PAN ARCHIVES OF ACOUSTICS
Vol. **43**, No. 2, pp. 307–311 (2018)

DOI: 10.24425/122378

Transfer Characteristics of Vocal Tract Closed by Mask Cavity

Milan VOJNOVIĆ[(1)], Miomir MIJIĆ[(2)], Dragana ŠUMARAC PAVLOVIĆ[(2)]

[(1)] *Life Activities Advancement Center*
Gospodar Jovanova 35, 11000 Belgrade, Serbia; e-mail: vojnovicemilan@yahoo.com

[(2)] *School of Electrical Engineering*
Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia; e-mail: {emijic, dsumarac}@etf.rs

(*received September 10, 2017; accepted December 31, 2017*)

This paper analyses the changes in transfer characteristics of the vocal tract when closed by a mask, i.e. a chamber. The analysis was performed in two ways: by analytical estimation and by measurements in the vocal tract physical model for the case of mask with inner volume $V = 430$ cm^3, corresponding to the oxygen masks used in combat airplanes. It was shown that closing the vocal tract with a mask cavity increases the first formant frequency by about 10% in front and high vowels (/e/, /i/, and /u/) and the frequencies of the first two formants by about 5% in the remaining two vowels (/a/ and /o/). It was also revealed that longitudinal and transversal resonances in the mask chamber can lead to errors in the recognition of the vowel formant frequencies. The results point to the need for additional knowledge about resonances in mask application.

Keywords: mask; transfer characteristics; vocal tract.

1. Introduction

Transfer function of a vocal tract is defined by its resonant frequencies where the maxima of response appear. During speech, the articulators change their positions and consequently the physical shape of the vocal tract changes. The correlation exists between the physical shape of the vocal tract and its transfer characteristics. Based on the vocal tract transfer function one can estimate its physical form, and *vice versa*.

Speech is a non-deterministic process and the pronunciation of a phoneme varies from person to person and even when repeatedly pronounced by the same person. However, as long as the relevant parameter changes in pronounced phonemes are within certain limits, the human perceptual system correctly detects what is spoken. The problem of defining the discrimination range for each phoneme is a complex task and is mainly based on listening tests performed by linguists, speech therapists, phoneticians, etc. (Umeda, 1975). Objective indicators of correct articulation of phonemes are based on statistically defined parameters obtained by measurement with a number of subjects. In preliminary studies only percentage differences of resonance frequencies are analysed in order to explore possible variations in each phoneme pronunciation (Hillenbrand *et al.*, 1995).

When the pronunciation of phonemes is out of their discrimination ranges, an atypical or pathological pronunciation results. There are some circumstances when speech degradation occurs due to specific conditions in which the communication is performed. Such circumstance is when the speaker uses some type of mask: oxygen, diving, or protective. Wearing a mask changes the vocal tract configuration because it is closed by a chamber formed between the mask wall and the speaker's face. Thus, its transfer characteristics also changes, i.e. there is a shift of resonant frequencies.

Moreover, there is a series of additional difficulties in speech caused by wearing a protective or an oxygen mask described in literature, such as: Lombard effect (Bond, Moore, 1990), change in mouth opening impedance (Vojnović *et al.*, 2017), change in speaking rate (Bond *et al.*, 1989), changes in articulation (Bond *et al.*, 1989), mask respiratory valves noise (Vojnović, Mijić, 1997). There is also some influence of various breathing gas mixtures, positive pressure breathing, etc. Automatic recognition of speech with mask is complex (Wang *et al.*, 2016) and requires analysis of each individual problem.

The cross-section of the vocal tract tube is less than 20 cm^2. In the frequency range where the formants occur one can conclude that the waves travel longitudinally along the vocal tract tube as plane due to the re-

lation between the sound wavelength and its transversal dimensions. However, the transverse dimensions of the mask are usually much larger than the diameter of the vocal tract so, in addition to longitudinal, there are also sound waves moving in transverse directions. Transverse resonances occur in the acoustic structure where the sound wavelength is shorter than the transverse dimension. Thus, both longitudinal and transverse resonances in the mask chamber appear as integral parts of the vocal tract overall transfer characteristics.

For the mask chamber in the form of a cylindrical tube with length l and radius a, the resonances are defined by (Morse, 1986):

$$f_{m,n,n_x} = \frac{c}{2}\sqrt{\left(\frac{n_x}{l}\right)^2 + \left(\frac{\alpha_{mn}}{a}\right)^2}, \quad n_x = 0, 1, 2, \ldots, \quad (1)$$

where c is the speed of sound. Parameter n_x defines longitudinal chamber resonances and α_{mn} presents transversal resonances. The coefficient α_{mn} is mainly given in the table (Morse, 1986). For a chamber that is not in the form of a cylindrical tube, the resonances differ from those defined by Eq. (1) to a certain extent, but the equation can be used for rough estimation.

The paper presents a part of the results of a complex analysis of machine speech recognition in aircraft systems control. It is concerned with a study of the vocal tract transmission characteristics when closed by a mask. In the analysis, a vocal tract acoustic model described in the literature was used (Lin, 1994). The transfer characteristics analysis of the vocal tract when closed by a mask was performed in two ways: by measurements on its physical model and by analytical estimation using an equivalent electrical scheme.

2. Method of vocal tract transfer characteristics analysis

Five different vocal tract configurations were analysed corresponding to five Russian pronunciations of vowels for which the precise dimensions can be found in literature (Fant, 1970). In the analysis, the vocal tract shape for each vowel is approximated by a tube made of cylindrical segments 0.5 cm in length and corresponding cross-sections. At its end a mask is assumed in the form of a calotte with internal volume $V = 430$ cm^3 (5.5 cm height and 11.4 cm diameter). This size corresponds to a pilot oxygen mask. In vocal tract modelling the mask chamber was also approximated with cylindrical segments of 0.5 cm length and with appropriate diameters.

The speech communication with an oxygen mask means that the microphone is placed inside the mask and vocal tract output is monitored in the electrical domain. In the analytical estimation of the vocal tract transfer function each cylindrical segment is modelled with an appropriate equivalent electrical T-network (Fant, 1970; Flanagan, 1972). Impedances of each symmetrical T-networks were determined according to the cylindrical segment length and cross-section areas. With such a procedure, a vocal tract acoustical model is converted into an equivalent electrical model. The calculation of transfer function was conducted by an algorithm presented in the literature (Badin, Fant, 1984). The applied procedure of cylindrical tube modelling in the electrical domain included the viscous and thermal losses that occur in sound propagation. The impedance of mouth opening was presented in the model by an equivalent electrical circuit simulating the radiation of piston mounted on the sphere with a diameter of 18 cm (Vojnović, Mijić, 2005). The influence of the vocal tract walls impedance was neglected, as well as the impedance of glottis and subglottic system, all taken to be infinitely large. The speed of sound was taken in the range 342.89–344.65 m/s and the air density in the range $1.208 \cdot 10^{-3}$–$1.196 \cdot 10^{-3}$ g/cm^3, corresponding to the variation of ambient temperature of about 19–22°C. This was done in order to compare the results of the formant frequencies analytical estimation and results of their measurement in the physical model, due to the fact that during the measurement in the laboratory the temperature was in the range 19–22°C.

In parallel with the analysis based on equivalent electrical circuits, experimental measurements were performed on a vocal tract physical model closed with a mask whose geometry is identical to the model used in the analytical estimation. A block diagram of the experimental setup for measurement in the physical model is shown in Fig. 1. The model was realised using

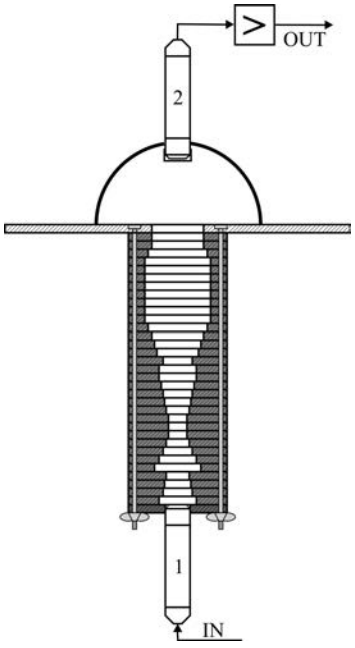

Fig. 1. Block diagram of the experimental setup by a physical model for determination of the vocal tract transmission characteristics: 1– microphone used as a sound source, 2 – measuring microphone.

0.5 cm length cylindrical segments made of hard plastic with an external diameter of 65 mm. In the centre of each segment there is an opening with an appropriate radius representing a particular part of the vocal tract.

As a sound source in the model a $1/2''$ condenser microphone was used, indicated by number 1 in the picture. As a receiver, a second microphone of the same type was placed in the mask chamber, marked with 2 in the picture. Both microphones were placed in the model structure axis.

3. Results

The results of the analytically estimated and measured vocal tract formant frequencies for five vowels are shown in Table 1. The frequency values preceded by label "(Long.)" and "(Trans.)" refer to longitudinal and transversal resonances in the mask chamber, respectively. Those resonant frequencies affect the change in formant structure, i.e. the order of estimated formants. All values are presented with a precision of 1 Hz.

As one can see from Table 1, all formant frequencies of vowels estimated by the equivalent electrical model are higher than measured in the physical model. The average difference between the analytically estimated and measured values is 5.3%, with the maximum difference of 10.8%. The largest percentage differences were obtained for the fourth and fifth formant frequencies (average 8.2% and 5.9%, respectively). The analysis of the difference for each vocal shows that the largest percentage difference is for vocal /u/ (6.7% on average) and vocal /e/ (5.9% on average). The average difference of formant frequencies for the remaining three vocals is practically consistent at around 4.6%.

Despite an effort made to ensure the same conditions in the analytical estimation by the equivalent electrical model and in the experimental analysis by the physical model (same acoustic structure, same ambient temperature, etc.), relatively large differences in the formant values were obtained by these two methods. The only major physical difference between the experimental measurement and the analytical estimation of formants was in the position of the measuring microphone inside the mask chamber (Fig. 1). In the analytical approach, it is assumed that the measuring microphone is at the level of the chamber wall, but in the experimental measurements the measuring microphone was positioned about ten millimetres inside the chamber. The difference also lies in the fact that in the analytical modelling it was assumed that the mask chamber had ideal rigid walls and fit perfectly on the end plate. In the experimental analysis of the physical model that certainly was not the case, because of the mask walls creating limited sound insulation and non-ideal sealing between mask edges and the end plate.

Both methods of vocal tract formant frequency analysis were repeated for the case when the same vocal tract was not closed by a chamber, but the sound of the mouth opening was radiated into the free space. The obtained formant frequency values are shown in Table 2. The values in the table are displayed with an accuracy of 1 Hz.

Table 1. Experimentally measured (exp.) and analytically estimated (est.) formant frequencies when the vocal tract is closed by a mask chamber of volume $V = 430$ cm^3.

Formant		/a/		/e/		/i/		/o/		/u/	
		exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]
F1		634	652	433	446	229	240	498	520	236	254
F2		1070	1079	1862	1928	2151	2218	838	868	561	586
F3		2302	2405	2600	2738	2846	3007	2250	2322	2210	2319
F4		3365	3576	3322	3631	3366	3650	3140	3339	3258	3610
F5		3711	4032	3807	4123	4622	4659	3691	3908	3732	3946
Chamber resonances	(long.)	3255	3332	3210	3287	3255	3385	3253	3388	3258	3363
	(trans.)	4270	3930	4489	3937	3939	3930	4112	3917	4087	3924

Table 2. Experimentally measured and analytically estimated vocal formant frequencies for the vocal tract without a mask.

Formant	/a/		/e/		/i/		/o/		/u/	
	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]	exp. [Hz]	est. [Hz]
F1	601	622	395	407	212	221	466	488	216	230
F2	1024	1049	1856	1920	2156	2218	802	837	557	583
F3	2297	2401	2608	2742	2880	3018	2248	2319	2211	2319
F4	3430	3521	3401	3539	3386	3630	3140	3358	3272	3610
F5	3740	4027	3915	4076	4575	4642	3708	3903	3731	3946

The data in Table 2 show that there is a slightly better match between the analytically estimated and experimentally measured formant frequencies when the vocal tract is without a mask. The average difference between the two is 4.7%, and the maximum difference is 10.3%. Also in this case the largest percentage difference was for vocal /u/ (6.2% on average). Differences in the vocal tract fourth formant frequencies estimation was 8.2% when closed by a mask and 6.2% without a mask. In the fifth formant frequency estimation there is also a difference of 1% (5.9% with a mask and 4.9% without a mask). These differences may be a result of longitudinal and transverse resonances in the mask chamber because they are located in the frequency range where the fourth and fifth formants of vocal tract would otherwise be. It can be observed that in the estimation of the first three formant frequencies there is a similarity regardless of whether the vocal tract is closed by a mask chamber or not.

4. Discusion

The results presented in the paper reveal how much vocal formant frequencies change when the vocal tract is closed by a mask chamber of the size of a pilot oxygen mask. In Fig. 2 the percentage change in vowel formant frequencies when closing the vocal tract with a mask is presented. The solid line and squares present the results of the experimental measurement and the dashed line and hollow circles present the results of the analytical estimation. One can see that there are no significant differences between the values obtained by experimental measurement and by estimation using the equivalent electrical circuit. Changes in vowel formant frequencies appear in the range between −1% and +10%. For the vowels /a/ and /o/ closing the vocal tract with a mask chamber affects the first and the second formant by about 5%. For the front and high vowels (/e/, /i/, and /u/) changes are visible only for the first formant by about 10%.

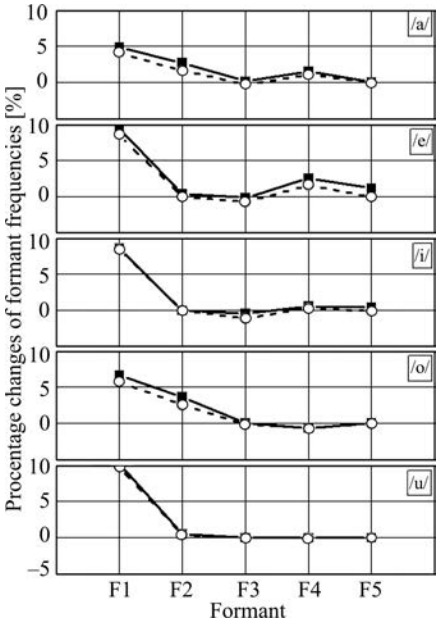

Fig. 2. Percentage change in vowel formant frequencies when the vocal tract is closed by a mask: solid line – measured in model, dashed line – results of the analytical estimation.

The presented results show that changes in formant frequencies introduced by a mask expressed in percentages are not significant. The largest changes are for the first two formant frequencies. However, the presence of longitudinal and transverse resonances inside the mask chamber disrupts the sequence of estimated formant frequencies and affects the recognition of the fourth and fifth formants. To illustrate such an occurrence, the calculated transfer characteristics of the vocal tract during the phonation of the vowel /e/ for normal speech (solid line) and for speech with a mask (dotted line) are presented in Fig. 3. The diagrams were obtained by analytical estimation with a vocal tract equivalent circuit.

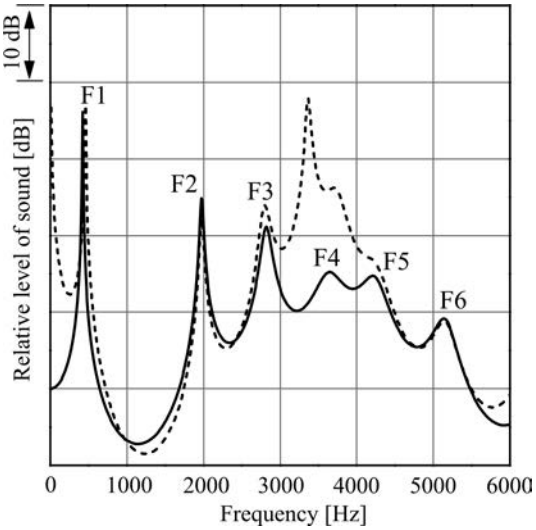

Fig. 3. Estimated transfer characteristics of the vocal tract during pronunciation of vowel /e/ for normal speech and speech with a mask of volume $V = 430\,\mathrm{cm}^3$: solid line – speech without mask, dashed line – speech with mask.

One can see that the resonance in mask cavity disturbed the formant structure of the vowel /e/. An "apparent" formant located around frequency 3400 Hz appeared in the transmission characteristics as a result of longitudinal resonance in the mask cavity. At the same time a transverse mask resonance near 4000 Hz made the fifth formant "invisible" because it has not appeared as a maximum in the characteristics, but only as an inflection point.

In circumstances when it is necessary to identify formants in speech signal for some reason, a possible consequence is that the longitudinal resonance of a mask cavity in the vicinity of 3400 Hz can be treated

as the fourth vowel formant, and all frequency formants above that will be moved one step up. The actual fourth formant be recognised as the fifth one, the actual fifth formant as the sixth one, and so on. The problem can be solved if the resonant frequency of the mask cavity is known in advance. Yet some caution should be taken with regard to circumstances when the resonances of cavity and of the vocal tract cannot be separated, as in the case when those two sets of frequencies overlap. In Fig. 3 that is the case with the transverse resonance of the mask cavity that covers the fifth formant of vowel /e/.

It is interesting to notice that speech from the microphone in an oxygen mask is clearly understandable and additional resonances in the mask cavity do not lead to misperceptions of either the vowels or other phonemes. It seems that the human perceptive system has an ability to eliminate stationary resonances by ignoring them because they do not carry speech information but only define the character and timbre of the voice.

5. Conclusion

Closing of the vocal tract by a mask chamber results in a formant frequencies change. That influence is twofold. Acoustic impedance of the chamber introduces certain increase of formant frequencies, mostly the first and the second formants. For the vowels /a/ and /o/ these changes imply an increase of the first and second formant by approximately 5%. For the front and high vocals (/e/, /i/, and /u/) changes occur only in the first formant by approximately 10%. The second aspect of the influence of the mask chamber is due to the introduction of its own longitudinal and transversal resonances. These resonances slightly shift the fourth and fifth formant, introducing formant reordering. The results have shown that the mask resonant frequency around 3400 Hz can be understood as the fourth vowel formant which can provoke certain errors in speech analysis.

Acknowledgments

This research was supported by grants TR32032 from the Ministry of education, science and technological development of the Republic of Serbia.

References

1. Badin P., Fant G. (1984), *Notes on vocal tract computation*, STL-QPSR 23/1984, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 53–108.
2. Bond Z., Moore T. (1990), *A note on loud and Lombard speech*, ICSLP, 969–972, Kobe, Japan.
3. Bond Z., Moore T., Gable B. (1989), *Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask*, Journal of the Acoustical Society of America, **85**, 2, 907–912.
4. Fant G. (1970), *Acoustic theory of speech production*, Mouton, The Hague.
5. Flanagan J.L. (1972), *Speech analysis, synthesis and perception*, Springer-Verlag, New York.
6. Hillenbrand J., Getty L.A., Clark M.J., Wheeler K. (1995), *Acoustic characteristics of American English vowels*, Journal of the Acoustical Society of America, **97**, 5, 3099–3111.
7. Lin Q. (1994), *Vocal-tract computation: How to make it robust and faster*, Journal of the Acoustical Society of America, **96**, 4, 2576–2579.
8. Morse P.M. (1986), *Vibration and sound*, Acoustical Society of America.
9. Umeda N. (1975), *Vowel duration in American English*, Journal of the Acoustical Society of America, **58**, 2, 434–445.
10. Vojnović M., Mijić M. (1997), *The influence of the oxygen mask on long-time spectra of continuous speech*, Journal of the Acoustical Society of America, **102**, 4, 2456–2458.
11. Vojnović M., Mijić M. (2005), *An improved model for the acoustic radiation impedance of the mouth based on an equivalent electrical network*, Applied Acoustics, **66**, 481–499.
12. Vojnović M., Mijić M., Šumarac Pavlović D. (2017), *A simplified model of mouth radiation impedance closed by mask cavity*, Applied Acoustics, **115**, 3–5.
13. Wang G.-Y., Zhao C.-Y., Xue X.-Z., Zhang J., Zhao X.-Q. (2016), *Correction of distortion mask speech based on parameter estimation of AR model*, International Conference on Audio, Language and Image Processing, 689–693.