FEBRUARY 16 2023

Speaking with mask in the COVID-19 era: Multiclass machine learning classification of acoustic and perceptual parameters

F. Calà; C. Manfredi; L. Battilocchi; ... et. al

J Acoust Soc Am 153, 1204–1218 (2023)
https://doi.org/10.1121/10.0017244

JASA ARTICLE

Speaking with mask in the COVID-19 era: Multiclass machine learning classification of acoustic and perceptual parameters

F. Calà,[1,a)] C. Manfredi,[1] L. Battilocchi,[2,b)] L. Frassineti,[1,c)] and G. Cantarella[2,b)]
[1]*Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy*
[2]*Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy*

ABSTRACT:
The intensive use of personal protective equipment often requires increasing voice intensity, with possible development of voice disorders. This paper exploits machine learning approaches to investigate the impact of different types of masks on sustained vowels /a/, /i/, and /u/ and the sequence /a'jw/ inside a standardized sentence. Both objective acoustical parameters and subjective ratings were used for statistical analysis, multiple comparisons, and in multivariate machine learning classification experiments. Significant differences were found between mask+shield configuration and no-mask and between mask and mask+shield conditions. Power spectral density decreases with statistical significance above 1.5 kHz when wearing masks. Subjective ratings confirmed increasing discomfort from no-mask condition to protective masks and shield. Machine learning techniques proved that masks alter voice production: in a multiclass experiment, random forest (RF) models were able to distinguish amongst seven masks conditions with up to 94% validation accuracy, separating masked from unmasked conditions with up to 100% validation accuracy and detecting the shield presence with up to 86% validation accuracy. Moreover, an RF classifier allowed distinguishing male from female subject in masked conditions with 100% validation accuracy. Combining acoustic and perceptual analysis represents a robust approach to characterize masks configurations and quantify the corresponding level of discomfort. © *2023 Acoustical Society of America.*
https://doi.org/10.1121/10.0017244

(Received 9 August 2022; revised 23 January 2023; accepted 26 January 2023; published online 16 February 2023)
[Editor: James F. Lynch] Pages: 1204–1218

I. INTRODUCTION

The intensive use of personal protective equipment (PPEs), social distancing, and isolation represent the most important strategies that the World Health Organization (WHO) and local governments put in place in response to the SARS-CoV-2 pandemic outbreak to reduce the spread of the virus. However, as a major negative consequence, communication between individuals was deeply affected, both as far as the quality and the intelligibility of voice are concerned. This aspect is particularly stressed in occupational voice users, i.e., people whose job demands a higher-than-normal voice load[1] such as actors, singers, teachers, and healthcare professionals. This latter category of subjects wears face masks for most of the working time to protect themselves from possible contaminations, especially in otolaryngology departments where visual investigations of the vocal tract and the phonatory apparatus can cause patients' sneezing and coughing. Moreover, wearing face masks in a clinical setting further compromises the already difficult human interaction in noisy environments and especially for patients affected by hearing impairments. Indeed, besides acoustic alterations, PPEs significantly reduce visual cues and feedbacks as they prevent lip-reading and limit facial expression interpretation: as a consequence, the listener has to give more attention to correctly perceive the emitted utterances while the speaker usually needs to increase voice intensity, which can lead to hyper-functionality and possible development of voice disorders

The main types of PPEs are as follows:

- Surgical masks. They are fluid resistant and consist of the overlap of three layers of non-woven fabric. Their design fits loosely on the face and WHO recommends their use in low risk situation, as they reduce the spread only of large droplets.
- FFP2 masks (or N95) are designed to be tightly attached to the wearer face and prevent the inhalation of smaller airborne particles. They should be used only in high-risk contexts.
- FFP3 masks are characterized by a valve-filter system that further reduces small particle inhalation with respect to FFP2 masks.
- Visors and windowed masks. They extend the protection from coronavirus-carrying droplets of the nose and mouth to the eyes. In particular, transparency makes are particularly advisable when communicating with deaf people or patients affected by hearing loss.

In Swanepoel *et al.*,[2] surveys and semi-structured online interviews have highlighted that for healthcare professionals

[a)]Electronic mail: federico.cala@stud.unifi.it
[b)]Also at: Department of Otolaryngology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.
[c)]Also at: Department of Medical Biotechnologies, Università di Siena, 53100 Siena, Italy.

https://doi.org/10.1121/10.0017244

the use of PPE is one of the most common causes of disrupted speech communication. Robeiro *et al.*,[3] with the same methodology applied on more than 400 adults, found out that communication while wearing PPEs has determined an increase in vocal effort, difficulty in coordinating speaking and breathing, and reduction of auditory feedback. These symptoms were underlined as well in several works cited in Ref. 2. From a socio-behavioral point of view, the use of surgical masks had little to no impact on speech understanding even in noisy environments for both healthy and acoustically impaired subjects. Nevertheless, at the expense of voice quality, the use of a face shield or a mask provided with a clear window instead of opaque coverings helped to improve lip reading with these individuals and cochlear implant users and prevents increase in concentration efforts and decrease in confidence.[2,4,5]

In order to objectively evaluate such conditions, over the last two years researchers' efforts were focused on the acoustic analysis of phonation and articulation tasks while wearing different masks configurations. Porschmann *et al.*[6] investigated the influence of masks on sound radiation, reporting how their use leads to a significant loss of transmission at frequencies above 2 kHz. In the article by Goldin *et al.*[7] vocal samples were measured according to the type of the mask worn: the authors found out that each mask acts as a low-pass filter by attenuating high frequencies in the 2–7 kHz bandwidth, which is critical for most vowels and fricative consonants recognition and understanding.[8] The attenuation was from 3–4 dB for the surgical mask up to 12 dB for the N95 mask. These results were confirmed in the work of Corey *et al.*[9] and Balamurali *et al.*:[10] both have used various types of masks including cotton masks and transparent visors, but the latter replaced real speakers with a dummy head mounted with a loudspeaker at its mouth to generate a broadband signal.

Magee *et al.*[11] evaluated the impact of different types of masks (surgical, cotton, N95) on the acoustic output and speech perception. Significant differences were found between the masked and unmasked signals in the power spectral density distribution for frequencies above 3 kHz. Moreover, significant differences were found between the various protection devices in the average time of pauses in the reading task. No significant differences were highlighted by Cavallaro *et al.*,[12,13] who investigated the impact of the surgical mask on speech parameters F0, Jitter, Shimmer, and HNR on a group of subjects who were asked to repeat the vowel /a/ for the maximum phonatory time with and without the surgical mask. The same task was applied by Lin *et al.*[14] that evaluated F0, jitter, and HNR: a decrease in jitter and an increase in HNR were reported when wearing surgical and KN95 masks, though these deviations were not significant. The authors stated that these alterations depend on the level of mask fitting on wearer's face as they were more evident in KN95 and could be associated with an adjustment of speaking habits. Similar results were obtained by Gojayev *et al.*,[15] where a sustained /a/ emitted by control subjects showed significant differences in HNR values when wearing a FFP3 mask. Joshi *et al.*[16] analyzed F0, formants F1-F3, and cepstral peak prominence (CPP) of sustained vowels /a/ and /i/ with five masks configurations: no significant difference was reported, but for F2 (only in male subjects) and CPP (in both genders) the presence of a face shield worn above a surgical mask together with a KN95 determined a statistical difference with the two masks used alone.

In hospitals, communication can even be harder as the protocols for healthcare personnel often require the use of two devices to be worn together: one for the respiratory tract and one for the eyes. Bandaru[17] focused on the analysis of the effects of the N95 mask and the visor among healthcare personnel aged between 20 and 60, without PPE and wearing N95 and visor at the same time. The results show a worsening of the intelligibility of the voice which makes the interactions between clinicians and patients more difficult. Finally, a study by McKenna *et al.*[18] has underlined that no statistical differences were found out in various acoustic parameters (F0, F1, F2, HNR, jitter, shimmer, CPP) in sustained vowels, words, sentences and in the "Rainbow Passage" before and after a working day for mask-wearing healthcare professionals.

This paper is aimed at investigating how the use of different types of masks, worn alone or together with the use of a protective visor, may affect the emitted sound. In contrast to previous studies, we evaluated whether the use of different types of PPEs caused alterations in sustained vowels and/or in a sequence of vowels, combining both acoustic and perceptual parameters: specifically, we analyzed the sustained vowels /a/, /i/, and /u/ and the vowel sequence /a'jw/ inside an almost vocalic sentence emitted by Italian speakers.

Voice quality can be evaluated non-invasively: with perceptual scales (like GIRBAS[19]), subjective scales (like VHI[20]), objective acoustic indexes computed with dedicated software tools such as MDVP,[21] PRAAT,[22] BIOVOICE,[23] etc., or with minimally invasive devices (such as electroglottography[24]). In this work only non-invasive measures are considered, based on the acoustic analysis of the signal obtained with BioVoice and on self-perceptual evaluation based on a specific questionnaire developed in analogy to similar works.[3,25]

Section II illustrates the recording procedure, signal processing methods and the applied statistical analysis applied separately to female and male subjects. A set of multiclass artificial intelligence (AI) experiments was developed to understand whether both acoustic and subjective parameters are able to distinguish between masks configurations and masked-unmasked conditions. To the authors' knowledge, this is the first attempt to perform such an automatic decision task mixing objective and subjective parameters. Section III shows the results of statistical analysis and of classifiers for males and females. Discussion and conclusions are presented in Secs. IV and V.

II. MATERIALS AND METHODS

A. Recordings

The study examines ten adult subjects, five female and five male (age range 25–30 years for both groups) with: mean = 27.8 years, std = 0.836 years; mean = 26.8 years,

std = 1.923 years, respectively. They were asked to utter (in Italian) three sustained vowel sounds (/a/, /i/, and /u/) and a sentence of particular interest for acoustic analysis as it is rich in vocalic sounds: /il bam'bino 'ama 'le a'jwɔle 'del:a 'mam:a/ ("the child loves mother's flower beds"). Vowels /a/, /i/, and /u/ represent Italian cardinal vowels as they are characterized by well-defined vocal tract configurations which make their phonation quite stable regardless of dialectal inflections.[26] The recording of each sound was repeated three times for each subject at conversational tone and intensity in the following 7 different PPE configurations, for a total of 210 voice signals:

(a) Absence of PPE (baseline)
(b) Surgical mask
(c) Surgical mask and visor (surgical + shield)
(d) FFP2 mask (FFP2)
(e) FFP2 mask and visor (FFP2 + shield)
(f) FFP3 mask (FFP3)
(g) FFP3 mask and visor (FFP3 + shield)

Recordings were made in a non-protected but controlled (quiet) environment after a working day inside the Ospedale Maggiore Policlinico Milano, Milano, Italy. Subjects are Italian speaking otolaryngologists working in the hospital.

The signals were manually segmented using the AUDACITY audio editor software. To avoid transient in the sustained vowels only the "stable" central part of /a/, /i/, and /u/ was selected (mean duration /a/ = 1.7 s, std /a/ = 0.8 s; mean duration /i/ = 1.7 s, std /i/ = 0.9 s; mean duration /u/ = 1.6 s, std = 0.9 s). The vocalic sequence /a'jw/ was selected from the word /a'jwɔle/ of the standardized sentence to perform a specific analysis on articulation in continuous speech (mean duration = 0.42 s, std = 0.09 s). Recordings were made with the Voice Recorder built-in app of a Samsung smartphone model A50[27]: the distance between the integrated microphone and the mouth was kept constant at 15 cm and with 45° inclination to reduce lateral distortions.[28] Audio files were recorded in .m4a format at 44.1 kHz sampling frequency and 128 kbps bitrate; .m4a is a lossy digital audio compression format that quantizes the signal on the basis of a psychoacoustic model. Therefore, frequencies above 20 kHz are usually cut off.[29] Audacity software was used to check sampling frequency and convert files into .wav format.

B. Acoustic analysis

The acoustic analysis is performed with the BIOVOICE open-source software tool.[23] In contrast to other tools, BioVoice automatically selects proper frequency ranges for the analysis of adults, children, and newborns voices. For adults, the gender of the subject must be specified: male or female. BioVoice performs both time and frequency analysis and estimates more than 20 acoustic parameters with advanced and robust analysis techniques specifically developed.

In the time domain, the number, length and percentage of voiced and unvoiced segments (V/UV) and jitter are detected and saved in an Excel table. In the frequency domain, the fundamental frequency F0, formant frequencies F1–F3, and the noise level are estimated.

Jitter J is computed applying Eq. (1),

$$J = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i}, \tag{1}$$

where N represents the number of considered time windows and T_i is the fundamental period (the reciprocal of F0) in the i-th window. For F0 and for each formant, the mean, median, standard deviation, maximum, and minimum values are calculated. Moreover, the power spectral density (PSD) is computed in the frequency range of each gender and normalized with respect to its maximum value; therefore, the range is 0 dB downward: this allows comparison among different PSDs. For statistical analysis, in this work the PSD frequency spectrum was divided into 500 Hz-wide intervals where the average power is computed.[30] Noise variations are tracked by means of an adaptive version of the normalized noise energy method, named adaptive normalized noise energy (ANNE), which relies on a comb filtering approach optimized to deal with data windows of varying length. Large negative ANNE values (in dB) correspond to good voice quality, while values close to zero reflect the presence of strong noise.[31]

According to similar studies[12,14,16] only the mean values of sustained vowels were considered. Jitter and NNE were included to analyze phonation irregularities and noise levels. Formants F1–F3 were computed and used to evaluate the following specific parameters related to articulation.

Formant ratios: Introduced by Shapir *et al.*[32] to further analyze tongue movements.

$$\frac{F1a}{F1i}, \tag{2}$$

$$\frac{F1a}{F1u}, \tag{3}$$

$$\frac{F2i}{F2u}, \tag{4}$$

where Fxy denotes the x-th formant of vowel y (x = 1, 2, 3; y = a, i, u). Equations (2) and (3) are more sensitive to vertical tongue movements, whereas Eq. (4) is used to study horizontal tongue movements.

Vowel space area (VSA):[33] Corresponds to the area of the vowel triangle and represents an important measure to monitor articulatory workspace and detect possible articulation difficulties. VSA is shown in Eq. (5),

$$VSA = 0.5 \times \left|F1i \times (F2a\text{–}F2u) + F1a \times (F2u\text{–}F2i) + F1u \times (F2i\text{–}F2a)\right|. \tag{5}$$

Formant centralization ratio (FCR): Proposed by Sapir *et al.*[34] as a normalization procedure that maximizes sensitivity to

https://doi.org/10.1121/10.0017244

TABLE I. Acoustic parameters considered for sustained vowels /a/, /i/, and /u/.

Feature	Description
F0 mean /a/	Mean fundamental frequency of /a/
F0 mean /i/	Mean fundamental frequency of /i/
F0 mean /u/	Mean fundamental frequency of /u/
Jitter /a/	Jitter (in %) for /a/
Jitter /i/	Jitter (in %) for /i/
Jitter /u/	Jitter (in %) for /u/
NNE /a/	Normalized noise energy of /a/
NNE /i/	Normalized noise energy of /i/
NNE /u/	Normalized noise energy of /u/
F1a/F1i	Formant ratio between F1 mean /a/ and F1 mean /i/
F1a/F1u	Formant ratio between F1 mean /a/ and F1 mean /u/
F2i/F2u	Formant ratio between F2 mean /i/ and F2 mean /u/
VSA	Vowel space area
FCR	Formant centralization ratio

formant centralization degree and minimizes intra- and inter-variability. It is shown in Eq. (6),

$$FCR = \frac{F1i + F1u + F2a + F2u}{F1a + F2i}. \quad (6)$$

Table I summarizes the 14 acoustic parameters used in this work for sustained vowels.

As for sustained vowels and in line with literature, for the vocalic sequence /a'jw/ only the mean values of F0, F1, and F2 were considered.[35] Jitter was excluded as it is less suitable for quantifying glottal cycles fluctuations during articulation changes.[36] The PSD divided into 500 Hz-wide intervals allowed performing a detailed analysis of the spectral energy. Table II displays the 15 parameters used for the statistical analysis of /a'jw/.

Thus, a total of 29 parameters were considered.

C. Statistical analysis

Statistical analysis was performed, aimed at finding whether one or more acoustic parameters are significantly different from the baseline (no mask) and to highlight possible alterations brought up by adding a shield over the face mask. A preliminary Shapiro-Wilk test was carried out to decide whether to apply a one-way ANOVA test or its corresponding nonparametric Kruskal-Wallis test. In case of a statistically significant level ($p < 0.05$), *post hoc* multiple comparison was performed using t-test and Tukey correction method or Dunn-Bonferroni test to compare mask configurations with each other.

TABLE II. Acoustic parameters considered for the vocalic sequence /a'jw/.

Feature	Description
F0 mean	Mean fundamental frequency
F1 mean	Mean first formant
F2 mean	Mean second formant
NNE	Normalized noise energy
FR I	Mean PSD in frequency range 0–500 Hz
FR II	Mean PSD in frequency range 500–1000 Hz
FR III	Mean PSD in frequency range 1000–1500 Hz
FR IV	Mean PSD in frequency range 1500–2000 Hz
FR V	Mean PSD in frequency range 2000–2500 Hz
FR VI	Mean PSD in frequency range 2500–3000 Hz
FR VII	Mean PSD in frequency range 3000–3500 Hz
FR VIII	Mean PSD in frequency range 3500–4000 Hz
FR IX	Mean PSD in frequency range 4000–4500 Hz
FR X	Mean PSD in frequency range 4500–5000 Hz
FR XI	Mean PSD in frequency range 5000–5500 Hz

D. Subjective ratings

An innovative aspect of the present study concerns the combination of subjective indices and objective parameters to train machine learning models. To this aim, an *ad hoc* questionnaire was administered to the participants. It concerns several aspects related to possible discomfort occurring when wearing masks during a working day: in this case discomfort generically referred to the overall perceived intensity, effort or fatigue during vocal emissions. Answers were given after task completion, and they are rated from 0 (no discomfort) to 4 (high discomfort). The eight questions focused on the following:

(1) Self-perceived vocal effort.
(2) Self-perceived voice alteration.
(3) Difficulty in voice projection.
(4) Difficulty in being understood by patients.
(5) Difficulty in being understood by colleagues in the ward or in the outpatient department.
(6) Difficult communication with colleagues in the operating room.
(7) Self-perceived speech articulation effort.
(8) Self-perceived hyperarticulation required for intelligibility.

Each question was asked for each of the 7 PPEs as well as for the baseline. Thus, the questionnaire is made of 56 questions for a total of 560 answers.

E. AI—Machine learning

Machine learning based voice assessment (MLVA) represents an effective tool with applications spreading over several fields such as the detection of phonatory apparatus disorders, neurological problems, cardiovascular diseases, diabetes, etc., and recently in COVID-19 detection as well.[37] As compared to conventional methods, MLVA allows simultaneous analysis of high-dimensional data through properly trained algorithms; furthermore, machine learning techniques can identify possible relationships between objective and perceptual parameters. In this study, k-nearest neighbors (KNN), support vector machines (SVM), and random forest (RF) classifier were chosen as they are among the most used models for acoustic analysis.[38] To carry out a detailed analysis, more parameters than those reported in Tables I and II were considered. Specifically, predictors concern: 24 parameters extracted from /a'jw/, the 11 PSD frequency ranges, 24 parameters from each corner vowel, 5 articulatory parameters, and the

https://doi.org/10.1121/10.0017244

FIG. 1. (Color online) Workflow of machine learning experiments. The five-steps procedure applied for each artificial intelligence experiment is depicted. After z-score normalization, three supervised classifiers were trained and hyperparameters tuning was performed with Bayesian optimization. Models were validated with K-fold cross-validation (K = 10).

8 self-perceptual questions, for a total of 120 features. Due to low numerousness of the dataset, all observations made the whole training set: to limit overfitting and obtain reliable prediction, trained models were cross-validated with leave-one-subject-out (LOSO) method. Bayesian optimization, with 60 iterations,[39] was carried out to maximize global accuracy: modifications were made to avoid the choice of $k = 1$ for KNN and minimum leaf size = 1 for RF. At each iteration, results were saved in an array and the outcome was obtained by averaging them. A MATLAB® 2020b code[40] was developed to calculate, for each class: recall, specificity, precision, F1-score, accuracy and the area-under-the-curve, i.e., the underlying area of the ROC curve (AUC). Global accuracy is determined as well. Figure 1 summarizes the used AI workflow applied in this paper.

III. RESULTS

Separate analyses were carried on for female and male voices both for sustained vowels and the vocalic sequence /a'jw/; results are presented in separate sections. As female subjects are characterized by higher F0 and formant frequencies, Z-score normalization was performed.

To test possible differences between the acoustic parameters for the seven mask configurations, a one-way ANOVA test was performed for normally distributed parameters extracted from sustained vowels. Multiple comparisons with Tukey correction were performed to detect which mask condition presents significant differences in objective acoustic parameters as compared to the others. For F1a/F1u (only for the female group) and jitter /u/, the Shapiro-Wilk test rejected the null hypothesis of normal distribution, therefore in these cases Kruskal-Wallis (KW) and Dunn-Bonferroni's tests were performed.

In PSD analysis, data were normally distributed in each range, so only a one-way ANOVA test was performed and Tukey correction was applied in multiple comparisons.

A. Sustained vowels

Table III summarizes multiple comparison results for female and male groups; they are reported in the left and right side of the table, respectively. Black boxes represent statistically significant differences when acoustic parameters extracted from sustained vowels and uttered wearing PPEs are compared with respect to the baseline (no mask worn). Grey boxes represent statistically significant difference when acoustic properties differ between a mask type and the same mask type worn together with a shield (e.g., for female subjects F1a/F1i shows a significant difference between surgical mask and surgical mask + face shield). The symbol α denotes parameters tested with Kruskal-Wallis test. F-statistics are displayed for both groups and directions of effects are provided in the form of + and − signs, representing upward and downward differences respectively in regard to the mask condition. Table III shows that for female subjects the mask vs same mask + shield difference is more common, especially when surgical mask and FFP2 are concerned. In both genders, articulation seems to be compromised when more onerous mask configurations are worn (particularly FFP2 + shield).

Figure 2(a) shows the vowel triangles for the 7 PPE configurations for the female group: the baseline is represented by the solid line. In Fig. 2(b) the boxplots for the

TABLE III. Black boxes = statistically significant differences between PPEs and baseline; gray boxes = statistically significant differences between a mask type and the same mask type worn together with a face shield; α = Kruskal-Wallis tested acoustic parameter; signs +/− represent directions of effects; degrees of freedom = 6.

	Female							Male						
	F-statistic	Surgical	Surgical +shield	FFP2	FFP2 +shield	FFP3	FFP3 +shield	F-statistic	Surgical	Surgical +shield	FFP2	FFP2 +shield	FFP3	FFP3 +shield
Jitter /u/$^{\alpha}$	13.6						−							
NNE /a/								2.36				−		
NNE /i/	4.22		−			−								
NNE /u/	2.54		−											
F1a/F1i$^{\alpha}$	26.8	−	−				−	3.45				−		−
F1a/F1u	4.81	+	−				−	3.80						−
F2i/F2u	4.24			+	−			2.34				−		
VSA	7.65	+		−				5.53	−			−		−
FCR	5.68			−	+			3.03				+		

https://doi.org/10.1121/10.0017244

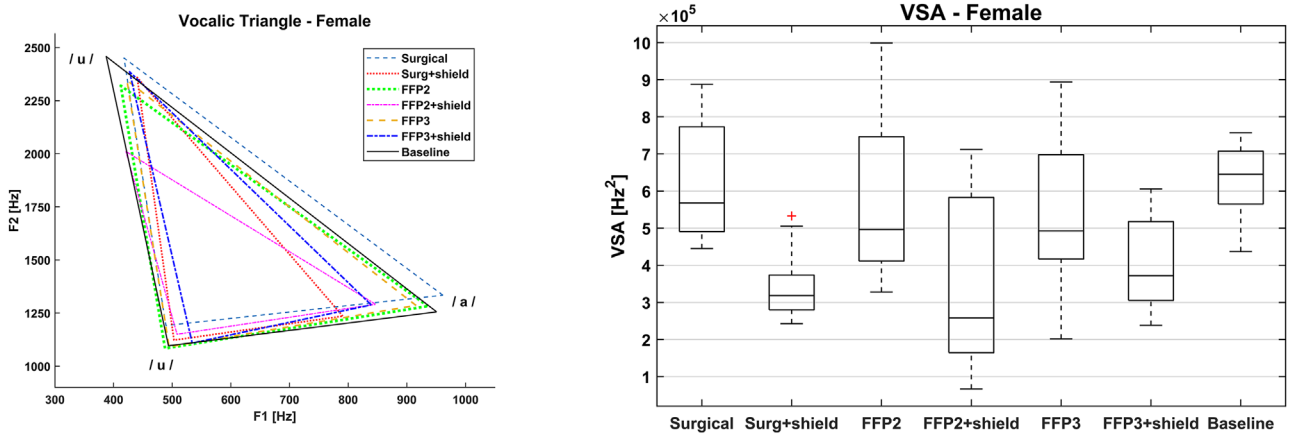

FIG. 2. (Color online) Articulatory results for the female group. (a) Vowel triangles for each PPE configuration; (b) VSA boxplots for each PPE configuration; surg = surgical.

VSA are displayed for the female group. Figure 2(a) shows that a face shield worn together with a PPE, especially FFP2 and FFP3, reduces the vocalic triangle area, which reflects a decrease in articulatory capabilities. This result is supported by boxplots in Fig. 2(b) as well.

Analogously for the male group vowel triangles and boxplots are shown in Figs. 3(a) and 3(b), respectively. As for female subjects, a relevant alteration of the vocalic triangle area can be observed for more onerous masks configuration, i.e., FFP2+shield and FFP3+shield.

F1 and F2 boxplots of each cardinal vowel both for female and male subjects are presented in Appendixes A and B, respectively.

B. Vocalic sequence /a'jw/

Figure 4 shows the PSD profile for female subjects whereas Fig. 5 concerns the male group. Slices of 500 Hz width are highlighted: a marked decrease in the PSD is observable for tighter PPEs such FFP2 masks, especially when used in combination with face shields, for both genders. Specifically, for female subjects a relevant decrease in PSD for more onerous configurations (FFP2+shield and FFP3+shield) starts above 3–3.5 kHz, whereas for male subjects it starts above 2–2.5 kHz.

Table IV illustrates 500 Hz-wide frequency ranges and possible statistically significant differences ($p < 0.05$), alongside with F-statistic values. Black boxes represent statistically significant differences when acoustic parameters extracted from sustained vowels and uttered wearing PPEs are compared with respect to the baseline (no mask worn). Grey boxes represent statistically significant difference when acoustic properties differ between a mask type and the same mask type worn together with a shield. Directions of effects are provided in the form of + and − signs, representing upward and downward differences respectively in regard to the mask condition. For both groups statistically significant differences of PSD with respect to the baseline condition are shown, starting from 2 kHz on. For female subjects differences are concentrated in the high frequency regions

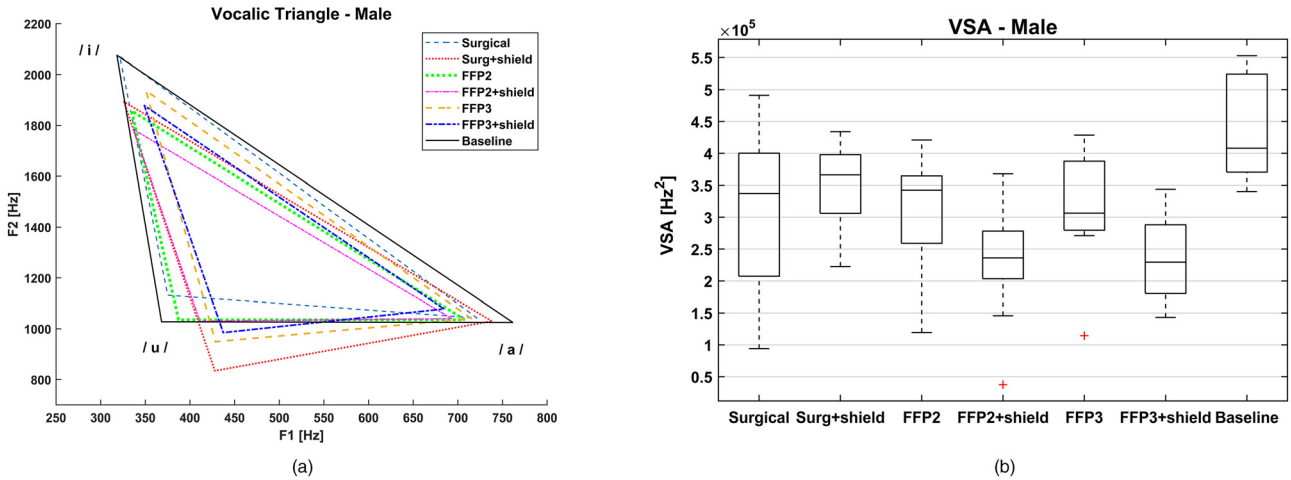

FIG. 3. (Color online) Articulatory results for the male group. (a) Vowel triangles for each PPE configuration; (b) VSA boxplots for each PPE configuration; surg = surgical.

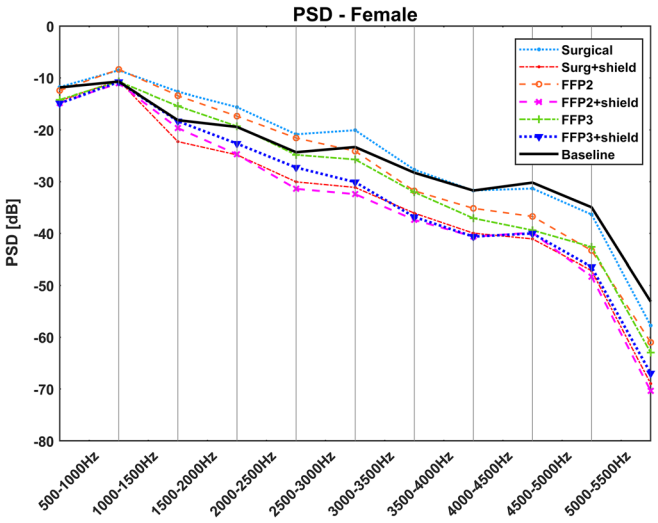

FIG. 4. (Color online) PSD plot for the female group. Average power over 500 Hz-wide intervals, normalized with respect to its maximum value; surg = surgical.

of the spectrum (>4 kHz) whereas, for male subjects differences are focused in the 2–4 kHz range. It is interesting to notice that PSD alterations with respect to the baseline exist only when a visor is worn together with a mask.

Regarding parameters extracted from /a'jw/ and highlighted in Table II, *post hoc* analysis detected a significant difference between FFP3+shield and baseline configurations only for F2 with p = 0,008.

C. Questionnaire

With reference to the 7 configurations (a)–(g) and questions 1–8 described above, results for female and male groups are reported in Table V. The table displays the mean values of the scores for all questions; standard deviation is reported in brackets. Table V shows an increasing trend for all ratings related to the use of more onerous PPE configurations for both genders, especially as far as articulation and intelligibility are concerned.

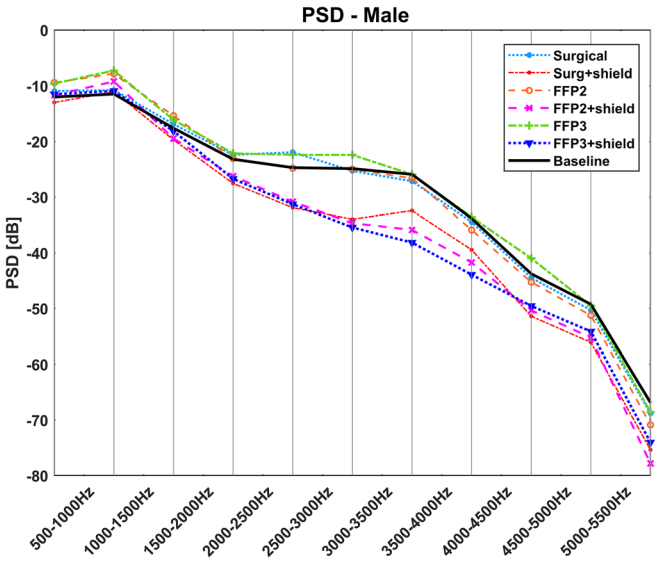

FIG. 5. (Color online) PSD plot for the male group. Average power over 500 Hz-wide intervals, normalized with respect to its maximum value; surg = surgical.

D. AI—Machine learning

Statistical analysis suggested that possible articulation irregularities exist due to the presence of face masks, especially when worn together with a face shield. To understand whether each PPE configuration determines a certain degree of voice alteration, KNN, SVM, and RF classifiers were automatically trained in sequence in the following cases using both objective and subjective measures:

(1) The seven different types of PPEs.
(2) Absence or presence of a face shield.
(3) Masked or unmasked conditions.
(4) Male and female voice features with PPEs.

For each experiment only the model with the highest validation accuracy was further analyzed.

In Tables VI and VII, performances of the first AI experiment are shown. They considered 120 features (24 parameters extracted from /a'jw/, 11 PSD frequency ranges, 24 parameters from each corner vowel, 5 articulatory parameters, and 8 self-perceptual questions) and the 7 classes (surgical mask, surgical mask+shield, FFP2, FFP2+shield, FFP3, FFP3+shield) for female and male groups, respectively. For both genders the best models belong to the RF group. Based on the number of correctly detected subjects, the evaluation metrics precision, recall, specificity, and F1 score are computed and displayed in percentage. Table VI shows that voice properties in female subjects seem to be modified differently when wearing various types of PPEs, since the RF classifiers is able to distinguish seven classes with high validation accuracy. Furthermore, recall (for surgical masks) and precision (for baseline condition) show lower values with respect to the baseline when a surgical mask is worn alone: this might suggest that this latter type of mask does not change relevantly voice quality. For males, similar considerations can be made, as shown in Table VII, but with a lower validation accuracy.

As face shields seem to compromise articulation, in the second AI experiment male and female groups were divided in two classes (absence or presence of the shield) and all the 120 features were considered. Best results were again obtained with RF. Table VIII displays the performances of the second experiment and provides a comparison between genders. In particular, percentages of precision, recall, specificity, F1-score and AUC are presented for the face shield absence (no shield) and presence (with shield) conditions. In line with statistical analysis results, machine learning was capable to detect voice properties alterations when wearing face shields, especially for the male group, as the RF classifier achieved 90% accuracy.

Tables VI and VII show that, even if with low numerousness, classifiers were able to correctly separate the baseline condition from those with masks. Therefore, in the third

https://doi.org/10.1121/10.0017244

TABLE IV. Black boxes = statistically significant differences between PPEs and baseline (no mask worn); gray boxes = statistically significant differences between a mask type and the same mask type worn together with a shield; signs +/− indicate directions of effects; degrees of freedom = 6.

	Female							Male						
	F-statistics	Surgical	Surgical +shield	FFP2	FFP2 +shield	FFP3	FFP3 +shield	F-statistics	Surgical	Surgical +shield	FFP2	FFP2 +shield	FFP3	FFP3 +shield
Range 0–500 Hz														
Range 500–1000 Hz														
Range 1000–1500Hz	2.94	+												
Range 1500–2000 Hz	5.23	+		+										
Range 2000–2500 Hz	6.19	+		+	−			8.01	+	−			+	−
Range 2500–3000 Hz	7.69	+	−	+	−			5.89			+	−	+	−
Range 3000–3500 Hz								6.82			+	−	+	−
Range 3500–4000 Hz								6.89				−	+	−
Range 4000–4500 Hz								5.48		−			+	
Range 4500–5000 Hz	5.19	+	−		−		−							
Range 5000–5500 Hz	8.85	+	−	+	−		−	4.09				−		

TABLE V. Questionnaire—Mean values for female group and for male group. Standard deviation is reported in brackets. *V* = vocalic, *Pat* = patient, *Col* = colleagues, *Int* = intelligibility, *OR* = operating room, D = difficulties.

	Baseline	Surgical	Surgical+shield	FFP2	FFP2+shield	FFP3	FFP3+shield
				Female			
V. Effort	0 (0)	0 (0)	1.2 (0.4)	1.2 (0.4)	2.4 (0.5)	2.4 (0.5)	3 (1)
V. Alteration	0 (0)	0.4 (0.5)	1.6 (0.5)	1 (0.7)	2 (0.7)	2.2 (0.4)	3.2 (0.4)
V. Projection	0 (0)	0.4 (0.5)	1.8 (0.4)	2 (1)	3 (0)	2.2 (0.4)	3.4 (0.5)
Pat. Int.	0 (0)	0.8 (0.4)	1.6 (0.9)	1.6 (1.1)	2.2 (0.8)	2.8 (0.4)	3.2 (0.4)
Col. Int. Ward	0 (0)	0.6 (0.5)	1.6 (0.9)	1.8 (1.3)	2.6 (0.9)	2.2 (0.8)	3.4 (0.9)
Col. Int. OR	0 (0)	0.8 (0.8)	1.6 (0.5)	1.8 (1.1)	2.2 (0)	2.6 (0.5)	3 (0.7)
Articulation D.	0 (0)	0 (0)	0.6 (0.5)	1.8 (0.4)	2 (0.4)	3 (0)	3.2 (0.4)
Hyperarticulation	0 (0)	0 (0)	0.4 (0.5)	1.8 (0.4)	1.8 (0.4)	3 (0)	3 (0)
				Male			
V. Effort	0 (0)	0.8 (0.4)	1 (0.5)	1.4 (0.8)	2.6 (0.5)	2.5 (1)	3.25 (0.9)
V. Alteration	0 (0)	0.4 (0.5)	1.75 (0.8)	1.8 (1.3)	2.2 (0.4)	2.5 (1)	2.75 (0.5)
V. Projection	0 (0)	1 (0)	1 (0.5)	1.4 (0.8)	2.6 (0.5)	2.5 (1)	3.25 (0.9)
Pat. Int.	0 (0)	1.2 (0,4)	2 (0)	2.2 (1.3)	3.2 (0.4)	2.75 (0.9)	3.75 (0.5)
Col. Int. Ward	0 (0)	0.4 (0,5)	1.75 (0.5)	1.2 (0.4)	2.2 (1.1)	1.75 (0.5)	3.25 (1.5)
Col. Int. OR	0 (0)	0.4 (0.5)	1.25 (0.9)	1.2 (0.4)	2.2 (1.1)	1.75 (0.5)	3.25 (1.5)
Articulation D.	0 (0)	0.6 (0.5)	1 (0)	2.2 (0.4)	2.2 (0.4)	3 (0)	3.25 (0.5)
Hyperarticulation	0 (0)	0.6 (0.5)	0.75 (0.5)	2 (0.7)	2.2 (0.4)	3 (0)	2.75 (0.5)

TABLE VI. Performance evaluation of the AI experiment where objective acoustic parameters and subjective ratings were used as features to distinguish 7 PPE configurations, for the female group.

Parameter	Surgical	Surgical +shield	FFP2	FFP2 +shield	FFP3	FFP3 +shield	Baseline
Precision	100%	100%	100%	100%	100%	100%	88%
Recall	87%	100%	100%	100%	100%	100%	100%
Specificity	100%	100%	100%	100%	100%	100%	98%
F1-score	93%	100%	100%	100%	100%	100%	94%
AUC	99%	100%	100%	100%	100%	100%	99%
Validation Accuracy				98%			

TABLE VII. Performance evaluation of the AI experiment where objective acoustic parameters and subjective ratings were used as features to distinguish 7 PPE configurations, for the male group.

Parameter	Surgical	Surgical +shield	FFP2	FFP2 +shield	FFP3	FFP3 +shield	Baseline
Precision	100%	92%	93%	86%	100%	100%	100%
Recall	100%	100%	100%	100%	83%	83%	100%
Specificity	100%	99%	99%	98%	100%	100%	100%
F1-score	100%	96%	97%	92%	91%	91%	100%
AUC	100%	100%	99%	99%	99%	99%	100%
Validation accuracy				95%			

TABLE VIII. Performances evaluation for the AI experiment where objective acoustic parameters and subjective ratings were used as features to distinguish between the absence/presence of the face shield. Left: females; right: males.

	Female		Male	
Parameter	No shield	With shield	No shield	With shield
Precision	93%	88%	90%	89%
Recall	87%	93%	90%	89%
Specificity	93%	87%	89%	90%
F1-score	90%	90%	90%	89%
AUC	96%	96%	96%	96%
Validation Accuracy	86%		90%	

experiment we tried to generalize the first one. Thus, male and female datasets were divided into two classes: masked and unmasked conditions. All the 120 features were considered. Amongst the three types of classifiers, RFs were again those with the best performance for each group.

Table IX displays the results of the third AI experiment separately for the two genders. Evaluation metrics explain how the baseline condition (no mask), i.e., no mask worn, is correctly detected with respect to the presence of any PPE worn by the subject (mask). Such generalization has allowed the development of a model with high validation accuracy (100% for male subjects), however, no method was applied to take into account unbalanced data.

The fourth experiment investigated whether masked configuration affected differently male and female subjects. Indeed, PSD evaluation as well as articulatory measures suggested different vocal efforts based on gender, as already stated in Ref. 14. Thus, excluding the baseline condition, female and male databases were considered altogether and the three chosen classifiers were sequentially trained. RF performed best also for this task. Table X shows the final results: again, precision, recall, specificity, and F1-score summarize the difference of voice features for female and male groups (female and male, respectively) while wearing protective masks. This last classifier obtained a 100% validation accuracy: even if vocal folds vibrate with different frequencies depending on gender, wearing masks might imply different adjustments of voice production between males and females.

TABLE IX. Performance evaluation for the AI experiment where objective acoustic parameters and subjective ratings were used as features to distinguish between baseline condition (no mask) and the presence of PPEs. Left: females; right: males.

	Female		Male	
Parameter	No mask	Mask	No mask	Mask
Precision	98%	100%	100%	100%
Recall	100%	98%	100%	100%
Specificity	98%	100%	100%	100%
F1-score	94%	99%	100%	100%
AUC	100%	99%	100%	100%
Validation Accuracy	98%		100%	

TABLE X. Performance evaluation of the AI experiment where objective acoustic parameters and subjective ratings were used as features to distinguish between female and male group wearing PPEs.

Parameter	Male	Female
Precision	100%	100%
Recall	100%	100%
Specificity	100%	100%
F1-score	100%	100%
AUC	100%	100%
Validation Accuracy	100%	

IV. DISCUSSION

This study focused on the identification of significant differences between masked and unmasked conditions and presence/absence of the face shield. Possible differences between single PPEs configurations (e.g., between surgical mask and FFP3 with face shield) were also explored, but they are not reported here as they are beyond the scope of this work.

Concerning sustained vowels, the following considerations can be made.

For female subjects, the mask configuration that gave most significant deviations from the baseline is the FFP2+shield that presents differences in

- F1a/F1i ($p = 0.003$)
- F2i/F2u ($p < 0.001$)
- VSA ($p < 0.001$)
- FCR ($p < 0.001$)

Articulation is thus strongly influenced by this PPEs combination even if the shield itself does not compromise articulatory movements. These alterations may be associated with more onerous frameworks that induce greater discomfort in female healthcare professionals, possibly linked to gender-related higher voice frequencies or because PPE are mostly designed with male anthropometric measures.

In general, FCR mean values for masked situation are higher than in the baseline condition. This might be due to restriction of movement for lips and jaw (and consequently tongue) as in Gustin *et al.*[41] that found a decrease in VAI (inverse of FCR). These restrictions, though minimal, might cause acoustically different sound production with respect to normal conditions, according to the quantal theory of speech.[35] FFP2 and FFP3 masks show significant statistical difference with the baseline when worn together with a face shield (Table III): their combination negatively affects articulation parameters such as F1a/F1i, F1a/F1u, and FCR. Interestingly, articulatory capabilities seemed to be impaired without face shield as well, as statistical differences were found with respect to the baseline in VSA for FFP2 masks and in F1a/F1i for FFP3 masks. This result is supported as well by subjective ratings related to articulation difficulties (Q7 of the questionnaire) and the need to hyperarticulate to be understood (Q8) when wearing FFP2 masks with face shields and FFP3 respirators. On the other hand, surgical masks alone show significant differences only when acoustic

https://doi.org/10.1121/10.0017244

parameters are compared with the surgical mask + shield configuration, e.g., in F1a/F1u ($p = 0.004$) and in VSA ($p < 0.001$), in accordance with Ref. 13 and Q1-Q3 subjective ratings. However, surgical masks in combination with face shields alter articulation as well, especially when F1a/F1i and F1a/F1u are considered.

Concerning formants, Fig. 2(a) shows that the presence of a face shield critically reduces the vowel triangle dimensions both when the baseline and the mask alone configuration are considered. In particular, as compared to the baseline, this additional PPE causes a decrease in F1 /a/ especially in combination with a surgical mask and of F2 /i/ when worn with a FFP2 mask. Alterations occurring when using a face shield are illustrated as well in the boxplots of Fig. 2(b).

Fundamental frequency F0 of cardinal vowels is not affected by mask, though some irregularities in phonation were found: FFP3 and face shield determine a statistical difference in jitter /u/ with respect to the baseline. Furthermore, surgical mask with shield and FFP3 cause NNE /i/ to significantly decrease with respect to the baseline. These results are in line with Gojayev *et al.*[15] and Lin *et al.*,[14] who stated that covering the mouth makes dysphonia less evident and raises HNR, also as a consequence of high frequencies attenuation and indicates that wearing a mask improve frontal resonance.[42] These results could be helpful for ENT specialists who can ask patients to phonate wearing surgical masks to prevent projection and aerosolization, as vocal properties seem unaffected by this type of PPE. Also, FFP2 and FFP3 masks, which offer higher protection, do not alter F0 and its correlated measures, although, as reported in some studies,[12,14,16] there is a tendency of F0 to rise, which might be associated with a compensation of voice intensity decrease, caused in turn by difficulties in coordination in the respiratory-laryngeal system during speech that reduce airflow intake.[42] This occlusion effect might bring people to speak louder than normal, especially when PPEs become more onerous: the subjective rates in Table V highlight this condition especially for self-perceived voice alteration (Q2) and difficulty in voice projection (Q3).

For male subjects, FFP2 and shield present the highest number of statistical differences with the baseline both in articulation and phonation; alterations have been found in F1a/F1i, F2i/F2u, VSA, FCR, and NNE /a/. Figure 3(a) shows that F2 mean /u/ for surgical mask (broken line) deviates from the baseline, in line with the results found by Georgiou[35] for the same type of PPE. This might be caused by articulators' readjustments in order to improve intelligibility. Furthermore, in the work by Joshi *et al.*,[16] F2 mean /i/ alterations were highlighted when a FFP2 mask is worn together with a face shield only in male subjects. In our work this is shown in Figs. 2(a) and 3(a) for FFP2+shield (thin dash-dotted line) and FFP3 (thick dash-dotted line). Anomalies from the baseline can be observed as well in Fig. 3(a), particularly for FFP2+shield configuration. The relevance of F2 alteration suggests a possible mask's impact on front-back movement of the tongue.[35] In line with Cavallaro *et al.*[13] surgical masks loosely affect both articulation and phonation, as shown by the answers to question 1–3 in Table V.

Subjective ratings show similar trends in both groups: for surgical masks, vocal effort alterations (Q1) are minimal as well as the need to hyperarticulate (Q8) in order to be understood by patients or colleagues, while scores to questions 1 and 8 significantly rise with more tightening PPEs, underlying that more onerous mask configurations impair communication, which is further compromised by the presence of a face shield. Indeed, Table V shows a marked increase in vocal effort, alteration and projection values between mask type and the same mask type worn together with a face shield. It is interesting to notice that, on the other hand, articulation in subjective ratings (Q7 and Q8) does not change when the same comparison is considered.

As far as the PSD is concerned, the presence of PPEs results in an effective variation of the PSD, usually lowering the average power. For a more detailed analysis, in our work a new investigation of the PSD is performed that does not evaluate the total average power, but the variations of the average power as a function of frequency. This analysis was implemented by dividing the frequency spectrum into 500 Hz intervals and calculating the average power over each interval for the different configurations of masks considered. In Ref. 30 such strategy was applied on each corner vowel; in this work, to focus on possible articulation deficits using protective masks, it was applied only on the vocalic sequence /a'jw/ extracted from the word /a'jwɔle/ of the vocalic sentence. As expected, the configuration with the surgical mask has the least influence for both groups of subjects (males and females). Below 1 kHz attenuation effects are limited for all masks configurations whereas beyond that threshold PPEs act differently on the PSD, in agreement with Shekaraiah and Suresh,[42] Magee *et al.*,[11] and Toscano and Toscano.[43] Specifically, the three types of face masks worn together with a face shield determine strongest attenuations for frequencies > 1 kHz, while it becomes relevant for masks alone above 2.5 kHz, as shown in Figs. 4 and 5. Between 2.5 and 3.5 kHz the mean PSD decreases when using a surgical mask and a FFP2 that, respectively, amounts to 2.1 and 3.8 dB. This latter value is lower with respect to the one reported by Nguyen *et al.*,[44] possibly because of the strategy implemented for PSD analysis in this work. The filtering capability and fitting degree on the wearer's face, which is higher for FFP2 and FFP3 masks, causes a more relevant reduction and consequently stronger signal degradation. Indeed, FFP2 masks decrease up to the 90% the amount of particles projection during speaking,[45] but also causes a more difficult perception in listeners. Corey *et al.*[9] discovered how attenuation properties are linked to mask material, layer number, thickness, and weave pattern. Interestingly, our results show that the presence of the shield causes PSD upward distortions in the 3–3.5 kHz frequency range only for the male group when worn with surgical and FFP2 masks, as supported by statistical differences in the same interval (Table IV).

Overall, configurations involving the simultaneous presence of a face mask and a shield are more onerous, with sharper drops in average power and marked differences in acoustic vocal measures. The most invasive configuration, in terms of alteration of the power of the voice signal and articulation impairment, is the presence of an FFP2 mask and a

shield, both for the male and female groups. This may be caused by the different type of materials (usually rigid plastic) of which shields are made.[10] Even if these PPEs can restore visual cues for people with hearing impairments, healthcare professionals must be aware of the relevant drop in PSD, especially in frequency ranges critical for vowel decoding.[10] These results are in agreement with Corey *et al.*[9]: at high frequencies windowed masks have shown PSD drops of about 8 dB for real speaker and up to 14 dB for loudspeaker inserted in dummy heads. Furthermore, Gama *et al.*[46] highlighted that, regardless of the mask type, significant alterations in acoustic properties exist when wearing a PPE, especially as far as HNR, F2, and PSD are concerned.

Although with different performances, the machine learning classification outcomes presented in this work are in line with the results in Ref. 14. Models were able to discriminate among four different conditions:

(1) The seven different types of PPEs.
(2) Absence or presence of a face shield.
(3) Masked or unmasked conditions.
(4) Male and female voice features with PPEs.

For males, in the first experiment an RF model was able to correctly discriminate among surgical mask, surgical mask with shield, and no PPE. Other configurations gave less relevant results, possibly because FFP2 and FFP3 masks have similar shapes and materials, thus the fitting level of the model does not change much. However, high specificity suggests that some peculiar voice parameters may have avoided large misclassification. In female subjects we got better results, however, recall and precision values show that surgical mask and baseline, as well as FFP3 and FFP3+shield observations, were misclassified.

In the second experiment, AUC evaluated from Table VIII showed better performances in distinguishing between masks without and with a face shield. This result is in line with statistical analysis that has demonstrated how articulatory parameters and PSD differ if the PPEs are worn without shield. Comparing the three masks configurations with the corresponding masks + shields ones, RF classifiers were able to distinguish these two cases with high accuracy for both female and male groups. Similar results were obtained by considering 29 features (Tables I and II parameters), with validation accuracy equal to 83% and 93% for male and female group, respectively, with RF classifiers: including measures such as cardinal vowels F0, jitter and NNE, along with subjective ratings, may have improved the results, especially for the female group where the RF model achieved an accuracy of 90%.

Regardless of the mask type, wearing a mask alters both perceptually and objectively the acoustic properties of voice. This was partially proved by good classification performances of the baseline models in the first experiment that were successfully identified with high performances despite the small number of observations. In the third experiment this capability remains valid in a more general context, in which female and male groups were split into masked and unmasked classes. Table IX shows that masked configurations are better distinguished: more onerous PPE, such as FFP2 and FFP3 alone or with face shield, determine more relevant, and thus more detectable, anomalies in the acoustic parameters. However, we point out that classes were unbalanced and, as a first attempt, no strategy (such as oversampling or synthetic sampling) was applied: therefore, these outcomes need to be further investigated in the future.

The fourth AI experiment showed that masks determine differences between genders: this may be also due to the fact that face coverings have universal shapes which may differently adapt on wearers' faces.

To the authors' knowledge, this is the first attempt to apply AI techniques for the acoustic characterization of voices while wearing PPEs that takes into account both objective and subjective voice parameters. The proposed approach and the achieved results could help ENT specialists and occupational voice users to adjust speech production to avoid stressing their phonatory apparatus while being anyway intelligible. The results also suggest that at least for professionals some vocal exercise such as bubbling[47] and face gym for articulation[48] would be advisable.

V. CONCLUSION

In this paper results about acoustic parameters estimated without and with protective masks are reported. Both sustained vowels and a vocalic sentence are recorded and analyzed acoustically and perceptually with both statistical and machine learning techniques. To the authors' knowledge, this is the first time that such task was performed considering seven masks configurations.

Interesting differences with respect to the baseline (no mask) are found, especially as far as VSA and spectral energy are concerned. The first one shows a reduced articulatory workspace when face mask and shield are worn together and the latter is characterized by consistent spectral power decrease above 1 kHz for female and 2 kHz for male subjects, in line with existing literature. Acoustic parameters are able to distinguish mask configurations in four different experiments: although promising, these results were obtained with limited observations and further studies, with a larger dataset, are required. Moreover, feature selection and data balancing correction need to be applied in the future.

Additional research is ongoing to support these results and find out most relevant predictors and in future work a hierarchical classifier could be developed to effectively differentiate the seven PPE configurations.

The proposed methods and the results achieved in this work might help to improve verbal communication without causing excessive phonatory strain, in particular for the healthcare professional in the operating room or in interviews with patients with hearing impairment.[5]

ACKNOWLEDGMENTS

This work was partially funded under the project 2018.0976 Fondazione Cassa di Risparmio di Firenze, Firenze, Italy.

https://doi.org/10.1121/10.0017244

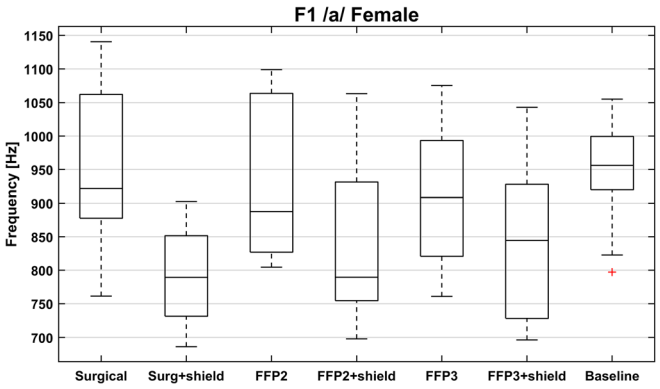

FIG. 6. (Color online) F1 mean /a/ boxplots for female subjects.

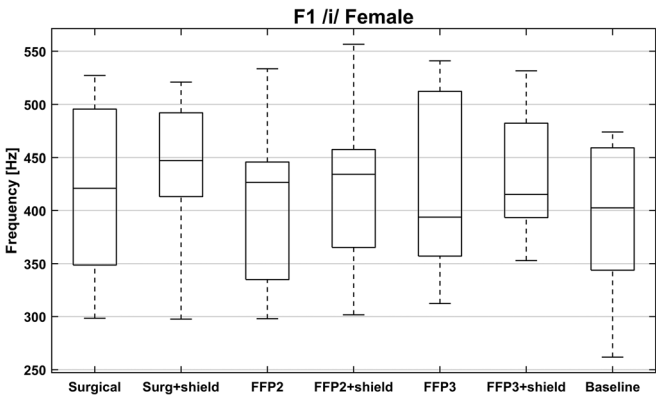

FIG. 7. F1 mean /i/ boxplots for female subjects.

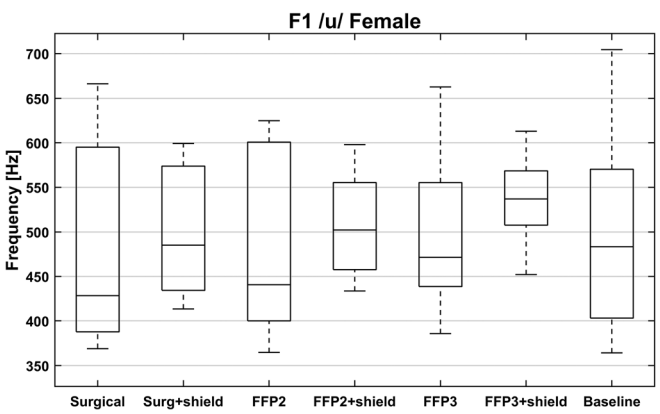

FIG. 8. F1 mean /u/ boxplots for female subjects.

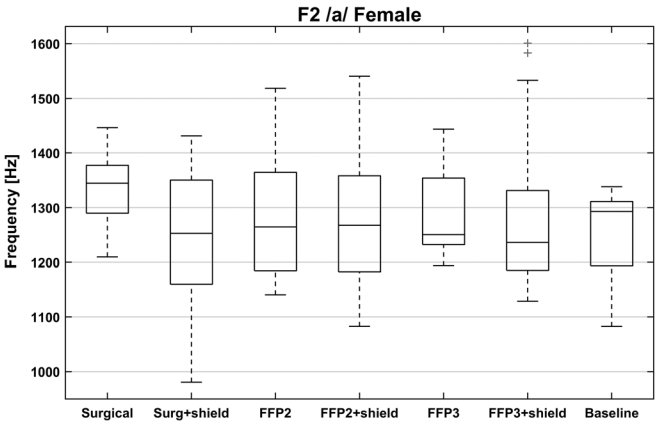

FIG. 9. F2 mean /a/ boxplots for female subjects.

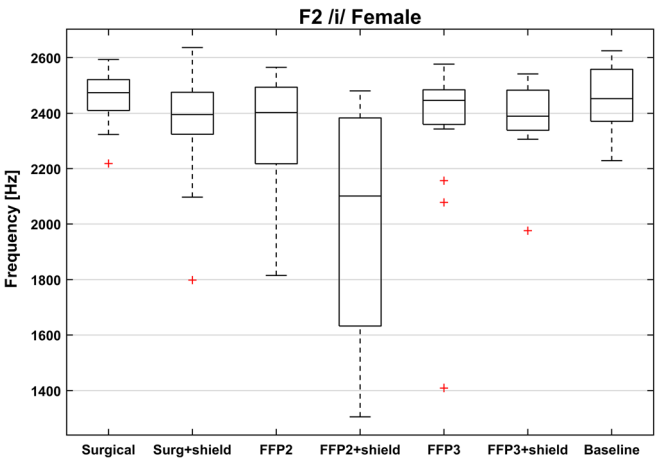

FIG. 10. (Color online) F2 mean /i/ boxplots for female subjects.

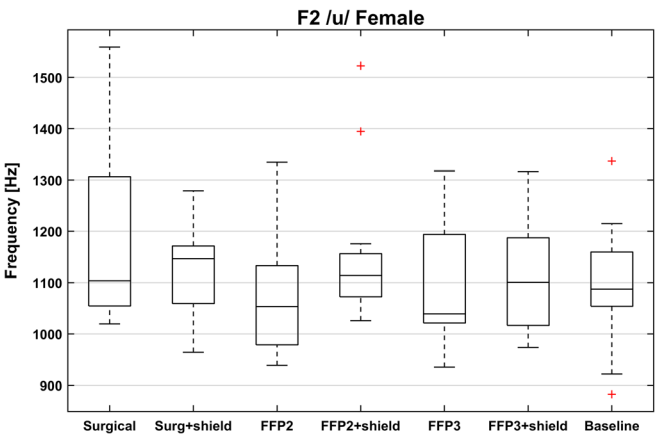

FIG. 11. (Color online) F2 mean /u/ boxplots for female subjects.

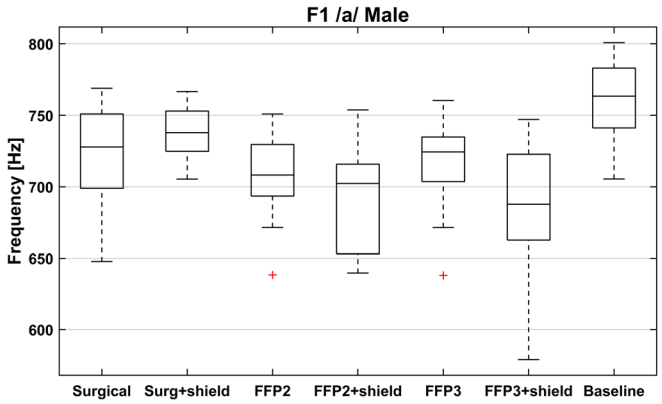

FIG. 12. (Color online) F1 mean /a/ boxplots for male subjects.

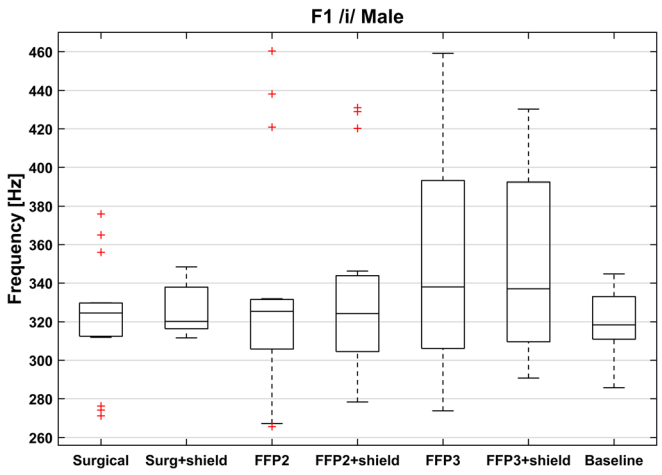

FIG. 13. (Color online) F1 mean /i/ boxplots for male subjects.

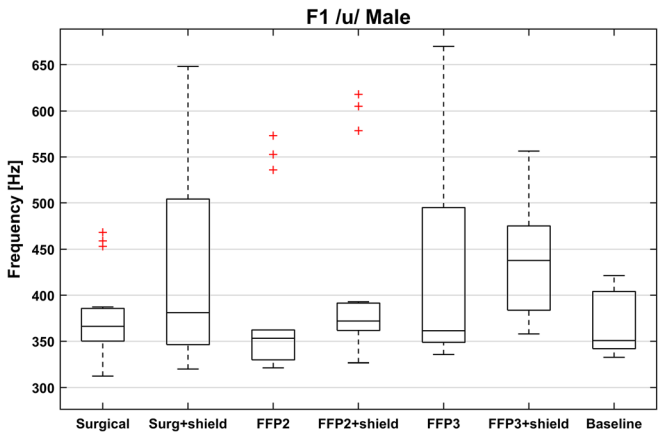

FIG. 14. (Color online) F1 /u/ mean for male subjects.

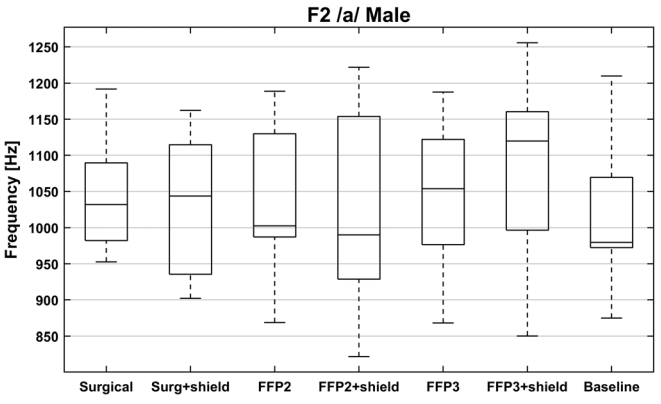

FIG. 15. F2 /a/ mean for male subjects.

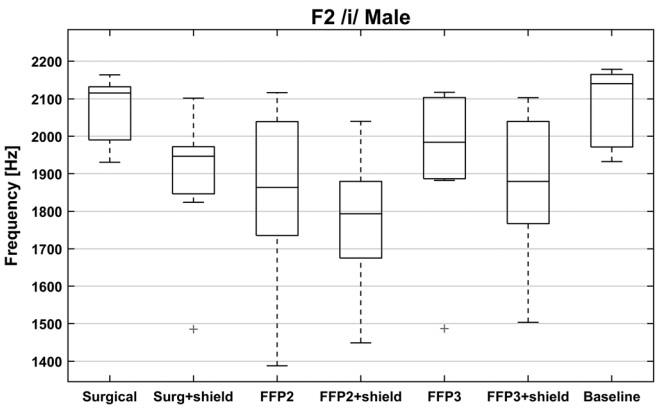

FIG. 16. F2 mean /i/ for male subjects.

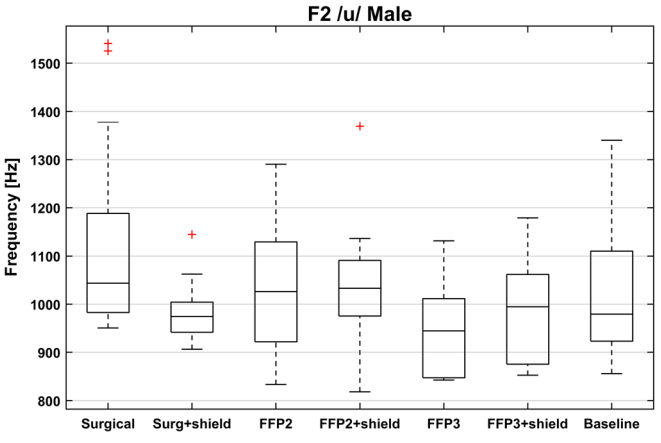

FIG. 17. (Color online) F2 mean /u/ for male subjects.

https://doi.org/10.1121/10.0017244

APPENDIX A: FORMANTS F1 AND F2 BOXPLOTS FOR FEMALE SUBJECTS

Figures 6, 7, and 8 show the F1 mean (Table II) boxplots of cardinal vowels /a/, /i/, and /u/ to support visual interpretation of the vocalic triangle presented in Fig. 2(a). Especially for F1 mean /a/, it is possible to observe how the presence of a face shield worn together with a mask determines a decrease in frequency of F1 in respect both to the baseline and the PPE worn without the shield, while for F1 mean /i/ and F1 mean /u/ the opposite trend is noticeable.

Figures 9, 10, and 11 show the F2 mean (Table II) boxplots of cardinal vowels /a/, /i/, and /u/ to support visual interpretation of the *y* axis values from vocalic triangle presented in Fig. 2(a). It is possible to notice relevant F2 alteration for /i/ in Fig. 10 when subjects wears surgical mask with shield, FFP2, and FFP2 mask with shield.

APPENDIX B: FORMANTS F1 AND F2 BOXPLOTS FOR MALE SUBJECTS

Figures 12, 13, and 14 show the F1 mean (Table II) boxplots of cardinal vowels /a/, /i/, and /u/ to support visual interpretation of the *x* axis values from the vocalic triangle presented in Fig. 3(a), for male subjects. In Fig. 12, for F1 mean /a/, it is possible to observe how the presence of a face shield worn together with FFP2 and FFP3 masks determines a decrease in frequency, analogously for female subjects. In Fig. 14, a relevant difference in F1 mean /u/ can be noticed between the baseline (no PPE worn) and the FFP3+shield configuration.

Figures 15, 16, and 17 show the F2 mean (Table II) boxplots of cardinal vowels /a/, /i/, and /u/ to support visual interpretation of the *y* axis values from vocalic triangle presented in Fig. 2(b), for male subjects. It is possible to notice relevant F2 mean /a/ in Fig. 15 when subjects wear the most onerous PPE configurations (i.e., FFP3 and FFP3 with shield), and in F2 mean /i/ for all mask condition.

[1]I. R. Titze, J. Lemke, and D. Montequin, "Population in the U.S. workforce who rely on voice as a primary tool for trade: A preliminary report," J. Voice **11**, 254–259 (1997).
[2]I. Oosthuizen, G. H. Saunders, V. Manchaiah, and D. W. Swanepoel, "Impact of SARS-CoV-2 virus (COVID-19) preventative measures on communication: A scoping review," Front. Public Health **10**, 652–662 (2022).
[3]V. V. Ribeiro, A. P. Dassie-Leite, E. C. Pereira, A. D. N. Santos, P. Martins, and R. de Alencar Irineu, "Effect on wearing a face mask on vocal self-perception during a pandemic," J. Voice **36**, 878.e1–878.e2 (2020).
[4]T. G. Vos, M. T. Dillon, E. Buss, M. A. Rooth, A. L. Bucker, S. Dillon, and M. M. Dedmon, "Influence of protective face coverings on the speech recognition of cochlear implant patients," Laryngoscope **131**(6), E2038–E2043 (2021).
[5]J. Chodosh, B. E. Weinstein, and J. Blustein, "Face masks can be devastating for people with hearing loss," BMJ **370**, m2683 (2020).
[6]C. Porschmann, T. Lubeck, and J. M. Arend, "Impact of face masks on voice radiation," J. Acoust. Soc. Am. **148**(6), 3663–3670 (2020).
[7]A. Goldin, B. E. Weinstein, and N. Shiman, "How do medical masks degrade speech perception?," Hear. Rev. **27**(5), 8–9 (2020).
[8]D. D. Nguyen, A. Chacon, C. Payten, R. Black, M. Sheth, P. McCabe, and C. Madill, "Acoustic characteristics of fricatives, amplitude of formants and clarity of speech produced without and with a medical mask," Int. J. Lang. Commun. Disorders **57**(2), 366–380 (2022).
[9]R. M. Corey, U. Jones, and A. C. Singer, "Acoustic effects of medical, cloth, and transparent face masks on speech signals," J. Acoust. Soc. Am. **148**(4), 2371–2375 (2020).
[10]B. T. Balamurali, T. Enyi, J. C. Clarke, S. Harn, and M. J. Chen, "Acoustic effects of face mask design and material choice," Acoust. Aust. **49**(3), 505–512 (2021).
[11]M. Magee, C. Lewis, G. Noffs, H. Reece, J. C. S. Chan, C. J. Zaga, C. Paynter, O. Birchall, S. Rojas-Azocar, and A. Ediriweera, "Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols," J. Acoust. Soc. Am. **148**(6), 3562–3568 (2020).
[12]G. Cavallaro, V. D. Nicola, N. Quaranta, and M. L. Fiorella, "Acoustic voice analysis in the COVID-19 era," Acta Otorhinolaryngol. IItal. **41**(1), 1–12 (2021).
[13]M. L. Fiorella, G. Cavallaro, V. Di Nicola, and N. Quaranta, "Voice differences when wearing and not wearing a surgical mask," J. Voice (published online 2021).
[14]Y. Lin, L. Cheng, Q. Wang, and W. Xu, "Effects of medical masks on voice assessment during the COVID-19 pandemic," J. Voice (published online 2021).
[15]E. K. Gojayev, Z. C. Buyukatalay, T. Akyuz, M. Reham, and G. Dursun, "The effects of masks and respirators on acoustic voice analysis during the COVID-19 pandemic," J. Voice (published online 2021).
[16]A. Joshi, T. Procter, and P. A. Kulesz, "COVID-19: Acoustic measures of voice in individuals wearing different facemasks," J. Voice (published online 2021).
[17]S. V. Bandaru, A. M. Augustine, A. Lepcha, S. Sebastian, M. Gowri, A. Philip, and M. D. Mammen, "The effects of N95 mask and face shield on speech perception among healthcare workers in the coronavirus disease 2019 pandemic scenario," J. Laryngol. Otol. **134**(10), 895–898 (2020).
[18]V. S. McKenna, T. H. Patel, C. L. Kendall, R. J. Howell, and R. L. Gustin, "Voice acoustics and vocal effort in mask-wearing healthcare professionals: A comparison pre-and post-workday," J. Voice (published online 2021).
[19]M. Hirano, "Clinical examination of voice," in *Disorders of Human Communication* (Springer-Verlag, Wien, 1981), pp. 1–99.
[20]B. H. Jacobson, A. Johnson, C. Grywalski, A. Silbergleit, G. Jacobson, M. S. Benninger, and C. W. Newman, "The voice handicap index (VHI) development and validation," Am. J. Speech-Lang. Pathol. **6**(3), 66–70 (1997).
[21]Kay Elemetrics Corporation, *Operations Manual: Multi-Dimensional Voice Program* (MDVP, Lincoln Park, NJ, 1993).
[22]P. Boersma and V. Van Heuven, "Praat, a system for doing phonetics by computer," Glot. Int. **5**(9/10): 341–345.
[23]M. S. Morelli, S. Orlandi, and C. Manfredi, "BioVoice: A multipurpose tool for voice analysis," Biomed. Sign. Process. Control **64**, 102302 (2021).
[24]P. Fabre, "Etude comparée des glottogrammes et des phonogrammes de la voix humaine" ("Study comparing the glottogram and the phonogram of the human voice"), Ann. Oto-rhino Laryngol. **75**, 767–775 (1958).
[25]L. T. D. Siqueira, J. D. S. Vitor, A. P. Dos Santos, R. L. F. Silva, P. A. M. Moreira, and V. Veis Ribeiro, "Influence of the characteristics of home office work on self-perceived vocal fatigue during the COVID-19 pandemic," Logoped. Phon. Vocol. **47**, 279–283 (2021).
[26]G. Lenoci, C. Celata, I. Ricci, A. Chilosi, and V. Barone, "Vowel variability and contrast in childhood apraxia of speech: Acoustics and articulation," Clin. Ling. Phon. **35**(11), 1011–1035 (2021).
[27]C. Manfredi, J. Lebacq, G. Cantarella, J. Schoentgen, S. Orlandi, A. Bandini, and P. H. DeJonckere, "Smartphones offer new opportunities in clinical voice research," J. Voice **31**, 111.e1–111.e7 (2017).
[28]J. Lebacq, J. Schoentgen, G. Cantarella, F. T. Bruss, C. Manfredi, and P. H. DeJonckere, "Maximal ambient noise levels and type of voice material required for valid use of smartphones in clinical voice research," J. Voice **31**(5), 550–556 (2017).
[29]C. C. Burgel, R. Bartholomaus, W. Fiesel, J. Hilpert, A. Hoelzer, and K. Linzmeier, "Beyond CD-quality: Advanced audio coding (AAC) for high resolution audio with 24 bit resolution and 96 kHz sampling frequency," in *Audio Engineering Society Convention 111* (Audio Engineering Society, New York, 2001).
[30]C. Manfredi, V. Altamore, A. Bandini, S. Orlandi, L. Battilocchi, and G. Cantarella, "Effect of protective masks on voice parameters: Acoustical

analysis of sustained vowels," in *Proceedings of Models and Analysis of Vocal Emissions for Biomededical Applications* (December 2021), Vol. 171–174, pp. 14–16.
[31]H. Kasuya, S. Ogawa, K. Mashima, and E. S, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," J. Acoust. Soc. Am. **80**(5), 1329–1334 (1986).
[32]S. Shapir, J. L. Spielman, L. O. Ramig, B. H. Story, C. Fox, and C. Blog, "Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings," J. Speech Lang. Hear. Res. **50**, 899–912 (2007).
[33]Y. J. K. R. D. Kent, "Towards an acoustic typology of motor speech disorders," Clin. Ling. Phon. **17**(6), 427–445 (2003).
[34]S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech," J. Speech Lang. Hear. Res. **53**(1), 114–125 (2010).
[35]G. P. Georgiou, "Acoustic markers of vowels produced with different types of face masks," Appl. Acoust. **191**, 108691 (2022).
[36]M. Vasilakis and Y. Stylianou, "Voice pathology detection based on short-term jitter estimations in running speech," Folia Phoniatr. Logopaed. **61**(3), 153–170 (2009).
[37]C. Robotti, G. Costantini, G. Saggio, V. Cesarini, A. Calastri, E. Maiorano, D. Piloni, T. Perrone, U. Sabatini, V. V. Ferretti, I. Cassaniti, F. Baldanti, A. Gravina, A. Sakib, E. Alessi, M. Pascucci, D. Casali, Z. Zarezadeh, V. Del Zoppo, A. Pisani, and M. Benazzo, "Machine learning-based voice assessment for the detection of positive and recovered COVID-19 patients," J. Voice (published online 2021).
[38]J. A. Gomez-Garcia, L. Moro-Velazquez, and J. I. Godino-LLorente, "On the design of automatic voice condition analysis systems. Part I: Review of concepts and an insights to the state of the art," Biomed. Sign. Process. Control **51**, 181–199 (2019).
[39]P. Harar, Z. Galaz, J. B. Alonso-Hernandez, J. Mekyska, R. Burget, and Z. Smekal, "Towards robust voice pathology detection," Neural Comput. Appl. **32**(20), 15747–15757 (2020).
[40]*MATLAB and Statistics Toolbox Release 2020b* (The MathWorks, Inc., Natick, MA, 2020).
[41]V. S. McKenna, C. L. Kendall, T. H. Patel, R. J. Howell, and R. L. Gustin, "Impact of face masks on speech acoustics and vocal effort in healthcare professionals," Laryngoscope **132**(2), 391–397 (2022).
[42]S. Shekaraiah and K. Suresh, "Effect of face mask on voice production during COVID-19 Pandemic: A systematic review," J. Voice (published online 2021).
[43]J. C. Toscano and C. M. Toscano, "Effects of face masks on speech recognition in multi-talker babble noise," PloS one **16**(2), e0246842 (2021).
[44]D. D. Nguyen, P. McCabe, D. Thomas, A. Purcell, M. Doble, D. Novakovic, and C. Madill, "Acoustic voice characteristics with and without wearing a facemask," Sci. Rep. **11**(1), 1–11 (2021).
[45]S. Asadi, C. D. Cappa, S. Barreda, A. S. Wexler, N. M. Bouvier, and W. D. Ristenpart, "Efficacy of masks and face coverings in controlling outword aerosol particle emission from expiratory activities," Sci. Rep. **10**(1), 1–13 (2020).
[46]R. Gama, M. E. Castro, J. T. van Lith-Bijl, and G. Desuter, "Does the wearing of masks change voice and speech parameters?," Eur. Arch. Oto-Rhino-Laryngology **279**, 1701–1708 (2021).
[47]A. Bandini, S. Orlandi, F. Giovannelli, A. Felici, M. Cincotta, D. Clemente, and C. Manfredi, "Markerless analysis of articulatory movements in patients with Parkinson's disease," J. Voice **30**(6), 766.e1–766.e11 (2016).
[48]V. Di Natale, G. Cantarella, C. Manfredi, A. Ciabatta, C. Bacherini, and P. H. DeJonckere, "Semioccluded vocal tract exercises improve self-perceived voice quality in healthy actors," J. Voice **36**(4), 584.E7–584.E14 (2022).