Journal of the Acoustical Society of Nigeria (JASON) Vol. 7, 2020

ACOUSTIC INVESTIGATION OF SPEECH WITH AND WITHOUT FACEMASK: SOME EXPLANATIONS

Onyinye Anulika Chiemezie
Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.
oa.chiemezie@unizik.edu.ng

Abstract
Speech has been an integral part of human communication. However, the use of facemask during the Covid-19 pandemic has posed some challenges on humans' ability and freedom to use this medium effectively. This study compares speech with and without facemask to ascertain, the physical differences and the factors responsible for these differences. Eight monosyllabic root verbs in Igbo language serve as data. 7 respondents are used in this study; 2 males and 2 females between the ages of 19-40 rendered voice recordings while 3 normal hearing males and females each of 10-25 age range listened to the data for perceptual assessment. Surgical and double-layered fabric facemasks are used. Recorded data are analyzed using Praat software while respondents for perceptual assessment underwent a hearing test using Marcin Masalski's Hearing Test App. version 1.2.4. The results show that the mean of F_1 and F_2 of speech without facemask is slightly higher than those with the type3 facemasks used. The surgical and some double-layered fabric facemasks have similar formants with that of normal speech, while there is no significant perceptual difference in the perception of speeches from the different types of facemasks. However, perception and spectrogram indicate muffled nasal voiced speech and completely blurred spectrograms with strong overlapping formants of respondents wearing tightly-fitted facemasks also, there isgross low value inthe physical properties of respondents with tightly-fitted facemasks (at the nose region)marking acul-de-sac resonance disorder. Findings show that properly spaced facemasks do not hinder speech rather unlike tightly-fitted facemasks.

Keywords: Acoustic properties, Speech, Facemasks, Resonance

Introduction
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) known as Covid-19, is a type of pathogenic coronavirus with high degree of transmissibility. It originated in 2019, characterised basically by acute respiratory breakdown. The pandemic spread round different countries and is widely known for its main communicable channel, through respiratory particles. Transmission can be direct or indirect. Direct is by inhaling the droplets deposited in the air, from the sneeze, cough, talk, shout or singing of an infected person or indirectly coming in contact, by touching, with the virus deposited on some surfaces. The symptoms are also not immediately measured or manifest on a patient, making it difficult to place some form of control before transmission occurs. Additionally, the fact that it can be transmitted by pauci-symptomatic (infected people showing symptoms) pre-symptomatic (infected people, yet to develop symptoms), and asymptomatic (infected but do not show symptoms) persons. However, studies have also shown that the use of facemasks accompanied with other precautions, reduces the rate of the virus transmission. As the spread of the virus became a pandemic, the need to secure lives necessitated different governments and WHO to introduce minimizing physical interactions and the use of facemasks amongst other measures to fight the spread of the virus. This step to secure life consequently had a ripple effect on humans as social beings (Mheidly *et al.*, 2020). The effect goes beyond the inability to see and appreciate facial expressions, to affect psychological freedom, fashion, communication and speech. To this end, the following questions come up: To what degree does facemasks affect our speech production and perception? Are there acoustic differences in monosyllabic words of speech with or without facemasks, what are their degrees of similarity? Are speeches produced with facemask different in pitch, formant values, loudness and noise filtration? This paper seeks to find answers to these questions.

Objectives

The aim of this paper is to ascertain the similarities and dissimilarities in the physical properties of speeches with and without facemasks. The study sets specific objectives in order to achieve this aim. They are to:

(i.) ascertain the physical properties: F_o,F_1, F_2, and intensity of speeches produced with and without facemasks; and

(ii.) determine the factors responsible for significant differences in the spectral, formants and pitches of different types of facemasks.

Literature Review

Facemasks have been for a long time used by nurses and doctors in wards and operating theatres or/and by electrolyte engineers. These fields of study have also managed the problems of communication emanating from the use of facemasks. Today, facemask has become part of our everyday dressing worn by everyone, giving rise to a surge in communication problems. These communication or perceptual problem arises due to different factors. Carbon (2020) conducted a test on the impact of facemasks in understanding and communicating emotions. He used a random sample of 41 consultants and employed 12 different facial emotions in his test. The findings show that facemask hinders easy readability of emotions and sometimes enhances the misinterpretation of emotions. He suggests that other body actions like gestures could be used to close this communication gap.

Studies have shown that facemask may reduce speech perception given the type and the environmental factors. Mendel *et al* (2008) opine that while surgical mask do not have negative impact on speech perception for normal and mild impaired hearing groups, noise from health care machines can be a factor especially in view of patients with hearing loss challenges. Atcherson *et al* (2017) attest that while there are significant differences in the spectral forms of speeches produced with and without facemasks, they observe that this did not deter communication quality. Rather, noise from the environment serve as a major factor in lowering the rate of perception. Another factor that may also deter speech perception is opaque facemask covering the lips. Not just that the facemask may obstruct or constrain airstream activity, the covering hinders lip reading which could have aided the listener to decode what is being said. Literature argues that such coverings affect the decoding of sounds with similar phonetic properties given peculiar environments and surrounding sounds. It was proposed that it is easier to perceive and differentiate the syllables /ta/ and /ka/ than the syllables /ti/ and /ki/ without lip reading, thus transparent facemask are somewhat recommended. Saedi *et al.* (2018) opine that the material for the face cover and the closeness of the cover to the organs of speech affect the sound production quality which will in turn reduce sound perception.

Magee *et al.* (2020) investigate speeches with and without facemask and their perceived intelligibility based on these acoustic properties: timing, frequency, perturbation and power spectral density. Speech intelligibility was conducted using three types of facemask- N95- an electrostatic non-woven polypropylene fiber containing a filtration layer, surgical and two layered cloth. This study employs words and sentences as speech samples. Findings show that there are no significant differences across these acoustic measurements for sentences and words while using these different facemask types. Also, data above word level attracted different levels of assimilation and other phonological processes, which agree that there is constant interaction in speech where sounds influence neighbouring sounds (Ladefoged 2003; Eme, 2008). Again, the characteristic factors responsible for these changes were not pointed out. This opens up a research gap that this study seeks to fill. The study compares the physical properties of speech produced with and without facemask using monosyllabic verb roots in Igbo language to minimize interference and wrong production of English words as second language learners.

Methodology

Eight (8) monosyllabic verb roots in Igbo language are used to elicit data. Two-segment words comprising plosives, fricatives and affricates and the front low vowel [a]. The idea is to reduce sound interaction and influence of anticipatory sound oriented movements and waves (Ladefoged 2003; Eme, 2008).Voiced and voiceless counterparts /pa,ba; ta,da; sa,za; tʃa, dʒa/ from each of these sound classes are used in the data presentation.2 males and 2 females rendered voice recording and 2 males and 2 females assessed the perception quality of the recordings. Nexton digital recorder with 44.1 wav setting is used for data elicitation while Praat software (Weenik and Boersma, 2011) was used for sound analysis. Hearing Test version 1.2.4 (Masalski,) was used to establish the hearing ability of the respondent for perceptual assessment. Formants 1 and 2, the Spectrogram, Pitch and Intensity form the bases for the analysis. The wax fabric, double layered and surgical are the three types of facemasks used for this analysis. Pulse lines are used to distinguish noise from sounds in the wave form.

Data Presentation

Speech samples are collected and the physical properties extracted using speech analyzer software. The different features are represented in the appendices' section. This section will only employ the mean values relevant in the analysis. Table 1 presents the mean values of speeches produced with different facemasks.

Table 1: Mean values of F1 produced with different facemasks

	Mean of all four consultants for F1(Hz)							
F1	**[pa]**	**[ba]**	**[tá]**	**[dá]**	**[sá]**	**[zá]**	**[tʃá]**	**[dʒá]**
NS	816	849	815	808	818	944	931	883
SFS	802	829	804	806	794	936	906	865
FaFS	798	824	799	798	788	922	894	857
TTFS	492	485	456	450	460	467	495	471

Where NS= Normal Speech; SFS= Surgical Facemasked Speech; FaFS= Fabric Facemasked Speech; TTFS=Too Tight Facemasked Speech.

It is observed that F1 of the Normal speech has similar value to the speech produced with surgical facemask. While the fabric facemask is a little different, it is not significant, does not hinder perception and it is also not consistent. It is more similar to the speech produced with surgical facemask than that of normal speech. It was also observed that some fabric facemasks generated very low F1 which led to categorizing them into another group given the mild discomfort it gives the consultants at the back of the ear, bridge of the nose or pressing against the lips. This relationship and degree of similarity is shown on the line graph below.

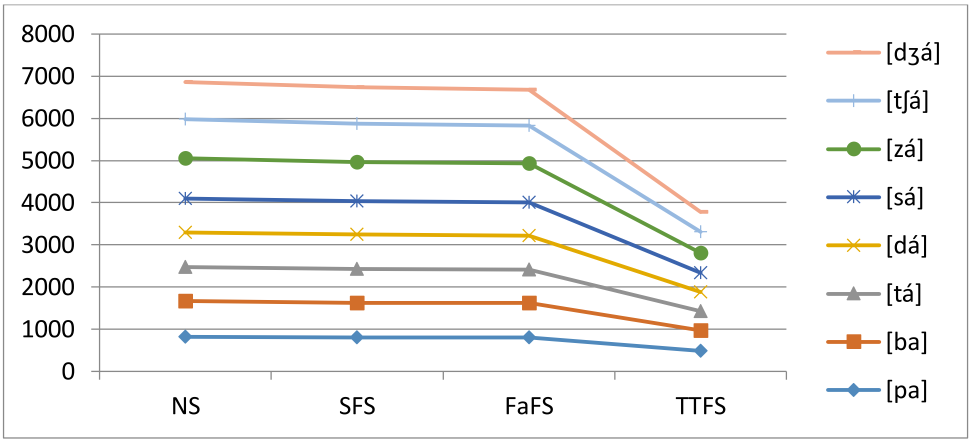

Figure 1: Comparing the F1 of different speech samples.

From the lines, it is clear that there is a steady but mild decline in value from normal speech to that of surgical facemasks then to a comfit fabric and that of a tightly-fitted fabric. The tightly-fitted is significantly different from the other three. The same pattern of value difference is observed in formant 2, however, on a lighter degree

Table 2: Mean values of F2 produced with different facemasks

	Mean of all four consultants for F2 (Hz)							
F2	[pa]	[ba]	[tá]	[dá]	[sá]	[zá]	[tʃá]	[dʒá]
NS	1511	1555	1555	1588	1569	1559	1534	1605
SFS	1507	1545	1552	1585	1556	1568	1564	1600
FaFS	1501	1538	1544	1578	1559	1559	1556	1590
TTFS	1214	1328	1272	1351	1246	1316	1362	1389

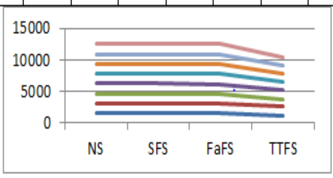

Figure 2: Mean value F2 of different speech patterns

Where NS= Normal Speech; SFS= Surgical Facemasked Speech;
FaFS= Fabric Facemasked Speech; TTFS=Too Tight Facemasked Speech.

Table 3: Mean values of Pitch produced with different facemasks

	Mean of all four consultants for pitch (Hz)							
Fo	[pa]	[ba]	[tá]	[dá]	[sá]	[zá]	[tʃá]	[dʒá]
NS	165	166	167	167	167	167	167	164
SFS	164	165	164	166	166	166	166	163
FaFS	165	165	165	167	167	167	168	165
TTFS	152	151	149	143	152	150	149	147

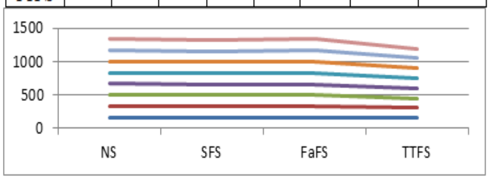

Figure 3: Mean values of Fo of different speech samples

Where NS= Normal Speech; SFS= Surgical Face masked Speech;
FaFS= Fabric Face masked Speech; TTFS=Too Tight Face masked Speech.

Table 3 shows the mean value of the fundamental frequencies of the different speech samples, while figure 3 gives a visual relationship they share which is consistent with formants 1 and 2.

All these productions could be said to be comparable given that they are averagely produced with the same frequency of the vibration of the vocal folds characterized in F_o and the same degree of loudness. See Table 4, Figure. 4, for the mean values on Intensity.

Table 4: Mean values of Intensity produced with different facemasks

	Mean of all four consultants for Intensity (dB)							
Intensity	[pa]	[ba]	[tá]	[dá]	[sá]	[zá]	[tʃá]	[dʒá]
NS	72	72	71	71	72	72	71	70
SFS	74	73	72	73	73	73	73	72
FaFS	76	75	74	74	74	74	73	72
TTFS	69	69	69	68	70	71	70	70

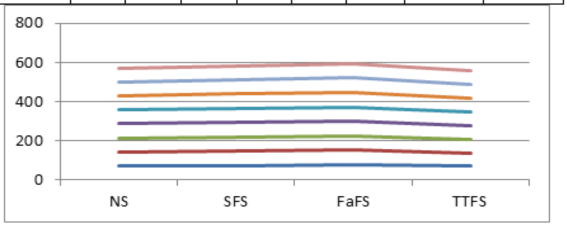

Figure 4: Mean value of loudness in different speech samples

Where NS= Normal Speech; SFS= Surgical Face masked Speech; FaFS= Fabric Face masked Speech; TTFS=Too Tight Face masked Speech.

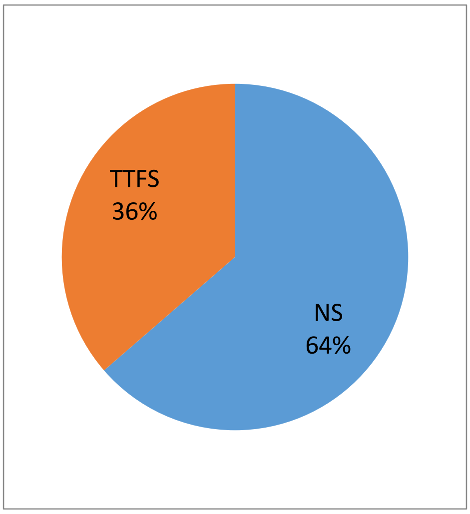

Figure 5: Percentage value of F1 in NS and TTFS

While F1 remains the highest degree of difference, it is still not significant when NS compared with the SFS and FaFS. However, when NS, SFS and FaFS are compared with TTFS, there is a significant difference. A sample of NS, 64% in value and TTFS, takes 36% value, giving a very high difference of 0.27. Surprisingly, this huge difference is not indicated in some the sound spectral. The spectrum to the left is NS while the spectrum to the right is TTFS. The study employs intensity and nasality values to tackle this problem.

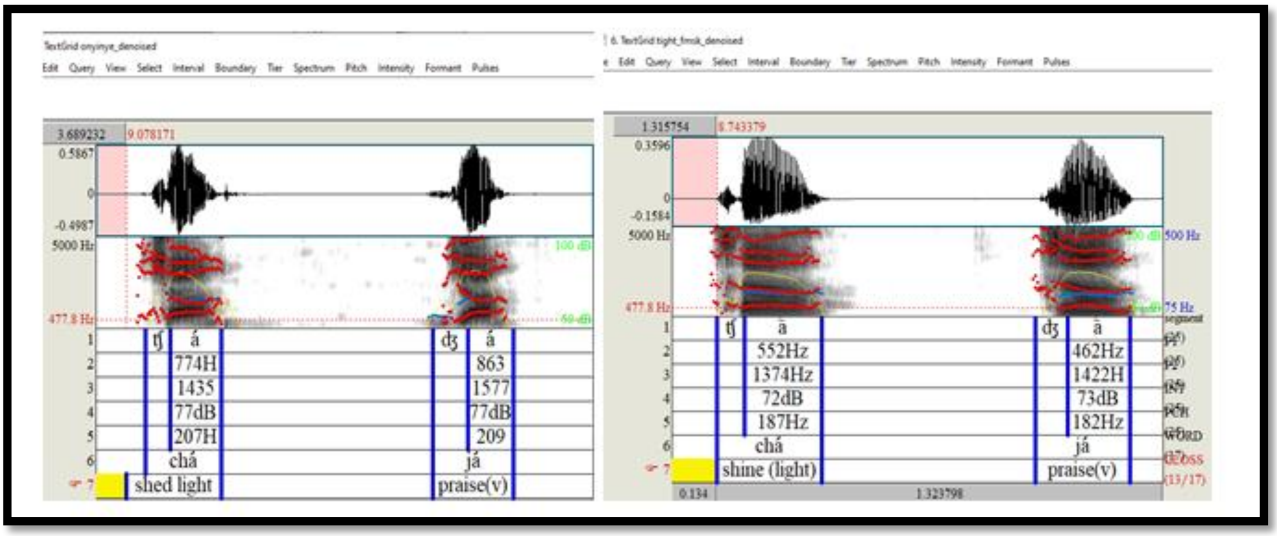

Figure 6: Spectral of speech samples, left for NS and right, TTF

Nasality is the measure of the degree of velopharyngeal opening in voiced speech formed by computing the ratio of the amplitude of the acoustic energy, results to low intensity (Sadjad et al., 2010). This explains the low intensity. It is attested in literature that the physical properties of nasalized vowels are usually higher than their oral counterparts (Veniranda, 2015; Alerechi and Aniagboso, 2019). Veniranda (2015) claims that the nasalized open front vowel [a], has a significant higher value than its counterpart evidenced in Teochew, one of the Min Chinese dialects. Manyah (2011) attests that F2 and F3 are lower in nasalized vowels compared to that their oral counterparts in Twi language. Logically, these scholars generally agree that F1 of nasalized vowels are higher than the oral pairs. However, the data in this paper, not corresponding with these assertions confirms that normal nasality may not be the case observed. The discomfort at the nose region closes down the nostrils thereby blocking the free flow of air through the nostrils in the production of vowel sounds. According to Cincinnati Children's Hospital Medical Center, such obstruction may likely result to a type of disorder called Cul-de-sac resonance which occurs when resonating sound is trapped in a cavity (Kummer and Lee, 1996).Formants correspond to the resonance in the vocal tract, therefore, when the closed nostril strap air in the nasal cavity, air builds up and naturally pushes back, the force lowers the larynx, sucking in the air (Yul-Ifode, 2008). Words with such ingressive intervention are likely to be muffled. This lowered larynx and low resonance results in low formants (Sundberg and Nordstrom, 1976). Again, Nasals have weak harmonics (Tarnóczy, 1948), andare also characterized by faint spectrogram (Ladefoged, 2003). The sound waves to the right of Fig.6 have shadowy images and their spectrograms a little smoky characterizing filtration of nasality. Given this arguments, the study concludes that cul-de-sac resonance disorder and lowered larynx are responsible for the low acoustic properties and muffled output in TTFS.

To drive home this point, a speech sample of TTFS, where tightness is at the mouth region was analyzed is shown in Fig.7. The formants are considerable closer in value to their other counterparts but the formants are greatly distorted. See Fig 7.

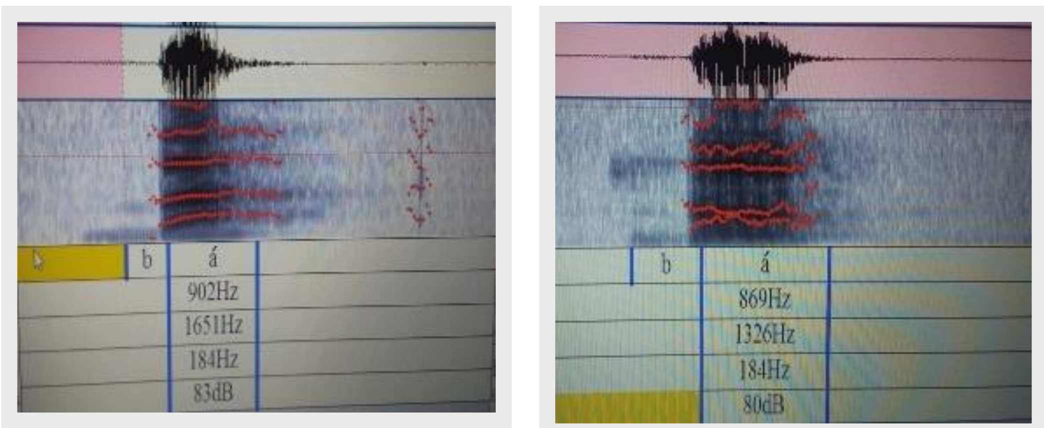

Figure 7: Spectrograms of NS AND TTFS at the mouth region. NS to the left and TTFS to the right.

Conclusion

This study has successfully compared the physical properties of sounds; bringing out differences between speeches with and without facemask. The study also determines the factors responsible for the differences observed. This study concludes that there is no significant difference in speech with and without facemask. The study notes that the comfort of the facemask is a major factor either in the reduction in measurements of acoustic properties or distortion of speech.

References

Alerechi, R.I.C. & Aniagboso, O. A. (2019). The syllable structure of Gbari. *The International Journal of Humanities and Social Sciences*, 7(12), 215-225.

Atcherson, S.R., Mendel, L.L., Baltimore, W.J., Patro, C., Lee, S., Pousson, M. & Spann, M.J. (2017). The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss. *J Am Acad Audiol*. 28, 58–67.

Carbon, C-C. (2020). Wearing facemasks strongly confuses counterparts in reading emotions. In Sokolowska, J. (Ed.), *Front. Psychol*. 11, 566-886.

Cincinnati Children's Hospital Medical Center (2015). Resonance disorders. *Speech-Language Pathology*. http://www.cincinnatichildrens.org/health/s/speech-disorder/

Eme, C. A. (2008). *Parameters of intersegmental co-ordination in speech. Insight from Igbo*. Awka: Amaka Dreams Ltd.

Kummer, A.W., Lee, L. (1996). Evaluation and treatment of resonance disorders. *Language Speech and Hearing Services in Schools* 27(3). DOI: 10.1044/0161-1461.2703.271

Ladefoged, P.(2003). *Phonetic analysis: an introduction to field work and instrumental techniques.* Malden: Blackwell Publishers.

Magee,M., C. Lewis, Noffs, G., Reece, H., Chan,J.C.S., Zaga, C.J., Paynter, C., Birchall, O., Azocar, S.R., Ediriweera, A., Caverlé, M.W., Schultz, B.G., & Vogel, A.P. (2020). *Effects of facemasks on acoustic analysis and speech perception: implications for peri-pandemic protocol*. Available from: https://www.researchgate.net/publication/344622563_Effects_of_face_masks_on_acoustic_analysis_and_speech_perception_Implications_for_peri-pandemic_protocols [accessed Dec 07 2021].

Manyah, A.K. (2011). Oral-nasal vowel contrasts: new perspectives on a debated question. *ICPhS* (XVII), 200-203.

Mendel, L.L., Gardino, J.A. & Atcherson, S.R. (2008). Speech understanding using surgical masks: a problem in health care?*J. Am Acad Audiol*, 19,686–695.

Mheidly N, Fares MY, Zalzale H and Fares J (2020). Effect of Facemasks on Interpersonal Communication during the COVID-19 Pandemic. *Front. Public Health* 8:582191. Doi: 10.3389/fpubh.2020.582191

Tarnóczy, T. (1948). Resonance data concerning nasals, laterals and trills. *WORD*, 4(2), 71-77.

Sadjad, V., Ghorban, A.A., Torabinezhad, F., Amiri, Y., & Keyhani, M.R. (2010). *The effect of vocal loudness on Nasalance of vowels in Persian adults. Iranian Rehabilitation Journal, 8,* 31-35.

Sundberg, J. & Nordstrom, P.E. (1976). Raised and lowered larynx - the effect on vowel formant frequencies. *TL-QPSR, 17*(2-3),35-39.

Veniranda, Y. (2015). Oral and nasal vowels in Pontianak Teochew. *A Journal of Language and Language Teaching*, 18(2), 107-124.

WHO (2020). *Transmission of SARS-CoV-2: implications for infection prevention precautions. Scientific Brief*. **https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions**.

Yul-Ifode, S. (2008). *Basic phonetics*. Port Harcourt: University of Port Harcourt Press

Appendices

Appendix 1: Sound properties for speech without facemask

		[pa]	[ba]	[tá]	[dá]	[sá]	[zá]	[tʃá]	[dʒá]
Male 1	F1 (Hz)	697	805	676	733	714	980	957	815
Male 2	F1 (Hz)	723	779	703	746	720	982	969	824
Female 1	F1 (Hz)	890	845	881	841	887	906	774	863
Female 2	F1 (Hz)	953	962	999	912	950	907	1023	1029
	Mean	816	849	815	808	818	944	931	883
Male 1	F.2 (Hz)	1451	1532	1465	1453	1448	1465	1410	1509
Male 2	F2 (Hz)	1516	1596	1651	1748	1646	1649	1675	1750
Female 1	F2 (Hz)	1530	1520	1531	1548	1565	1571	1435	1577
Female 2	F2 (Hz)	1546	1571	1572	1603	1577	1606	1617	1582
	Mean	1511	1555	1555	1588	1569	1559	1534	1605
Male 1	Pitch (Hz)	105	100	102	101	102	101	101	99
Male 2	Pitch (Hz))	130	132	141	139	142	140	143	142
Female 1	Pitch (Hz)	209	212	208	210	208	209	207	209
Female 2	Pitch (Hz)	217	219	219	217	215	221	218	206
	Mean	165	166	167	167	167	167	167	164
Male 1	Intensity (db)	72	69	68	67	70	70	68	66
Male 2	Intensity (db)	70	69	68	68	69	71	69	69
Female 1	Intensity (db)	76	78	76	77	76	78	77	77
Female 2	Intensity (db)	69	71	71	71	71	70	69	69

	Mean	72	72	71	71	72	72	71	70

Appendix 2: Sound properties for speech with surgical facemask

consultants	words → / Acoustic features ↓	[pá]	[bá]	[tá]	[dá]	[sá]	[zá]	[ʧá]	[dʒá]
Male 1	F1 (Hz)	695	806	674	730	714	970	915	803
Male 2	F1 (Hz)	718	752	710	735	711	967	949	812
Female 1	F1 (Hz)	875	825	861	841	827	898	770	843
Female 2	F1 (Hz)	921	932	970	919	925	907	989	1001
	Mean	802	829	804	806	794	936	906	865
Male 1	F2 (Hz)	1449	1530	1466	1450	1445	1450	1425	1513
Male 2	F2 (Hz)	1509	1556	1648	1745	1640	1637	1659	1737
Female 1	F2 (Hz)	1530	1525	1529	1535	1570	1573	1555	1577
Female 2	F2 (Hz)	1540	1569	1565	1611	1569	1610	1615	1572
	Mean	1507	1545	1552	1585	1556	1568	1564	1600
Male 1	Pitch (Hz)	107	110	105	103	107	108	105	102
Male 2	Pitch (Hz))	129	131	131	135	132	130	137	135
Female 1	Pitch (Hz)	205	207	208	211	208	209	209	207
Female 2	Pitch (Hz)	215	213	212	215	216	215	214	208
	Mean	164	165	164	166	166	166	166	163
Male 1	Intensity (db)	73	71	70	70	71	70	69	68
Male 2	Intensity (db)	72	71	70	70	70	71	70	69
Female 1	Intensity (db)	78	78	77	78	77	78	78	77
Female 2	Intensity (db)	72	71	72	73	73	74	73	72
	Mean	74	73	72	73	73	73	73	72

Appendix 3:Sound properties for speech with fabric facemask

consultants	words → / Acoustic features ↓	[pá]	[bá]	[tá]	[dá]	[sá]	[zá]	[ʧá]	[dʒá]
Male 1	F1 (Hz)	690	802	669	723	710	956	901	801
Male 2	F1 (Hz)	715	747	708	729	705	951	938	806
Female 1	F1 (Hz)	869	819	854	838	820	877	759	827
Female 2	F1 (Hz)	916	927	965	902	917	902	977	992

	Mean	798	824	799	798	788	922	894	857
Male 1	F2 (Hz)	1440	1525	1457	1445	1445	1443	1418	1500
Male 2	F2 (Hz)	1501	1548	1642	1737	1632	1630	1651	1729
Female 1	F2 (Hz)	1523	1518	1520	1527	1560	1562	1547	1568
Female 2	F2 (Hz)	1540	1562	1557	1602	1559	1600	1609	1562
	Mean	1501	1538	1544	1578	1559	1559	1556	1590
Male 1	Pitch (Hz)	108	110	107	107	107	108	108	107
Male 2	Pitch (Hz))	130	130	133	135	136	133	137	135
Female 1	Pitch (Hz)	208	208	208	210	209	210	211	209
Female 2	Pitch (Hz)	215	213	213	215	215	217	215	210
	Mean	165	165	165	167	167	167	168	165
Male 1	Intensity (db)	75	73	73	73	72	72	70	69
Male 2	Intensity (db)	73	72	72	71	73	73	72	70
Female 1	Intensity (db)	80	79	78	78	78	78	78	77
Female 2	Intensity (db)	75	74	73	73	73	74	73	72
	Mean	76	75	74	74	74	74	73	72

Appendix 4: Sound properties for speech with tight facemask

consultants	words → / Acoustic features ↓	**[pá]**	**[bá]**	**[tá]**	**[dá]**	**[sá]**	**[zá]**	**[tʃá]**	**[dʒá]**
Male 1	F1 (Hz)	450	445	437	431	446	451	472	469
Male 2	F1 (Hz)	475	468	435	426	443	453	467	465
Female 1	F1 (Hz)	523	516	481	476	480	489	552	462
Female 2	F1 (Hz)	518	509	469	468	470	475	489	487
	Mean	492	485	456	450	460	467	495	471
Male 1	F2 (Hz)	1105	1285	1228	1337	1232	1305	1348	1350
Male 2	F2 (Hz)	1213	1298	1275	1348	1244	1316	1359	1365
Female 1	F2 (Hz)	1273	1374	1295	1361	1259	1322	1374	1422
Female 2	F2 (Hz)	1265	1356	1288	1359	1249	1322	1368	1419
	Mean	1214	1328	1272	1351	1246	1316	1362	1389
Male 1	Pitch (Hz)	101	100	100	99	105	106	107	106

Male 2	Pitch (Hz))	120	121	119	117	128	129	127	125
Female 1	Pitch (Hz)	199	199	194	185	188	183	187	182
Female 2	Pitch (Hz)	187	185	182	172	185	181	175	173
	Mean	152	151	149	143	152	150	149	147
Male 1	Intensity (db)	67	67	67	66	70	70	69	68
Male 2	Intensity (db)	68	68	68	67	70	71	69	69
Female 1	Intensity (db)	72	71	72	71	71	71	72	73
Female 2	Intensity (db)	70	70	70	69	70	70	69	69
	Mean	69	69	69	68	70	71	70	70