EFFECT OF PROTECTIVE MASKS ON VOICE PARAMETERS: ACOUSTICAL ANALYSIS OF SUSTAINED VOWELS

C.Manfredi[1], V.Altamore[2], A.Bandini[3], S.Orlandi[4], L.Battilocchi[5], G.Cantarella[5]

[1]Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[2]School of Engineering, Università degli Studi di Firenze, Firenze, Italy
[3] Scuola Superiore Sant'Anna, The Biorobotics Institute, Pisa, Italy
[4]Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione "Guglielmo Marconi", Bologna, Italy
[5]IRCCS Ca' Granda Foundation, Ospedale Maggiore Policlinico Milano, Milano, Italy
claudia.manfredi@unifi.it, virginia.altamore@stud.unifi.it, andreabandini87@gmail.com, orlandisilvia85@gmail.com, ludovicabattilocchi@gmail.com, giovanna.cantarella@policlinico.mi.it

Abstract: **During the last two years the use of masks as personal protective equipment became necessary and mandatory to deal with the SARS-CoV-2 epidemiological emergency with impact on the quality and efficiency of verbal communication.**
This paper compares for the first time 7 different mask configurations. The sustained vowels /a/, /i/ and /u/ emitted by Italian speakers are considered. The purpose of this work is to evaluate whether the use of different types of masks, by themselves or worn together with a protective shield, may affect the acoustical parameters and thus voice quality. This is exploited by means of acoustical analysis performed with the BioVoice tool that estimates more than 20 parameters. For each vowel, the values of the fundamental frequency F0, the first two formant frequencies F1 and F2, jitter and noise are compared among the 7 configurations.
Preliminary results show that for the three vowels there are few statistically significant differences among masks when worn alone, while the presence of the shield has a relevant impact on the signal energy above 1 kHz. Further studies are ongoing to analyze vocalic sentences in order to detect possible influence of the masks on vowel articulation.
Keywords: **Face masks, SARS-CoV-2, acoustical analysis, BioVoice, F0, formants.**

I. INTRODUCTION

Most personal protective equipments (PPEs) have a relevant impact both on the quality and the intelligibility of the voice signal especially in the case of noisy environments or hearing impairments. Moreover the inter-personal mandatory distance often leads to raise the voice, increasing voice fatigue especially for professionals that have a high daily voice load. Therefore, in the last two years the scientific community has examined the influence of masks on vocal acoustic characteristics. [1-6]. This paper compares for the first time 7 different mask configurations. The study is preliminary and limited **to** sustained vowels /a/, /i/ and /u/ emitted by Italian speakers, but further work is ongoing to analyze vocalic sentences in order to detect possible influence of the masks on vowel articulation.

The main characteristics of vowels are the fundamental frequency F0 and the formant frequencies, that is, the resonant frequencies of the vocal tract. In particular, the first two formants, F1 and F2, are related to the position of the tongue: F1 is linked to height, while F2 is linked to the front-to-back movement. Formants position may vary according to age and gender, but is also related to the language under consideration. In particular, the Italian language comprises just 7 vocalic sounds and does not make a distinction between rounded and non-rounded vowels. Figure 1 shows the vowel trapezoid of the Italian language according to the International Phonetic Alphabet (IPA):

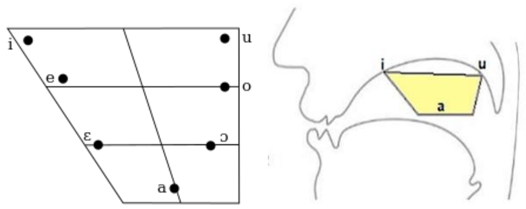

Figure 1 – Vowel trapezoid for the Italian language

In this work only the three vowels at the corners of the trapezoid are analyzed: /a/, /i/ and /u/. They roughly correspond to American English vowels /a/ ("father"), /I/ ("it") and /U/ ("foot") as reported in [7].

II. METHODS

Voice signals were recorded from 10 subjects (5 males and 5 females, mean age: 27,3, std= 1,494) of Italian mother tongue. Specifically, for females: mean=27,8; std= 0,836. For males: mean=26,8; std= 1,923.
Each subject was asked to emit the sustained vowels /a/, /i/ and /u/ at conversational amplitude for at least

5s. Seven configurations of different masks without or with protective shield were considered:
- No mask (Baseline)
- Surgical mask (Surgical)
- Ffp2 mask (Ffp2)
- Ffp3 mask (Ffp3)
- Surgical mask and visor (Surgical+shield)
- Ffp2 mask and visor (Ffp2+shield)
- Ffp3 mask and visor (Ffp3+shield)

Each vowel was repeated 3 times. Thus 210 recordings were collected.

Each recording was processed using the BioVoice tool that automatically distinguish among newborn cry, children, adults and singers voices, and in case of adults performs separate analyses for male and female voices [8, 9].

In this work, for each vowel the following acoustical parameters were considered: F0, F1, F2, jitter, and Adaptive Normalized Noise Energy (ANNE), along with their descriptive statistics (mean, median and standard deviation). A high-resolution method for F0 estimation is implemented in BioVoice, based on parametric AutoRegressive (AR) models applied to time-windows of varying length. AR models were also implemented to estimate the Power Spectral Density (PSD). PSD is automatically computed in the frequency range specific of each category, and normalized with respect to its maximum value, therefore the PSD range is 0dB downward. The first two PSD peaks correspond to F1 and F2.

ANNE relies on a comb filtering approach, optimised to deal with data windows of varying time duration. Large negative ANNE values correspond to good voice quality, while values close to zero reflect the presence of strong noise.

Concerning PSD, statistical analysis was implemented by dividing the frequency spectrum into 500 Hz intervals and calculating the average power over each interval for each mask configuration and each vowel, distinguishing between males and females. Also overall results (males and females) were considered and they are reported in this paper.

As data were not normally distributed, a non-parametric Friedman test was performed to find possible differences between the acoustical parameters obtained with the 7 configurations of face masks. In case of significant level of the Friedman test (p value < 0.05), a post-hoc multiple comparison was applied using the Dunn-Sidak method.

III. RESULTS

Results reported here concern male and female voices altogether. Separate analysis will be reported elsewhere.

A. F0, formant, jitter and noise

For all vowels and all configurations, F0 mean and median values are similar, both ranging between 168 Hz and 182 Hz, with a slight increase from the baseline to masks with shield. This could be related to an increasing effort in vocal emission due to protective equipments. However, no statistically significant differences were obtained for F0.

For the mean values of F1 and F2, the following ranges were found. Median values are not reported as they gave similar results.

Vowel /a/

F1: 720 Hz - 810 Hz. F2: 1080 Hz - 1180 Hz.

Vowel /i/

F1: 340 Hz - 370 Hz. F2: 1830 Hz - 2200 Hz.

Vowel /u/

F1: 400 Hz - 440 Hz. F2: 1010 Hz - 1120 Hz.

For all vowels jitter ranges between 0.50% and 1.4% and ANNE ranges between -23 dB and -27 dB.

The 6 configuration with masks were compared to the baseline with the Friedman test. Concerning jitter, no statistically significant difference was found for the three vowels.

Only the statistically significant results are reported here:

Vowel /a/

- F1 mean of FFP3 + shield was significantly lower than the baseline result.
- NNE of FFP2 + shield was significantly higher than the baseline result.
- NNE of FFP3 + shield was significantly higher than the baseline result.
- ANNE of FFP2 + shield and FFP3 + shield was significantly higher than the baseline result.

Vowel /i/

- F2 mean obtained with FFP2 + shield was significantly lower than the baseline.
- F2 median obtained with FFP2 + shield was significantly lower than the baseline.

Vowel /u/

No statistically significant differences were found with respect to the baseline.

B. Power Spectral Density

Figures 1-3 show the PSD trend in steps of 500 Hz for the three sustained vowels /a/, /i/ and /u/ and all the 7 configurations. The mean PSD of male and female values are presented, without differentiating between the two genders.

Baseline (no mask) is indicated with a solid line. Each dot corresponds to the mean value of the PSD values in each frequency step.

IV. DISCUSSION

Results show that F0 is only slightly influenced by masks and shield, while formants values exhibit statistically significant differences. This is especially true for F2 and for the vowel /a/. Indeed, F2 tends to be lowered by a back tongue constriction and raised by a front tongue constriction [7], therefore changes might be due to the presence of face mask and shield.

Also, higher ANNE values with shield with respect to the baseline might indicate a higher effort required in vowel emission. No statistically significant difference was found for jitter.

Concerning PSD: For vowel /a/ (Figure 1) the decrease in PSD is quite evident for the three configurations with shield already around 1kHz, where about 10dB of decrease is shown. Even larger differences are found from 2 kHz on for the same configurations. Less evident decrease is shown in Figures 2 and 3 that concern vowel /i/ and /u/ respectively.

It should be taken into account that these plots concern cumulative average values of men and women calculated over 500 Hz intervals, so they are somewhat different from the traditional power spectrum. In fact, when gender data were considered separately, a greater decrease and higher frequency values were observed for women. This might indicate a greater effort required to females with respect to males, possibly related to their shorter vocal tract and higher formant frequencies.

Furthermore, the vowel / u / is one of the most difficult to analyze through automatic tools, due to the position of its formants that depends on the position of the tongue and the conformation of the vocal tract which are very particular in this case.

Though preliminary, results show that masks alone have a negligible influence on the power spectral density (PSD) of sustained vowels /a/, /i/ and /u/, while the presence of the shield causes a relevant energy decreases above 1 kHz that is directly related to voice energy. This is particularly evident for vowel /a/ while vowels /i/ and /u/ show a less strong PSD decrease especially for frequencies below 2 kHz.

However, high standard deviation was found for all the configurations and vowels, baseline included. This might be related both to the mixture of male and female parameters used here and to the time of day when the recordings were made, i.e. at the end of the working day. Consequently, also the baseline parameters may have suffered from some distortion due to voice fatigue.

Work is ongoing to detect the influence of masks and shield on articulation in conversational voice and speech.

Though mandatory, the use of masks and shields might have negative impact especially in professions that make large use of voice such as teachers. These preliminary results suggest that some vocal exercise such as bubbling and face gym would be advisable at least for professionals [10].

V. Conclusion

This paper presents preliminary results on the impact of protective masks and shield on voice parameters estimated on the three basic sustained vowels of the Italian language /a/, /i/ and /u/. Recordings were made after a working day and concern 10 adult healthy subjects. Results confirm that voice energy decreases above 1 kHz especially when masks are worn along with the protective shield.

To the authors knowledge this is the first attempt to compare seven different masks configurations.

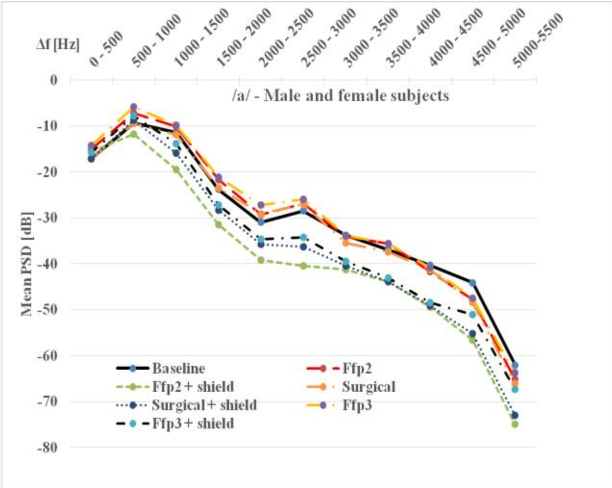

Figure 1 – /a/ vowel: dots correspond to the average PSD over 500Hz slices.

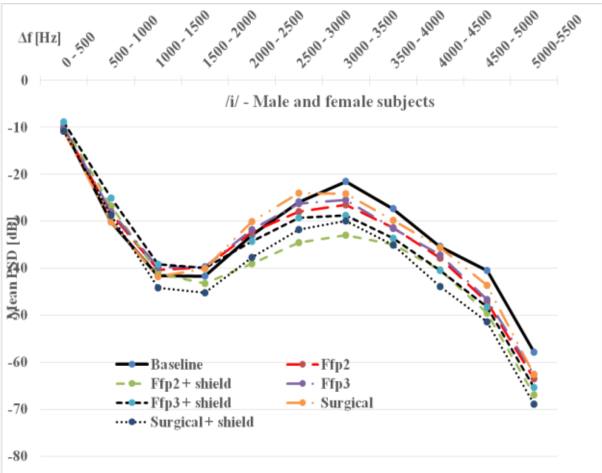

Figure 2 – /i/ vowel: dots correspond to the average PSD over 500Hz slices.

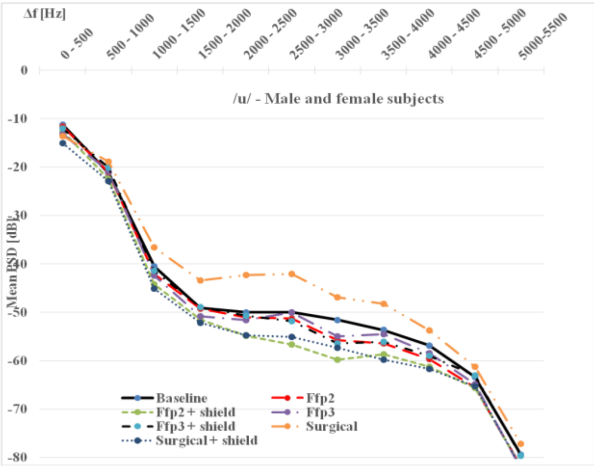

Figure 3 – /u/ vowel: dots correspond to the average PSD over 500Hz slices.

Future work will concern the analysis of vowels during articulation [11] and vocalic sentences. Moreover, self-perceptual evaluation based on a specific questionnaire will be performed.

In this work only non-invasive measures are considered, based on voice recordings and the acoustical analysis of the signal. More invasive analysis could be performed such as videokymography that might provide further helpful parameters [12].

REFERENCES

1. Pörschmann C, Lübeck T, Arend JM. Impact of face masks on voice radiation. J Acoust Soc Am. 2020 Dec;148(6):3663. doi: 10.1121/10.0002853. PMID: 33379881; PMCID: PMC7857507.
2. Goldin A, Weinstein BE, Shiman N. How do medical masks degrade speech perception? *Hearing Review*. 2020;27(5):8-9
3. Corey RM, Jones U, Singer AC. Acoustic effects of medical, cloth, and transparent face masks on speech signals. J Acoust Soc Am. 2020 Oct;148(4):2371. doi: 10.1121/10.0002279. PMID: 33138498; PMCID: PMC7857499.
4. Magee M, Lewis C, Noffs G, Reece H, Chan JCS, Zaga CJ, Paynter C, Birchall O, Rojas Azocar S, Ediriweera A, Kenyon K, Caverlé MW, Schultz BG, Vogel AP. Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. J Acoust Soc Am. 2020 Dec;148(6):3562. doi: 10.1121/10.0002873. PMID: 33379897; PMCID: PMC7857500.
5. Cavallaro G, Di Nicola V, Quaranta N, Fiorella ML. Acoustic voice analysis in the COVID-19 era. Acta Otorhinolaryngol Ital. 2021 Feb;41(1):1-5. doi: 10.14639/0392-100X-N1002. Epub 2020 Nov 24. PMID: 33231205; PMCID: PMC7982755.
6. Magee M, Lewis C, Noffs G, Reece H, Chan JCS, Zaga CJ, Paynter C, Birchall O, Rojas Azocar S, Ediriweera A, Kenyon K, Caverlé MW, Schultz BG, Vogel AP. Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols. J Acoust Soc Am. 2020 Dec;148(6):3562. doi: 10.1121/10.0002873. PMID: 33379897; PMCID: PMC7857500.
7. Deller J R, Proakis J G, Hansen J H L. Discrete-Time Processing of Speech Signals**,** Macmillan Coll Div, 1993, ISBN 10: 0023283017ISBN 13: 9780023283017,
8. Morelli M. S., Orlandi S., Manfredi C. BioVoice: A multipurpose tool for voice analysis. Biomedical Signal Processing and Control 2021 64,102302 doi:10.1016/j.bspc.2020.102302
9. Manfredi C, Barbagallo D, Baracca G, Orlandi S, Bandini A, Dejonckere P. H. Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers, Journal of Voice, 2015, Vol.29(4), p.517.e1-517.e9 ISSN: 0892-1997, 1873-4588; DOI: 10.1016/j.jvoice.2014.09.014
10. Di Natale V, Cantarella G, Manfredi C, Ciabatta A., Bacherini C, DeJonckere P.H. Semioccluded Vocal Tract Exercises Improve Self-Perceived Voice Quality in Healthy Actors. Journal of Voice 2020. ISSN: 0892-1997,1873-4588; DOI: 10.1016/j.jvoice.2020.07.024
11. Bandini, A., Orlandi, S., Giovannelli, F., Felici, A., Cincotta, M., Clemente, D., Vanni P., Zaccara G., Manfredi, C. Markerless analysis of articulatory movements in patients with Parkinson's disease. Journal of Voice 2016, 30(6),766.e1-766.e11. doi:10.1016/j.jvoice.2015.10.014.
12. Piazza, C; Mangili, S, Del Bon, F, Gritti, F, Manfredi, C, Nicolai, P, Peretti, G.,Quantitative analysis of videokymography in normal and pathological vocal folds: a preliminary study European archives of oto-rhino-laryngology. , 2012, Vol.269(1), p.207-212 ISSN: 0937-4477 , 1434-4726; DOI: 10.1007/s00405-011-1780-y