The Influence of Face Coverings on Speech

Maximilian Prinz, Carina Lozo
Univeristy of Vienna, 1010 Vienna, Austria

Abstract
The global pandemic of COVID-19 lead to mandatory face mask laws in many states of the EU, spawning a general conception that face masks alter speech intelligibility. In a two-part study we investigated the effect of three different face masks (FFP2, surgical and cloth mask) on acoustic features as well as intelligibility loss. Part one of the study involved the acoustic analysis of speech in different face mask conditions. Acoustic parameters that include harmonic relations (e.g., H1-H2), formant amplitudes (e.g., A1) and noise measures (CPP(S)) were extracted from the data of one female and one male speaker. To control for intra-speaker variability and safety reasons, we built an artificial head-loudspeaker system adapted to the human voice to play the sound samples. The experiment was performed in a sound attenuated booth using two condenser microphones (AKG C451). In part two of this study a perception test was carried out to test the effect of the face masks on speech intelligibility. The statistical analyses of the first experiment revealed that sounds are affected most in higher frequency bands. The perception test showed that the masks were perceived as equally intelligible.

Introduction
The SARS-CoV-2 outbreak in spring 2020 led to extensive cuts into everyday life as we knew it. Protective face coverings have become a daily companion and took a supposed toll on our ability to understand each other. The pandemic gave rise to research into the acoustic effects of these face coverings and numerous studies described how face masks challenge our communication. Previous studies (e.g., [1, 2, 3]) came to a similar conclusion regarding the effect of masks on the speech signal as they showed that face coverings tend to mostly attenuate speech in higher frequencies, while having little to no effect regarding lower frequencies [2, 4, 5, 6]. In order to understand the effects of different masks on speech, it thus becomes necessary to study acoustic parameters that describe these higher frequencies. It is believed that the quality of speech is decreased and problems in perception arise (e.g., [7]). Arguments as such are often used against wearing face coverings as it seemingly becomes more difficult to talk to each other. Perception tests performed in previous studies, showed that, while higher frequencies are attenuated more (e.g., [8]), the intelligibility of speech suffers only little to nothing (e.g., [9]). However, [10] found in their study that different face mask result in a different degree of intelligibility depending on the environmental context. For this contribution we recorded three different masks (surgical mask, FFP2 mask, and cloth mask) in a highly controlled environment. We first built a dummy head with an integrated loudspeaker in order to control for speaker variability. Afterwards a speech signal was played through the loudspeaker, using different masks as conditions, while keeping the distance to the recording device fixed. Acoustic parameters such as corrected harmonic amplitudes, harmonic relations and measures of noise were extracted and statistically analysed. Further, the intelligibility of different mask types was evaluated by 16 listeners in a perception experiment.

Data
For this study, recordings of one male and one female informant under the age of 25, were used by playing them through a loudspeaker setting. The original data was recorded with a condenser microphone, AKG C451 that was placed 30 cm right in front of the informants head, with a sampling rate of 44.1 kHz and 16 bits/sample. The recorded readings provided for this study included the parable *Nordwind und Sonne* and the none-sense poem *Karawane*. Target sounds are all vowels /a, e, i, o, u/ and the consonants /f, m, n , r, s/ and /z/. The recordings for this contribution were conducted through a loudspeaker and microphone setup described below in the methods section. In total 12 471 target sounds were analysed.

Methods
Ludwig – the dummy head
The experimental setup uses an artificial head-loudspeaker system, displayed in figure 1, which we named *Ludwig*. A commercial broadband loudspeaker with a frequency range from 50 Hz to 20 kHz and an isotropic directivity pattern (it emits sound uniformly in all directions in front of it) was used. The measurements were carried out in a sound attenuated booth at the Acoustic Research Institute of the Austrian Academy of Sciences in Vienna, Austria. Speech data of two speakers described above, was played through the loudspeaker using the following face coverings as conditions: FFP2 mask, cloth mask, surgical mask, and no mask. For each mask and speaker, five rounds of *Nordwind und Sonne* and *Karawane* were recorded, the control condition was recorded ten times per text. The measurements were conducted with a condenser microphone, AKG C451, placed 50 cm in front of the loudspeaker. For the extraction of the acoustic speech parameters, an adapted script by [11] for the software Praat [12] was used. In order to find significant differences, we used multivariate analysis of variance (MANOVA) and a post-hoc Tukey test. The MANOVA was conducted to state whether or not the tested face coverings differ from one another, considering following parameters (* indicate values corrected for the effects of the vocal tract with [13]): H1* (dB), H2* (dB), H4* (dB), H1*-H2* (dB), H2*-H4* (dB), A1* (dB), A2* (dB), A3* (dB), CPP(S)(dB), amplitude of the harmonic closest to 2 kHz and to 5 kHz, as well as

Figure 1: Different facial coverings on Ludwig during the recordings.

their relative difference (2 kHz - 5 kHz).

Pairwise Comparison

In order to assess the effect on intelligibility of the different masks for listeners, a two-part online evaluation was set up with [14]. In total, 16 listeners (8 female, 5 male, and 3 NA) in two age groups (18–30 years, N = 12, 31–45 years, N = 2, and 2 NA) participated in the survey (14 of which are completed). It has to be noted that the number of times a condition was compared is not equal among all conditions since some participants did not complete the survey or skipped certain questions. The data for this evaluation consisted only of read speech from the female speaker. Before the pairwise comparison started, the speaker's overall intelligibility was rated by the listeners on a scale from 1 (high intelligibility) to 5 (poor intelligibility). (i) and (ii) show the two sentences from the texts described above.

(i) Als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam. (*Nordwind*)

(ii) Blago bung, blago bung, bosso fataka. (*Karawane*)

The utterances were presented to the listener in five conditions. Alongside with the three different masks, we chose two control conditions: one with no mask and one with no mask but with background noise (restaurant chatter). In the first part the participants were asked to transcribe the two presented utterances, in order to ensure that intelligibility was given for the listener. The second part consisted of a pairwise comparison, in which each sample (N = 10) was compared with one another and the listener was asked to choose the more intelligible sample. Overall the listeners were presented with 20 comparisons.

Results

Acoustic measurements

We choose a significance level of $\alpha = 0.05$ for all statistical tests and took adjusted *p-values* into consideration where necessary. The results of the MANOVA for the acoustic measurements in the poem *Nordwind* can be seen in table 1. *Wilks' Lambda* (Λ) is a test-statistic which computes the proportion of variance in the dependent variables that can not be linked to the independent variables. The column *F-value* gives the value of the corresponding F-statistic; *p-value* gives the probability of the *F-value* being larger than what was measured. In order to be significant the *p-value* needs to be smaller or equal to 0.05. The column *Num Df* gives the numerator degrees of freedom of the corresponding F-statistic; the column *Den Df* gives the denominator degrees of freedom of the corresponding F-statistic.

Nordwind und Sonne

For the female speaker we found differences in all dimensions ($p \leq 0.001$), except for CPP(S) (*Wilk's Lambda* = 0.9984, $F(6, 8594) = 1.1716$, $p = 0.3182$). For the male speaker we found significant differences in all dimensions ($p \leq 0.0148$), for details see table 1. The post-hoc Tukey test reported 53 significant differences between all face coverings, which corresponds to 34% of possible differences. Of the 53 significant differences, 25 were recorded for the female speaker ($p \leq 0.0315$), and 28 for the male speaker ($p \leq 0.0337$). The most differences have been counted between the cloth mask and the control group with 16, the second most differences have been counted for the surgical mask and the control group with 12, the third most differences have been counted between the FFP2 mask and the control group with nine, followed by the FFP2 mask and the cloth mask with nine, the surgical mask and the FFP2 mask with six, and the cloth mask and surgical mask with two significant differences between each other. The amount of significant differences by parameter are 12 for A3*, ten for 5 kHz and for 2 kHz - 5 kHz, six for 2 kHz, five for H4*, three for A2* and H1*-H2*, two for A2* and one for H2*-H4* and CPPS. See table 1 for a summary of the multivariate effects for *Nordwind*.

Karawane

For the female speaker we found significant differences for the dimensions related to the amplitude of formants (*Wilk's Lambda* = 0.9426, $F(9, 5259.4568) = 14.3639$, $p \leq 0.001$), and for the dimensions related to the amplitudes of the harmonics closest to two and five kHz (*Wilk's Lambda* = 0.9033, $F(9, 4324) = 37.6039$, $p \leq 0.001$). For the male speaker we found significant differences in all dimensions ($p \leq 0.0276$), except for CPP(S) (*Wilk's Lambda* = 0.9961, $F(6, 3462) = 1.1294$, $p = 0.3423$). The post-hoc Tukey test reported 39 significant differences between all face coverings, which corresponds to 25% of possible differences. Of the 39 significant differences, 19 were recorded for the female speaker ($p \leq 0.0358$), and 20 for the male speaker ($p \leq 0.0436$). The most differences have been counted between the FFP2 mask and the control group with 12, the second most differences have been counted for the cloth mask and the control group with ten, the third most differences have been counted between the surgical mask and the control group with seven, followed by the surgical mask and the FFP2 mask with five, the FFP2 mask and the cloth mask with five, and the cloth mask and surgical mask with zero significant differences between each other. The amount of significant differences by parameter are ten for A3* and 5 kHz, eight for 2 kHz - 5 kHz, five for 2 kHz, two for H4* and H2*-H4* and one for A1* and A2*.

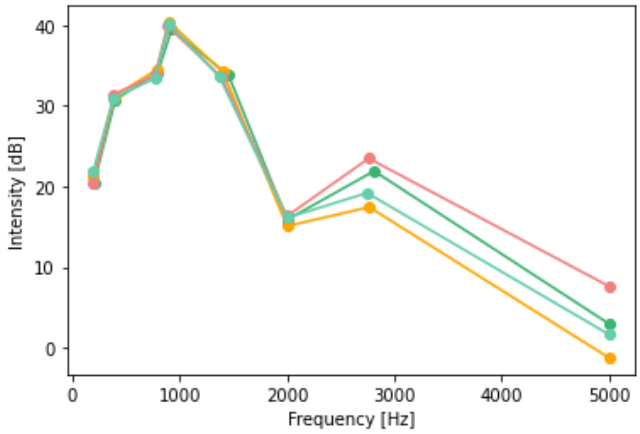

Figure 2: Average spectral energy distribution for the vowel [a] in the nonsense text *Karawane* for the female speaker. Colours indicate different masks: ● control, ● surgical mask, ● cloth mask, ● FFP2 mask.

Pairwise comparison

The overall intelligibility of the female speaker was rated good (avg = 1.75). The given percentages refer to the total number of comparisons which included the condition, further given with *N*. As for the pairwise comparisons, we found that most samples were rated to be equally intelligible (87% of all comparisons (N = 271). The condition with the most wins is the control condition without mask (40%, N = 113), followed by surgical mask (36% N = 116), FFP2 mask (29%, N = 115), cloth mask (28%, N = 116) and noise control with the least wins (0.04%, N = 117).

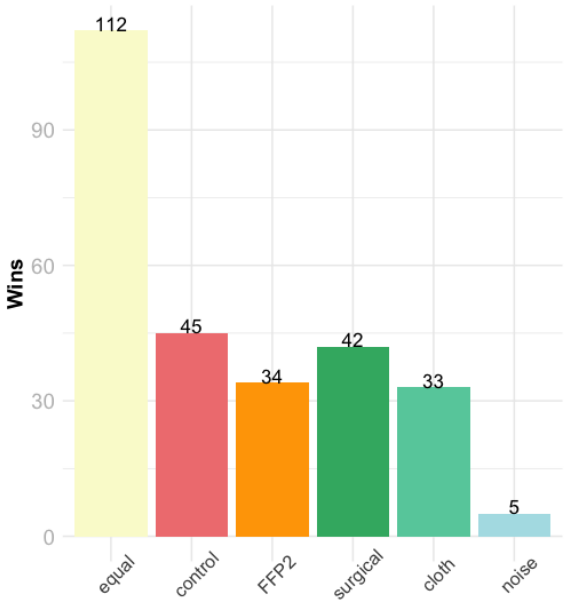

Figure 3: Total number of wins per condition (out of 271 comparisons in total).

Discussion

MANOVA

Our MANOVA calculates extremely high values for *Wilks' Lambda*, as only one value is below 0.9, see table 1. As *Wilks' Lambda* gives the proportion of variance that can not be attributed to the independent variables, up to 99% of variance can not be explained by different face coverings. This does not make sense, as the only difference between the recordings as described above are the face coverings used and thus any variation that is not noise or an unwanted signal has to be linked to the different guises.

Differences in Face Coverings

We found at least one difference between all masks, but were surprised regarding *Nordwind und Sonne*, as the cloth mask counted more significant differences when compared to the control measurements, than the surgical mask and the FFP2 mask. We expected the FFP2 mask to show the most differences compared to the control group, but it appears that the tested face coverings behave similar to another. This is certainly the case for the cloth mask and the surgical mask, as we found only two significant differences between them, suggesting similarities in behaviour. For contrast the FFP2 mask differs more from the other face coverings, which we expected due to requirements made regarding medical usage, whereas the others are primarily of commercial nature. For the poem *Karawane*, the FFP2 mask shows the most significant differences when compared to the control group, followed by the cloth mask and the surgical mask. The similarities between the cloth and surgical masks are further enhanced, as we found no significant differences between the two. The average spectral energy distribution of the female speaker is shown in figure 2 for the vowel [a]. Damping effects are clearly visible above 2 kHz, with the FFP2 mask performing the worst.

Differences in Parameters

As expected, high frequency parameters are affected most by the wearing of face coverings, as can be seen in the high amounts of significant differences for the parameters A3* and 5 kHz. The parameters CPP(S) do not seem to be influenced in as a significant way as the other parameters, which is surprising as those are indicators for noise and glottal opening. The parameter H1*-H2*, which serves as indicator of a breathy or creaky voice, accounted for overall three significant differences, suggesting that face coverings cannot be linked to a degradation in voice quality that would imitate those voice defects. Sounds more likely to be influenced by the high frequency attenuation are fricatives, as their key differences between one another are at higher frequencies. With our data we laid a foundation for further investigations regarding the specific behaviour of various sounds.

Pairwise comparison

The results of the pairwise comparison come as a surprise as we expected a higher win-rate for masks with little transmission loss. In the perception task most comparison

Table 1: Multivariate Effects for *Nordwind*. The column *Variables* denotes the acoustic features taken into consideration for the given MANOVA. *CPP(S)* stands for the (smoothed) cepstral peak prominence. *Harmonics* stands for the features H1*, H2*, H4*, H1*-H2* and H2*-H4*. *Formants* stands for A1*, A2* and A3*, *kHz* stands for the amplitude of the harmonics closest to two and five kHz as well as their difference.

Speaker	Nordwind					
	Wilk's Λ	F-Value	p-Value	Num Df	Den Df	Variables
female	0.9984	1.1716	0.3182	6	8594.0000	CPP(S)
female	0.9887	4.0830	0.0000	12	11363.7934	Harmonics
female	0.9150	43.1724	0.0000	9	10455.4858	Formants
female	0.9235	38.5991	0.0000	9	10455.4858	kHz
male	0.9959	2.9242	0.0075	6	8522.0000	CPP(S)
male	0.9932	1.9527	0.0148	15	11754.8644	Harmonics
male	0.8929	54.9122	0.0000	9	10367.8712	Formants
male	0.9188	40.7684	0.0000	9	10367.8712	kHz

conditions were rated equally intelligible, however. This can be explained by a highly controlled experiment setting and divergence from this would result in more pronounced differences between the masks depending partly on the acoustic properties of the area, as investigated by [10].

Summary

In line with our expectations, as well as with literature, higher frequency parameters are effected the most by the wearing of face coverings. As to why our values for *Wilks' Lambda* during the MANOVA are so high, we can not give any statement. Below 2 kHz the masks behave almost identical to each other, above that frequency they disperse significantly. The FFP2 mask shows most attenuation, however, as it is illustrated by figure 2, which is also reflected in the bad performance in the pairwise comparison. As to the other two face coverings, they behave more similar to one another. As for the perception, we can state that the face coverings in a highly controlled environment are similar in their degree of intelligibility. In order to grasp precise differences between the masks, the next step in this research should involve measurements closer to real life scenarios than our highly controlled setting. The focus on everyday life situations with different sets of room arrangements such as public transport and public places, or open plan offices as well as background noises would provide insights into how speech is affected by such circumstances.

Acknowledgments

We want to thank the Acoustic Research Institute of the Austrian Academy of Science for providing the lab facilities as well as DI (FH) Michael Mihocic, MSc for his help with experiment setup.

References

[1] Samuel R. Atcherson, B. Renee McDowell, and Morgan P. Howard. "Acoustic effects of non-transparent and transparent face coverings". In: *The Journal of the Acoustical Society of America* 149.4 (2021), pp. 2249–2254.

[2] Ryan M. Corey, Uriah Jones, and Andrew C. Singer. "Acoustic effects of medical, cloth, and transparent face masks on speech signals". In: *The Journal of the Acoustical Society of America* 148.4 (2020), pp. 2371–2375.

[3] Christoph Pörschmann, Tim Lübeck, and Johannes M. Arend. "Impact of face masks on voice radiation". In: *The Journal of the Acoustical Society of America* 148.6 (2020), pp. 3663–3670.

[4] Carmen Llamas et al. "Effects of different types of face coverings on speech acoustics and intelligibility". English. In: *York Papers in Linguistics Series 2* 9 (2008), pp. 80–104.

[5] Rahim Saeidi, Ilkka Huhtakallio, and Paavo Alku. "Analysis of Face Mask Effect on Speaker Recognition". In: *Interspeech 2016*. 2016, pp. 1800–1804.

[6] Rahim Saeidi et al. "Speaker recognition for speech under face cover". English. In: *Interspeech'15, Dresden, Germany, Sept. 6-10, 2015*. 2015.

[7] Vanessa Veis Ribeiro et al. "Effect of Wearing a Face Mask on Vocal Self-Perception during a Pandemic". In: *Journal of Voice* (2020).

[8] Alexander Goldin, Barbara Weinstein, and Nimrod Shiman. "How do medical masks degrade speech perception". In: *Hearing review* 27.5 (2020), pp. 8–9.

[9] Anastassia Loukina et al. "Do Face Masks Introduce Bias in Speech Technologies? The Case of Automated Scoring of Speaking Proficiency". In: *Interspeech 2020*. Interspeech 2020. ISCA, 2020, pp. 1942–1946.

[10] Klaus-Hendrik Lorenz-Kierakiewitz. "Zur Verringerung der Sprachverständlichkeit". In: *Akustik Journal* 3 (2020), pp. 42–52.

[11] Chad Vicenik, Patrick Callier, and Iseli. *praat_ voice_ measures.praat [source code]*. 2015.

[12] Paul Boersma and David Weenink. *Praat: doing phonetics by computer [Computer program]. Version 6.1. 26*. 2020.

[13] Markus Iseli, Yen-Liang Shue, and Abeer Alwan. "Age, sex, and vowel dependencies of acoustic measures related to the voice source". In: *The Journal of the Acoustical Society of America* 121.4 (2007), pp. 2283–2295.

[14] *Limesurvey GmbH. / LimeSurvey: An Open Source survey tool*. Hamburg, 2021. URL: `http://www.limesurvey.org`.