Vered Aharonson*, Craig S. Carlson, Michiel Postema, Hanna Putter-Katz, and Simona Tetin Schneider

Influence of wearing a face mask on speech properties

https://doi.org/10.1515/cdbme-2024-2004

Abstract: The increased use of face masks for infection prevention has resulted in difficulties in speech understanding. Thus, there is a need for tuning of acoustic plans in hearing aids to compensate for the loss of acoustic information due to the mask.Previous studies focused on isolated vowel sounds. Our study measured the effect of surgery masks on the acoustic attributes of vowels and compared the effect of isolated to co-articulated vowels. The voices of 45 male speakers were recorded whilst pronouncing isolated vowels and vowels within the co-articulation in a sentence. All recordings were repeated with the participants wearing a surgical mask. Acoustic factors of length, intensity, pitch, formants, shimmer, and jitter were extracted from the vowels in all the different conditions. The differences between masked and unmasked acoustics in the co-articulated vowels were found to be smaller than the isolated vowels acoustic features. These preliminary results improve the quantification of the effect of hearing a face mask in the real-life production of speech.

Keywords: speech perception, isolated vowels, surgical masks, prosody, co-articulation.

1 Introduction

The COVID-19 pandemic dramatically increased the use of face masks. The mask creates a physical barrier to the respiration and spray of particles from the nose and mouth during speech, cough, sneeze and singing [1, 2]. One symptom of mask wearing were complaints of people who experienced difficulty in understanding speech when speakers wear non-transparent surgical masks [3, 4]. As many individuals still wear masks as a requirement of their profession, their medical condition, or their personal preferences, the effect of mask on speech production and perception remains highly relevant. An understanding of the differences in the perception of speech when speakers wear masks may be useful in the design of acoustic plans of hearing aids. These tuned-to-mask plans could then better compensate the loss of acoustic information due to the mask for the hearing impaired. The physiology of speech production and perception and the source-and-filter model of speech production can explain the effect of masks [5] . When the face are covered from below the chin to over the nose root and across the cheeks, acoustic properties of the filter change. Previous studies suggested that surgical masks impact speech perception, particularly in noisy environments and for individuals with hearing loss, but yielded mixed results [3, 6–8]. The effects of surgical masks on speech production were studied using spectral and voice quality features which are commonly used by speech therapists. The studies reported mixed, sometimes contradicting effects of increase or decrease in the pitch (F0), formants, vocal intensity, jitter and shimmer in masked compared to non-mask speech [6, 8–13].

One cause for the differences in speech features when recorded in different speaker conditions is the language spoken. Nguyen who investigated speech production through masks in mandarin and English found that indeed different effects are present in different languages [10]. A solution to this issue is the use of discrete short sounds such as isolated vowels or fricatives which reduce these language differences. This focus may, in turn, limit the scope too much, as natural speech involves longer units of words and sentences which may alter the findings of changes in the artificially, unnaturally-isolated vowels. The above-mentioned studies differ in their speech units, which vary between isolated vowels and natural speech.

The current study investigates the intermediate case, where vowels are studied, but within their use in normal speech, where the vowels are co-articulated with consonants. Our study aims to provide a comprehensive understanding of how masks affect speech in real-life scenarios, which is crucial

***Corresponding author: Vered Aharonson,** School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, 1 Jan Smutslaan, Braamfontein 2050, ZA and Medical School, University of Nicosia, 93 Agiou Nikolaou Street, Engomi 2408, CY. email: vered.aharonson@wits.ac.za
Craig S. Carlson, Department of Electrical Engineering and Automation, Aalto University, Espoo, FI; Department of Biomedical Technology, Faculty of Medicine and Health Technology, Tampere University, Tampere, FI and School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, Braamfontein, ZA.
Michiel Postema, Department of Biomedical Technology, Faculty of Medicine and Health Technology, Tampere University, Tampere, FI and School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, Braamfontein, ZA.
Hanna Putter-Katz, Simona Tetin Schneider, Department of Communication Sciences and Disorders, Faculty of Health Professions, Ono Academic College, Kiryat Ono, IL.

for optimising hearing aids and communication strategies for the general population, not just those accustomed to wearing masks. The acoustic features studied are the ones identified in the previous speech through mask studies, as well as in speech production models as cues that affect speech perception. In addition this study considered the features in the context of their articulation type and place in the mouth [14]. Whereas previous studies investigated different types of face masks, our study focus on surgical masks due to their wide accessibility to study participants and to the population at large.

2 Materials and methods

Data were recorded from volunteering students at Ono Academic College (Kyriat Ono, IL). The study protocol was approved by the IRB committee of the participating institutes and all participants signed an informed consent. The participants inclusion criteria were native Hebrew speakers, non-smokers and persons with no speech or hearing disorder. The participants included 45 male speakers, age 28 ± 4. The participants provided their sex and age and received a written recording protocol and online guidance, upon which they recorded themselves while sitting, using the native recording application on their cellular phones, which was propped on a table in front of their mouth. The recording session included other speech utterances beyond the 10 vowels used for this preliminary study. The duration of the recording of the vowels less than 2 minutes. The protocol was repeated in two sessions, one with and one without masks, with a break of 5 minutes and the participants were instructed to drink a glass of water between the two sessions. The speech tasks were the pronunciation of five vowels: [a], [e], [i], [o], and [u], pronounced as AH (as in "hard"), EH (as in "bed"), EE (as in "see"), OH (as in "fork") , and OOH (as in "boot"), respectively. The vowels were first articulated as single, standalone sounds, and then in co-articulation with consonants. The co-articulation instances were taken from words in the frequently used and phonetically-balanced reading passage the "thousand islands". The words used were [e]lef, ha'[i]yim, ba'[o]ki[a]nus, and mech[u]sim for the vowels [e], [i], [o], [a], and [u], respectively. The recordings included the isolated and coarticulated vowels in one file. Each vowel was cropped and saved in a separate file for further analysis. Vocal features were extracted for each segment using MATLAB® (The MathWorks, Inc., Natick, MA, USA). The feature set included the segment's length, intensity, jitter, shimmer, pitch (F0) and the three first formants: F1, F2, F3 [16]. Short contextual definitions are as follows.

- Jitter: A voice quality feature, associated with the frequency variations in the voice signal.
- Shimmer: A voice quality feature, associated with the amplitude variations in the voice signal.
- Pitch (F0): Vocal folds rate of vibration during phonation. Faster vibrations result in a higher perceived pitch.
- F1: The lowest frequency band of resonance in the human vocal tract, associated with the vertical dimension of the tongue's position in the mouth. Lower tongue position results in a higher F1.
- F2: The second lowest frequency band of resonance in the human vocal tract, associated with the horizontal tongue's position in the mouth. Forward tongue position results in a higher F2.
- F3: The third lowest frequency band of resonance in the human vocal tract, associated with the shape of the tract and the position of the lips. With F1 and F2, it changes vowel qualities and plays a crucial role in speech intelligibility.

We employed a MATLAB® script to examine the feature values for potential outliers using box plot analyses. Outliers were defined based on the interquartile range (IQR) method : any data point that fell below Q1 −1.5IQR or above Q3 +1.5IQR was considered an outlier, where Q1 and Q3 represent the first and third quartiles, respectively. The outliers were verified across the all the values in the mask and no-mask, isolated and co-articulated speech conditions. Differential features were computed as quantifiers of the mask effect, by subtracting each feature value in a speaker's mask condition with its corresponding feature in the speaker's no-mask condition. This computation was performed for all features for the vowels in isolation and for the vowels in co-articulation. A paired t-test was performed to evaluate the difference between the isolated and co-articulated vowels conditions, for each one of the differential features. Despite the absence of evidence for a Gaussian distribution, we employed the paired t-test for our preliminary analysis due to its simplicity and widespread acceptance, while acknowledging that it provides a reasonable approximation for large sample sizes.

3 Results

Table 1 presents the the t-test values of statistically significant changes in vocal features due to a mask affecting isolated and co-articulated vowels, respectively. The plus and minus signs in the t values denote an increase or decrease in the difference between the vowels articulated with and without a mask. A statistically significant level of $p<0.001$ can be seen in the

Feature	[a]	[e]	[i]	[o]	[u]	[a]	[e]	[i]	[o]	[u]
Length	+0.16	+0.18	+0.81	–0.17	+0.53	–1.26	+2.01	–1.85	+1.65	+1.16
Intensity	+2.33*	–1.06	–1.19	–1.85	–0.55	–1.86	+2.08*	+1.17	–0.81	+0.51
Jitter	–0.68	+0.21	+0.11	+0.81	+0.26	–0.73	–0.21	+0.04	+0.34	–0.74
Shimmer	–2.56**	–0.15	–13.99***	+1.40	+0.82	+0.63	–1.87	–1.40	–2.06*	+1.11
F0 (Pitch)	+4.56***	+2.88**	+16.23***	+0.51	+0.92	+0.81	+1.19	+2.02*	+1.82	+0.14
F1 (1st formant)	+3.78***	+2.15*	–0.13	+0.38	+2.24*	+0.17	+1.11	–0.68	–0.46	–1.22
F2 (2nd formant)	–0.43	–2.07*	–0.68	+0.78	+1.89	+0.88	–0.45	+1.44	–1.37	–1.90
F3 (3rd formant)	–2.55*	+1.03	+1.44	–0.92	+0.66	–0.11	+0.05	+0.11	–1.19	–0.97

Tab. 1: T-test values for change in the features due to mask in isolated vowels (left column) and in co-articulated vowels (right column). Statistically significant values are stated in red, where *, **, and *** denote $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

pitch of the isolated vowels /a/, /e/ and /i/, the F1 format of the isolated /a/ and the shimmer of the isolated /a, /i/, when participants wore a mask. A statistical significance change at a level of p<0.05 can be seen in the formants F1 and F2 of the isolated vowel /e/ and in the F3 formant of the isolated /a/. No statistically significant changes are seen for the vowel /o/ and the vowel /u/ shows a single significant change (p<0.05) in the first formant F1. In the coarticulated vowels tests, only significant levels of p<0.05 can be seen, in the intensity of the vowel /e/ , the pith of the vowel /i/ and the shimmer of the vowel /u/ when participants wore a mask. There are 11 statistically significant cases where a mask affects isolated vowels, compared to three cases for co-articulated vowels. In eight of the cases, a significant effect appears in the isolated vowel where no effect appears in the respective co-articulated vowels. Two cases show a significant effect on the co-articulated vowels only and in one case the effect occurs in both isolated and co-articulated vowels. The significance of the latter three cases, is lower (p<0.05) than the significance shown in the isolated vowels and the t-values of these three cases are the lowest in the table.

4 Discussion and conclusions

Face masks have been and are here to stay in many professions such as medicine. In these conditions it is important that speech understanding, of both human and machine, as well as for the hearing impaired, will not be affected. Our preliminary results provided a distinction between the effect of masks in the unnatural, isolated vowels condition and the effect when those vowels are co-articulated within words. The results conveyed that while sporadic effects due to masks in co-articulated vowels exist, most mask effects in isolated vowels features are reduced or eliminated when the same vowels are co-articulated. while isolated vowels — an artificial experimental condition — undergo changes when speakers wear a surgical mask, the changes are reduced or eliminated for vowels within words, with the influence of adjacent sounds and their co-articulations with consonants. The results of the effects on co-articulated vowels enabled an explanation of some mixed results of previous studies on changes of speech acoustics parameters when speakers wear a surgical mask. No previous studies reported a change in vowel length when speakers wore masks. Our results confirm this finding for both isolated and co-articulated vowels. Gama reported an increase in speech intensity due to mask, and hypotesised that speakers when aware of the mask on their face reflexively raise their voice intensity [9]. Our results convey an increase due to masks in the vowel [a], when isolated and in the vowel [i] when co-articulated. Thus we confirm the latter two studies' findings, and further specify the vowels and conditions that portray the increase. Prior studies reported an increase in the pitch and the two first formants, F1 and F2 when speakers wore masks. Our results confirm this trend for the pitch and add the specific vowels where this increase occurred: in the vowels [a], [e] [i] when isolated and in the vowel [i] on,y when co-articulated [12, 13]. Moreover, the significance of the change for the co-articulated was smaller (p<0.05) than in the isolated vowel (p<0.001), and in the isolated [a] and [e] (p<0.0001 and p<0.01, respectively). This finding implies that the mask effect on pitch was reduced or eliminated in the co-articulated vowels. The F1 shows an increase for the vowels [e] and [u] but an increase in [a], all only for the isolated vowels, while no significant change in any co-articulated vowels. Corey found significant reduction in F3 for speakers wearing masks, and explained it by the low pass filter property of the mask that attenuates this higher formant's frequency [11]. Our results convey a decrease but only for the vowel [a] and only when isolated. A prior study reported no changes in either jitter or shimmer values masks [8], whilst another noted decrease in both jitter and shimmer [13]. Our results conveyed no significant changes in the jitter values, but found significant changes in the shimmer for the isolated vowels [a] and [i] and for the co-articulated [o]. The latter reduction had a smaller t values and a lower significance (p<0.05), however, compared to the isolated vowels changes (p<0.01 and p<0.001). All the significant changes due to mask occurred in the isolated vowels [a], [e], and [i]. These vowels are articulated in the rear part of the mouth [e] and [i] or in the centre [a] as opposed to the vowels [o] and [u] which are articulated at the forward part of the mouth, close to the lips [14]. The effect of the mask may thus be assigned to a physical blocking of the sound wave that propagated through some length, in the mouth cavity, before hitting the mask barrier. In comparison, the [o] and [u] which were articulated close to the mask penetrated through the barrier with less effect. The results are limited to the preliminary dataset used in this study. A larger cohort of speakers need to be collected to further validate these finding. Female speakers and more age groups of both sexes need to be recorded in order to assess sex and age differences for the vowels and masks conditions. Moreover, a pertinent aspect when collecting data at participants' homes, and in this case, recording them using their personal mobile phones, may induce variability due to phone type. Specifically, filters embedded in the mobiles phones' native recording applications may smooth out subtle differences in the speech variables, such as pitch and these filters differ across phone models. Future analysis should entail a dataset containing larger samples for each mobile phone types. The study focus on features extracted from the speech and thus pertain to speech production acoustics. Complementary studies should examine this acoustics in listening experiments and investigate how these changes affect the intelligibility of vowels to the listeners [15]. Our results on the case of co-articulated vowels confirm, however, the earlier findings that reported no detrimental effect on speech understanding in normal-hearing or hearing-impaired individuals [6]. It is noted that the difference between more than one co-articulated vowel in a word vs. the same vowels co-articulated in two words, limiting the scope of this study.

Our results convey an optimistic message, that in the course of natural speech, when vowels are articulated within words and sentences, the acoustical changes due to mask wearing are negligible and may not affect speech understanding. The finding that the acoustic differences are smaller in co-articulated vowels compared to isolated vowels provides new insights into how masks affect everyday communication, which has significant implications for the design of hearing aids and other auditory technologies.

Author Statement

This work was not supported financially. Authors state no conflict of interest. Informed consent has been obtained from all individuals included in this study. The research related to human use complied with all the relevant national regulations and institutional policies and has been approved by the IRB committee (certificate number 29077).

References

[1] Jefferson T, Foxlee R, Del Mar C, Dooley L, Ferroni E, Hewak B, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses: systematic review. BMJ 2008;336:77–80.
[2] Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020;395:1973–87.
[3] Alkharabsheh AA, Aboudi O, Abdulbaqi K, Garada S. The effect of wearing face mask on speech intelligibility in listeners with sensorineural hearing loss and normal hearing sensitivity. Int J Audiol 2023;62:328–33.
[4] Moradi S, Lidestam B, Danielsson H, Ng EHN, Rönnberg J. Visual cues contribute differentially to audiovisual perception of consonants and vowels in improving recognition and reducing cognitive demands in listeners with hearing impairment using hearing aids. J Speech Lang Hear Res 2017;60:2687–2703.
[5] Fant G. The source filter concept in voice production. STL-QPSR 1981;22:21–37.
[6] Mendel LL, Gardino JA, Atcherson SR. Speech understanding using surgical masks: a problem in health care?. J Am Acad Audiol 2008;19:686–95.
[7] Goldin A, Weinstein B, Shuman N. How do medical masks degrade speech perception? Hearing Rev 2020;27:8–9.
[8] Magee M, Lewis C, Noffs G, Reece H, Chan JCS, Zaga CJ, et al. Effects of face masks on acoustic analysis and speech perception: implications for peri-pandemic protocols. J Acoust Soc Am 2020;148:3562–68.
[9] Gama R, Castro ME, van Lith-Bijl JT, Desuter G. Does the wearing of masks change voice and speech parameters? Eur Arch Otorhinolaryngol 2021;279:1701–8.
[10] Nguyen DD, McCabe P, Thomas D, Purcell A, Doble M, Novakovic D, et al. Acoustic voice characteristics with and without wearing a facemask. Sci Rep 2021;11:5651.
[11] Corey RM, Jones U, Singer AC. Acoustic effects of medical, cloth, and transparent face masks on speech signals. J Acoust Soc Am 2020;148:2371–75.
[12] Georgiou GP. Acoustic markers of vowels produced with different types of face masks. Appl Acoust 2022;191:108691.
[13] Lin Y, Cheng L, Wang Q, Xu W. 2023. Effects of medical masks on voice assessment during the COVID-19 pandemic. J Voice 2023;37:802.e25–9.
[14] Pandey PC, Shah MS. Estimation of place of articulation during stop closures of vowel–consonant–vowel utterances. IEEE Trans Audio Speech Lang Process 2009;17:277–86.
[15] Muzzi E, Chermaz C, Castro V, Zaninoni M, Saksida A, Orzan E. Short report on the effects of SARS-CoV-2 face protective equipment on verbal communication. Eur Arch Otorhinolaryngol 2021;278:3565–70.
[16] Raphael LJ, Borden GJ, Harris KS. Speech Science Primer: Physiology, Acoustics, and Perception of Speech. 5th ed. Baltimore: Lippincott Williams & Wilkins; 2007.