0001	-9,2	<.0001
-3,46	Transparent	-12,24	<.0001	6,13	KN95 (FFP2)	15,6	<.0001	-7,78	<.0001

Between sessions 1 and 2, the paired two-tailed t-test comparison of the differences in spectral amplitude every 100Hz between the absence and presence of a given mask type (tab. 1: right column)

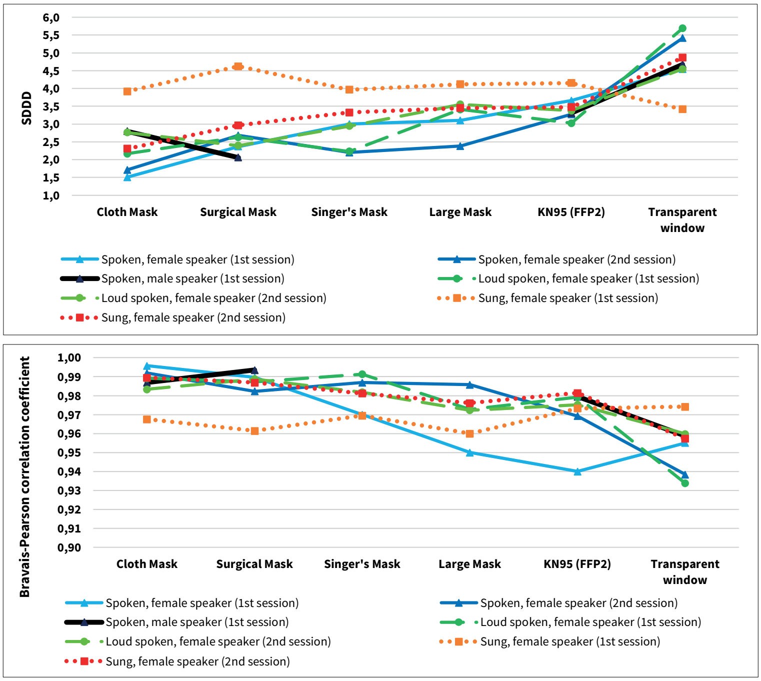

Figure 6 SDDD (top) and Bravais-Pearson correlation coefficient (bottom) as a function of session, task and mask type compared to production without mask

shows statistically significant differences between these sessions, except for the large mask and cloth mask for spoken productions. These differences are less significant for the spoken productions (except for the KN95 (FFP2) mask), for the surgical mask and the singer's mask.

Figure 6 quantifies these spectral differences using the *Standard Deviation of the Differences Distribution* (SDDD) (Harmegnies 1988), and the Bravais-Pearson correlation coefficients between each long-term averaged spectrum of a production without a mask and the same production with a mask type. Results show that the SDDD values are globally higher for the KN95 (FFP2) and transparent masks (except for the song recorded in the first session) and the lowest for the cloth and surgical masks. The opposite is observed for the Bravais-Pearson correlation coefficient values; however, it remains high and ranges between 0.934 and 0.996. Note that the cloth mask worn by the male speaker results in higher SDDD values than the spoken productions of the female speaker. The opposite is observed for the Bravais-Pearson correlation coefficients.

e-ISSN 2499-1562

4 Discussion

This exploratory study highlighted several anti-COVID face mask effects on the voice and speech of French speakers: subjective results have been verified by acoustic analyses.

The feedback from the 303 participants interviewed showed that 39-42% of them had difficulty producing plosives with a mask, which seems to correspond to significant acoustic alterations in bilabial and labio-dental plosives. However, more participants wearing KN95 (FFP2) masks and the other mask types we tested acoustically should be interviewed to confirm these answers. The participants also expressed other concerns such as the loss of voice carrying power, which leads to the need to repeat their words to their interlocutors more often when wearing the mask: acoustic attenuation from 2000Hz, whereas the determining frequencies in the voice range are between 1000Hz and 4000Hz (Leino 2009), explains these feelings.

As mentioned in our state of the art, to our knowledge, there is no acoustic study in French on the description of local acoustic distortions on plosives and fricatives produced with and without masks: Fecher's (2014) thesis deals with another language than French, with unvoiced fricatives and plosives only, and with other types of face veiling devices than most of our masks. However, Fecher (2014) results agree with the values of spectral centre of gravity and standard deviation that were not significantly modified whatever the type of face shield, for /s ʃ/; obtained in this study. Increased values of coefficient of asymmetry for /f/, constant values of coefficient of kurtosis for /s/, but higher and variable values for /ʃ/ were also attested. The acoustic changes in plosives we obtained, apart from a variable standard deviation of the burst for /p k/ and consistent for /p/, do not agree with Fecher's (2014) findings for the centre of gravity of the burst of /p t k/. However, this author found negligible acoustic changes between the control condition (without mask) and the condition of wearing the surgical mask, the only facial covering device common with our study.

Furthermore, it seems that consonants mobilising labial movements (/p b f v/ and to a lesser degree /ʃ ʒ/) undergo more acoustic distortions than other consonants, probably due to the acoustic disturbance by contact between the lips and the mask.

The overall acoustic changes also show that the transparent and KN95 (FFP2) masks attenuate frequencies above 1,000Hz to a greater extent, but less so for singing, especially in the first session, which is in contradiction with the study of Oren et al. (2021): the singer female speaker seems, here, likely to adapt to this external disturbance caused by the mask, in order to minimise its effects. However, given the mostly significant differences between sessions for this speaker, she seems to show some variability in this type of adaptation to an external disturbance like the use of a mask.

The acoustic effect of the fabric (cloth) mask is greater in male than in female speaker because fabric masks vary considerably in their composition and weave. In addition, surgical masks offer the best acoustic performance of all masks tested. If these masks are not available, loosely woven 100% cotton masks offer good acoustic performance (Corey, Jones, Singer 2020).

Finally, the singer's mask, which has the added advantage over the surgical mask does not interfere with the articulation of segments requiring rounding and protrusion of the lips. The singer's mask appears to be a good alternative to the surgical mask for subjects who need to vocalise and sing.

However, all of these preliminary results should be interpreted with great caution, because of the small number of speakers and the variability between the two production sessions for the female speaker. The variability is interesting, because it shows that the same speaker can adapt her voice and speech production differently depending on the recording session. This variability is an integral part of speech (Perkell, Klatt 1987). It cannot therefore be studied only with the help of talking heads, as most research on the effects of masking on speech has done.

5 Conclusion

The main contribution of this case study, compared to previous research, is the descriptive data on the feeling of French-speaking participants regarding the use of different types of masks. Our contribution also provides new information on the nature of acoustic distortions in French concerning voiced and unvoiced French fricative and plosive consonants, and concerning whole conversational, loud spoken and sung utterances actually produced by two speakers and in two sessions for the female speaker. In addition, the analysis of acoustic distortions caused by the large mask and the singer's mask has not been studied until now.

The results of the acoustic analyses in this case study confirm that, both locally (consonants, as felt by the participants of the questionnaire) and globally (the whole utterance), KN95 (FFP2) and especially transparent masks cause the greatest acoustic distortions in the spoken, loud and, to a lesser extent, sung signal, compared to equivalent productions without masks. The presence of these types of masks therefore constitutes an external source of disturbance to the speech production system (Sock 2001), but to a lesser extent in singing in the context of this study. On the other hand, the singer's mask minimises the energy attenuation in the majority of the tasks produced by the female speaker.

However, these preliminary results should be confirmed 1) by analyses carried out on similar productions by other speakers; 2) by more robust statistical analyses, and 3) by the study of the production of nasal and rounded vowels. The former are felt to be difficult to produce by the participants according to the questionnaire, and the latter are supposed to be disturbed in their emission because of the contact between the rounded and protruding lips against the mask.

We are currently seeking to complete these preliminary results on the effects of the mask by perceptual tests of consonant identification and discrimination of the same spoken extracts, declaimed and sung by the female speaker, in order to know if the local and global acoustic distortions highlighted here are perceptually relevant.

Acknowledgments

This work has been supported by *Partenariats Hubert Curien (PHC) TOURNESOL nos. 46241WH* and *ANR-10-LABX-0083*. It contributes to the IdEx Université de Paris (ANR-18-IDEX-0001). The authors thank Andrés Lara for proofreading an earlier draft of this paper.

The authors have declared that they have no competing or potential conflicts of interest in relationship to this article.

Bibliography

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 the American Academy of Audiology*, 28(1), 58-67. http://doi.org/10.3766/jaaa.15151.

Boersma, P.; Weenink, D. (1992-2021). Praat: doing phonetics by computer [Mac]. Version 6.1.16 retrieved 2 February 2021 from http://www.praat.org/.

Bottalico, P.; Murgia, S.; Puglisi, G.E.; Astolfi, A.; Kirk, K.I. (2020). “Effect of Masks on Speech Intelligibility in Auralized Classrooms”. *The Journal of the Acoustical Society of America*, 148(5), 2878-84. http://doi.org/10.1121/10.0002450.

Cohn, M.; Pycha, A.; Zellou, G. (2021). “Intelligibility of Face-Masked Speech Depends on Speaking Style: Comparing Casual, Clear, and Emotional Speech”. *Cognition*, 210, 104570. http://doi.org/10.1016/j.cognition.2020.104570.

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), 2371-5. http://doi.org/10.1121/10.0002279.

Erber, N.P. (1969). “Interaction of Audition and Vision in the Recognition of Oral Speech Stimuli”. *Journal of Speech and Hearing Research*, 12(2), 423-5. http://doi.org/10.1044/jshr.1202.423.

Fecher, N.; Watt, D. (2013). "Effects of Forensically-Realistic Facial Concealment on Auditory-Visual Consonant Recognition in Quiet and Noise Conditions". *12th International Conference on Auditory-Visual Speech Processing (AVSP)* (August 29-Sept, 1 2013). Annecy: Inria, 1-6.

Fecher, N. (2014). *Effects of Forensically-Relevant Facial Concealment on Acoustic and Perceptual Properties of Consonants* [PhD Dissertation]. York: University of York.

Fiorella, M.L.; Cavallaro, G.; Di Nicola, V.; Quaranta, N. (2021). "Voice Differences When Wearing and Not Wearing a Surgical Mask". *Journal of Voice*, online since 10 March 2021. http://doi.org/10.1016/j.jvoice.2021.01.026.

Harmegnies, B. (1988). "SDDD: A New Dissimilarity Index for the Comparison of Speech Spectra". *Pattern Recognition Letters*, 8(3), 153-8. http://doi.org/10.1016/0167-8655(88)90093-1.

Harrington, J. (2010). "Acoustic Phonetics". Hardcastle, W.J.; Laver, J.; Gibbon, F.E. (eds), *The Handbook of Phonetic Sciences*. 2nd ed. Malden; Oxford; Chichester: Wiley-Blackwell, 81-129. http://doi.org/10.1002/9781444317251.ch3.

Jongman, A.; Wayland, R.; Wong, S. (2000). "Acoustic Characteristics of English Fricatives". *The Journal of the Acoustical Society of America*, 108(3), 1252-63. http://doi.org/10.1121/1.1288413.

Howard, J.; Huang, A.; Li, Z.; Tufekci, Z.; Zdimal, V.; van der Westhuizen, H.M.; von Delft, A.; Price, A.; Fridman, L.; Tang, L-H.; Tang, V.; Watson, G.L.; Bax, C.E.; Shaikh, R.; Questier, F.; Hernandez, D.; Chu, L.F.; Ramirez, C.M.; Rimoin, A.W. (2021). "An Evidence Review of Face Masks Against COVID-19". *Proceedings of the National Academy of Sciences*, 118(4), 1-12. http://doi.org/10.1073/pnas.2014564118.

Leino, T. (2009). "Long-Term Average Spectrum in Screening of Voice Quality in Speech: Untrained Male University Students". *Journal of Voice*, 23(6), 671-6. http://doi.org/10.1016/j.jvoice.2008.03.008.

Magee, M.; Lewis, C.; Noffs, G.; Reece, H.; Chan, J.C.; Zaga, C.J.; Paynter, C.; Birchall, O.; Rojas Azocar, S.; Ediriweera, A.; Kenyon, K.; Caverlé, M.W.; Schultz, B.G.; Vogel, A.P. (2020). "Effects of Face Masks on Acoustic Analysis and Speech Perception: Implications for Peri-Pandemic Protocols". *The Journal of the Acoustical Society of America*, 148(6), 3562-8. http://doi.org/10.1121/10.0002873.

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(9), 686-95. http://doi.org/10.3766/jaaa.19.9.4.

Nguyen, D.D.; McCabe, P.; Thomas, D.; Purcell, A.; Doble, M.; Novakovic, D.; Chacon, A.; Madill, C. (2021). "Acoustic Voice Characteristics with and Without Wearing a Facemask". *Scientific Reports*, 11(1), 1-11. http://doi.org/10.1038/s41598-021-85130-8.

Oren, L.; Rollins, M.; Gutmark, E.; Howell, R. (2021). "How Face Masks Affect Acoustic and Auditory Perceptual Characteristics of the Singing Voice". *Journal of Voice*, online since 26 March 2021. http://doi.org/10.1016/j.jvoice.2021.02.028.

Palmiero, A.J.; Symons, D.; Morgan, J.W.; Shaffer, R.E. (2016). "Speech Intelligibility Assessment of Protective Facemasks and Air-Purifying Respirators". *Journal of Occupational and Environmental Hygiene*, 13(12), 960-8. http://doi.org/10.1080/15459624.2016.1200723.

Perkell, J.S; Klatt, D.H. (1987). “Invariance and Variability in Speech Processes”. *The American Journal of Psychology*, 100(2), 306. http://doi.org/10.2307/1422414.
Ribeiro, V.V.; Dassie-Leite, A.P.; Pereira, E.C.; Nunes Santos, A.D.; Martins, P.; de Alencar Irineu, R. (2020). “Effect of Wearing a Face Mask on Vocal Self-perception During a Pandemic”. *Journal of Voice*. http://doi.org/10.1016/j.jvoice.2020.09.006.
Sock, R. (2001). “La Théorie de la Viabilité en production-perception de la parole”. Keller, D.; Durafour, J-P.; Bonnot, J-F.; Sock, R. (eds), *Psychologie et Sciences Humaines*. Liège: Mardaga, 285-316.
Stevenson, R.; Sheffield, S.W.; Butera, I.M.; Gifford, R.H.; Wallace, M. (2017). “Multisensory Integration in Cochlear Implant Recipients”. *Ear and Hearing*, 38(5), 521-38. http://doi.org/10.1097/aud.0000000000000435.
Stuart-Smith, J.; Timmins, C.; Wrench, A. (2003). “Sex and Gender in Glaswegian /s/”. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)* (Barcelona, August 3-9, 2003). Barcelona: Universitat Autonoma de Barcelona, 1851-4.