INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Combining acoustic and aerodynamic data collection: A perceptual evaluation of acoustic distortions

Amélie Elmerich[1], *Jiayin Gao*[1], *Angélique Amelot*[1], *Lise Crevier-Buchman*[1,2], *Shinji Maeda*[1]

[1]LPP (UMR7018, CNRS/Sorbonne Nouvelle), France
[2]Hôpital Foch: Service de Laryngologie Phoniatrie, Université Paris Saclay, Suresnes, France
{amelie.elmerich, angelique.amelot}@sorbonne-nouvelle.fr, jiayin.gao@cnrs.fr, lise.buchman1@gmail.com, smaeda75@gmail.com

Abstract

Combining acoustic and aerodynamic data acquisitions is challenging. Devices for aerodynamic measurements often create severe acoustic distortions, which make it impossible to analyse the simultaneously recorded acoustic data. An improved technique, a pneumotachograph mask made of synthetic fibers, is acoustically transparent while ensuring a high-quality aerodynamic data acquisition. A previous acoustic study confirms the minimal acoustic distortions caused by this technique. The present study evaluates the impact of different aerodynamic devices on the human perception of vowels. Results show that vowels recorded with a fiber mask are almost as accurately categorised as acoustic-only recordings, compared with rigid masks that result in perceptual confusions. Listeners are also less likely to perceive the presence of a mask. Overall, our study provides the perceptual validation of the fiber mask technique, which will be of a great value in the field of speech sciences.

Index Terms: aerodynamic, speech intelligibility, speech perception, acoustic distortion, pneumotachograph mask

1. Introduction

1.1. Overview of aerodynamic measurements

Aerodynamic measurements offer invaluable data for the study of speech. Quantification of the dynamic change in air pressure, oral and nasal airflows contributes to the analysis of complex phenomena including nasalisation, change in the friction noises, change in the phonation mode, etc.

Until the 1960s, the kymograph was used for collecting aerodynamic data [1], but the analysis was restricted to qualitative observations. The pneumograph could also be used to record changes in the volume of the chest cavity [2] but was not adapted to speech analysis because the change in the quantity of air during phonation would require a greater precision than what could be captured by the lever drum of the instrument.

It was in the 1920s that aerodynamic measurements made a real leap forward with the invention of the pneumotachograph and the use of the manometric mask [3]. In 1925, a pneumotachograph invented by Fleisch (known as Fleisch pneumotachograph, or PTG) [4] was used to study breathing activities. With their measurements of dynamic and quantitative flows, PTGs have become the preferred instruments for ventilometric measurements in respiratory physiology, and several aerodynamic studies on speech have been conducted [5, 6]. However, PTGs are poorly adapted to the characteristics of fast and highly dynamic articulatory flows. They are also sensitive to condensation and saliva aerosol projection.

Solutions have been proposed such as coupling PTGs with pressure sensors. Phoneticians have therefore been using the respiratory instruments at their disposal such as the Aerophone (Kay Elemetric) or the Rothenberg mask [7] developed for phonetic use. However, these instruments all have a high dead volume and an anesthesia mask obstructing the speaker's face, which has significant drawbacks. To optimise the measurement of simultaneous airflow at the mouth and nostrils, Teston developed in 1982 a device more suitable for speech flow measurements: the Polyphonometer III. The flow sensors consisted of PTGs with the smallest possible dead volume, combined with very sensitive pressure sensors. All precautions of linearity were taken to ensure reliable and calibratable quantitative measurements [8]. It was the first measuring instrument to have a vertical device for measuring nasal airflow and a flexible mouthpiece that did not restrict the movements of the mandible. The creation in 1995 of a system to overcome its disadvantages by reducing the dead volume and adding more sensitive pressure sensors was born under the name of Assisted Voice Assessment (EVA2^{TM}) [9]. In recent decades, it remains, together with the Scicon R&D Mask (Scicon R&D Inc.) and the Rothenberg Mask [7], the main tool for acquiring aerodynamic data on speech tasks. The Super Nasal-Oral Ratiometry Sytem (SNORS) mask [10] is also marketed but it is limited to giving qualitative rather than quantitative values.

1.2. Overcoming the acoustic distortions

Despite a tremendous improvement in the precision of aerodynamic measurements, acoustic data collected with aerodynamic masks are often distorted due to the acoustic resistance of rigid materials, and thus cannot be properly analysed.

The Rothenberg mask, due to its shape and rigidity [11], leads to a lowering of the formants, particularly for low vowels such as [a]. The first two formants are lowered by about 50 to 100 Hz due to the increase in length between the vocal tract and the mask [7, 11]. In addition, other studies have shown a dip between 1600 and 2000 Hz [12], a general attenuation of the amplitude above 1000 Hz combined with a slight peak at 1300 Hz and a slight valley from 2000 to 2500 Hz in the data recorded with the mask [13]. Recognising these problems, improvements have been continually made by adding holes around the Rothenberg mask to reduce acoustic resistance and thus attenuate these artefacts. The EVA2^{TM} mask, although flexible, acts as a low-pass filter that creates a strong resonance around 1.4 kHz and an anti-resonance phenomenon around 3.5 kHz [14]. The recorded audio signal is also hardly usable for acoustic analysis or perceptual experiments.

To overcome these limitations, a technique has been developed to adapt a disposable soft mask made of synthetic paper fiber to collect aerodynamic data (2009 patent Nº 0900696) [15]. This type of protective mask is designed to have a rela-

10.21437/Interspeech.2023-1918

tively low airflow resistance in order to ensure comfort for the user. Furthermore, this low resistance allows to measure airflow without disturbing the sound propagation, making the acoustic signal analysable. Indeed, if the resistance within the mask is too high, this leads to a significant increase in the damping of the formants [7].

This adapted fiber mask presents several other advantages, which are continuously improved for the purposes of public research of speech sciences. First, it removes the constraint of holding the mask by the speaker, which can potentially lead to aerodynamic losses in the event of speaker movement. At the same time, it gives the speaker greater freedom of movement by not forcing them to remain in a fixed position facing a recording system. Second, this mask is disposable; it is an individual mask which can easily be used in a clinical setting. Third, oral and nasal airflow can be separately recorded thanks to two partitioned compartments. In each compartment, a probe tube of about 30 cm is attached to the mask on one end and connected to a differential air pressure sensor on the other. The air pressure inside the mask is measured relatively to the atmospheric air pressure through the sensor.

This design allows for high-quality recordings of aerodynamic data, while ensuring acoustic transparency. The aerodynamic performance of the fiber mask has been the object of several studies [16, 17, 18, 19]. For the purpose of the present study, for oral vowels, the fiber mask records overall the same amount of oral and nasal airflow as other devices with good performances (the EVA system and the Rothenberg mask). As for the acoustic quality, our previous study compared formant values and spectral forms of oral vowels among recordings with different types of masks [20]. Compared with data recorded with the EVA2™ mask and the Rothenberg mask, there is little damping of the mid- and high-frequency zones in the data recorded with the fiber mask. The F1/F2 vowel triangle generated based on the acoustic data recorded simultaneously with a fiber mask is minimally distorted, if at all, as compared to the data recorded with the other two types of masks.

1.3. Goal of the present study

The goal of the present study is to further assess the acoustic quality of the data recorded with the adapted fiber mask by providing a perceptual evaluation of oral vowels. Indeed, the intelligibility of speech requires more information than formant values or spectral compositions. For a better evaluation of this technique, we conducted a perception experiment to investigate whether and how the intelligibility and the perception of the naturalness of vowels are affected by the fiber mask technique, in comparison with the EVA2™ mask and the Rothenberg mask.

2. Experiment

This experiment belongs to a larger project of investigating nasality using novel techniques with different populations, approved by the ethics committee of Université de Paris.

2.1. Method

2.1.1. Stimuli

Stimuli were recorded at a soundproof room with two female speakers aged 26 and 49, who are native speakers of French, under four recording conditions: (1) Acoustic-only condition: with an AKG C520L microphone through the soundcard Protools (16 bits/44.1 kHz); (2) Fiber-mask condition (Fig. 1 left): acoustic recording with a Primo EMU-4520 microphone combined with aerodynamic data collected using an adapted fiber mask (16 bits/20 kHz); (3) EVA2 condition (Fig. 1 middle): acoustic recording with an AKG C520L microphone combined with aerodynamic data collected using the oral mask of the EVA2™ station (16 bits/25 kHz); (4) Rothenberg condition (Fig. 1 right): acoustic recording with a Primo EMU-4520 microphone combined with aerodynamic data collected using one of the first versions of the Rothenberg mask [11] (16 bits/20 kHz). The physiological inputs in the conditions (1)–(3) were simultaneously recorded with the acoustic signal through a multi-channel sound acquisition card (DT9800-EC-I, Data Translation®). All data were resampled to 20 kHz.

For the fiber-mask condition, a calibration of the pressure sensor module must be operated for each mask for each speaker, converting airflow values in the physical unit (liters/s). To perform the calibration, we attached the mask to a calibration plate, which was linked into a syringe of 1 liter, and recorded several "bumps" (peak of maximum or minimum airflow) with the syringe. Each bump corresponded to 1L of positive airflow (positive bump) or 1L of negative airflow (negative bump). We then extracted the air volume in Voltage and computed two scaling coefficients (one for each oral or nasal compartment). The two coefficients were then used for deriving the measured airflow values having the physical unit, liters/s.

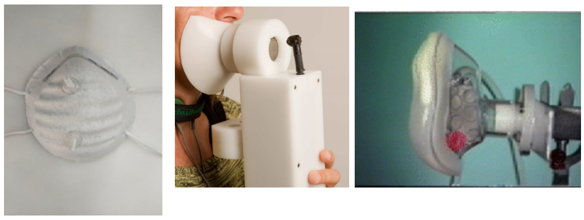

Figure 1: *The three masks used in the three recording conditions. From left to right: fiber mask, EVA2™, Rothenberg.*

Seven oral vowels of French /i, e, y, ø, u, o, a/ were recorded in two contexts. The first context was a sustained vowel in isolation. The second context was a syllable /p_p/ embedding the vowel, extracted from a carrier sentence "Il a dit V comme dans pV. pVp, **pVp**, pVp. Tu n'as pas dit pVpVp mais pVpVpVp". The syllable in bold was extracted from the sentence. Each utterance was repeated a few times, and the first two repetitions were used as stimuli.

Our perceptual evaluation consisted of investigating (1) how listeners' perception of each vowel was affected by the interaction between recording conditions (hereafter, **ACOUSTIC**, **FIBER**, **EVA**, **ROTH**) and context (hereafter isolated (or **V**), embedded (or **pVp**)) and (2) whether listeners could perceive the presence of a mask in these conditions.

2.1.2. Participants

Twenty native French-speaking listeners (mean age = 27.8; SD = 5.3; 16 women, 4 men) participated voluntarily in the experiment. We discarded the data from one listener due to a high error rate of 43% in the ACOUSTIC condition whereas the other listeners had a near-ceiling performance in this condition. Three listeners reported having a history of tinnitus or dyslexia, but their data were retained because they did not present atypical behaviours from the general pattern. No other listeners reported any history of speech, hearing, or motor-related disorders.

Table 1: *Confusion matrices for the categorisation of vowels by recording condition. Upper panel for isolated vowels, lower panel for vowels embedded in /p_p/. Rows for reference vowels, columns for listeners' responses. Bottom rows for accuracies with 95% CIs.*

	V: ACOUSTIC						
	i	y	u	e	ø	o	a
i	76	0	0	0	0	0	0
y	0	76	0	0	0	0	0
u	0	1	75	0	0	0	0
e	0	0	0	76	0	0	0
ø	0	0	0	1	75	0	0
o	0	0	0	0	0	76	0
a	0	0	0	0	0	0	76
	Accuracy: 0.996 (0.99, 1)						

	V: FIBER						
	i	y	u	e	ø	o	a
i	75	1	0	0	0	0	0
y	0	76	0	0	0	0	0
u	0	1	70	0	1	4	0
e	0	0	0	76	0	0	0
ø	0	0	0	2	74	0	0
o	0	0	0	0	1	75	0
a	0	0	0	0	0	0	76
	Accuracy: 0.981 (0.97, 0.99)						

	V: EVA						
	i	y	u	e	ø	o	a
i	57	9	7	0	3	0	0
y	0	70	2	0	3	0	0
u	0	2	74	0	0	0	0
e	0	2	2	61	8	3	0
ø	0	0	0	1	73	2	0
o	0	0	2	0	3	71	0
a	0	0	0	3	5	5	63
	Accuracy: 0.883 (0.85, 0.91)						

	V: ROTH						
	i	y	u	e	ø	o	a
i	60	0	12	0	2	2	0
y	12	31	29	0	3	1	0
u	0	0	58	0	0	18	0
e	4	0	4	17	11	40	0
ø	0	0	0	1	3	72	0
o	0	0	3	0	0	73	0
a	0	0	0	0	5	1	69
	Accuracy: 0.586 (0.54, 0.63)						

	pVp: ACOUSTIC						
	i	y	u	e	ø	o	a
i	76	0	0	0	0	0	0
y	0	76	0	0	0	0	0
u	0	1	75	0	0	0	0
e	7	0	0	69	0	0	0
ø	0	1	1	0	74	0	0
o	0	0	5	0	0	71	0
a	0	0	0	0	1	0	75
	Accuracy: 0.970 (0.95, 0.98)						

	pVp: FIBER						
	i	y	u	e	ø	o	a
i	76	0	0	0	0	0	0
y	0	76	0	0	0	0	0
u	0	1	74	0	0	1	0
e	7	1	0	68	0	0	0
ø	0	2	0	0	74	0	0
o	0	0	5	0	0	71	0
a	0	0	0	0	0	0	76
	Accuracy: 0.968 (0.95, 0.98)						

	pVp: EVA						
	i	y	u	e	ø	o	a
i	76	0	0	0	0	0	0
y	0	76	0	0	0	0	0
u	0	0	73	0	1	2	0
e	9	2	0	64	1	0	0
ø	0	1	1	0	73	0	0
o	0	0	23	0	1	51	1
a	0	0	0	3	11	5	57
	Accuracy: 0.885 (0.85, 0.91)						

	pVp: ROTH						
	i	y	u	e	ø	o	a
i	76	0	0	0	0	0	0
y	9	65	2	0	0	0	0
u	0	3	69	0	0	4	0
e	12	1	0	35	5	0	22
ø	1	3	16	10	17	28	1
o	0	0	12	0	2	60	2
a	0	0	0	0	3	0	73
	Accuracy: 0.744 (0.70, 0.78)						

2.1.3. *Procedure*

Data collection was performed online via PsyToolkit [21, 22]. Participants were instructed to use earbuds or headphones to complete the experiment in a quiet room. As detailed below, each trial consisted of (1) vowel identification; (2) mask-wearing judgement.

At the trial onset, the listener was invited to click on a PLAY button to hear an auditory stimulus twice in a row. The onset of the two repetitions were set apart by 1.5 seconds. At the offset of the second repetition, the listener was invited to choose the vowel they heard among the seven vowels orthographically represented on the screen, by clicking on the corresponding button. Because some graphemes correspond to more than one vowel phoneme, each grapheme was placed above a lexical item that contained the appropriate vowel in the following order: "i (lit), é (fée), ou (toux), u (bus), o (sot), eu (peu), a (chat)", for /i, e, u, y, o, ø, a/, respectively.

Once the listener performed the choice of the vowel, a question appeared on the screen "Is this person wearing a mask?". The listener needed to click on the "yes" or "no" button, before proceeding to the following trial.

Stimuli were randomised in a different order for each participant. Each trial was repeated twice, giving 224 trials in total (7 vowels $\times$ 2 contexts $\times$ 4 recording conditions $\times$ 2 talker voices $\times$ 2 repetitions). The test session was preceded by a practice session with 6 trials using stimuli that were very easy to identify: recordings of isolated /i, a, u/ from the two talkers in the ACOUSTIC condition. Participants mostly spent less than 30 minutes on the experiment.

2.2. Results

2.2.1. *Vowel categorisation*

Table 1 shows the confusion matrices for the categorisation of vowels by recording condition as well as the accuracies of categorisation. If we compare accuracies among the four recording conditions, ACOUSTIC and FIBER give fairly comparable results, which are 10% higher than for EVA, and much higher than for ROTH.

How the categorisation of vowels is perturbed differs according to both the recording condition and the context. There are a few noticeable differences between the isolated and embedded contexts. The first concerns two mid-high vowels /e, o/. While they are almost 100% correctly categorised in the isolated context, they can be miscategorised as high vowels /i, u/ in the embedded syllable /p_p/. This is likely due to phonotactic constraints, /e/ being inhibited in closed syllables, and /o/ dispreferred in this context [23] especially when the first and last consonants are stops. Regardless of this phonotactic effect, the recording condition still plays a role, giving a higher balanced accuracy for ACOUSTIC (/e/: 0.95; /o/: 0.97) and FIBER (/e/: 0.94; /o/: 0.96) than EVA (/e/: 0.91; /o/: 0.82) and ROTH (/e/: 0.71; /o/: 0.85) conditions.

The second difference concerns the high front vowels /i, y/. In isolated forms and the EVA and ROTH conditions, these vowels are confused with each other, perceived as further back and/or lowered. This is likely due to the acoustic distortions caused by the two recording conditions. The correctness of categorising /i, y/ is, however, increased in the embedded context, despite similar acoustic distortions. For ROTH only, the embedded context increases the categorisation correctness as compared to the isolated context for all the vowels except /o/, with an overall increase of 15%. In particular, non-focal vowels /e, ø/ are the most incorrectly and variably categorised in the ROTH condition, and their correctness is improved in the embedded context. For the other recording conditions, the accuracy rate is overall 1-3% lower for embedded than isolated vowels.

The miscategorisation of the vowel /a/ as a non-high and non-front vowel occurs for the EVA and ROTH conditions. For ROTH and the embedded context only, other non-high and non-front vowels can also be miscategorised as /a/.

Overall, the FIBER condition gives a very close accuracy rate to the ACOUSTIC condition. The only exception is the isolated /u/, which is miscategorised as /o/ 4 times out of 76 in the FIBER but not the ACOUSTIC condition. However, given that this miscategorisation only occurs with one talker's stimulus, it could be a token- or speaker-specific case.

2.2.2. *Mask-wearing judgement*

A generalised linear mixed model (GLMM), using the `lmerTest` package [24] in R [25], was fitted to listeners' judgement response as to whether the talker was wearing a mask. The Helmert-coded predictors VOWEL, CONTEXT, RECORDING condition, and the CONTEXT $\times$ RECORDING in-

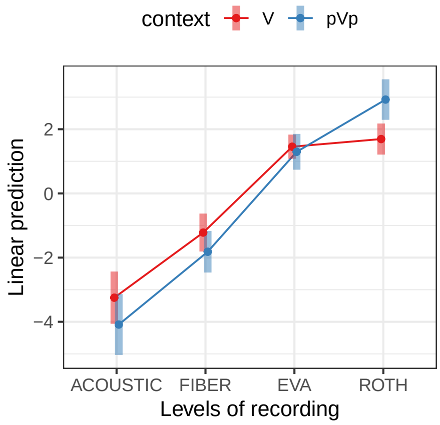

Figure 2: *Interaction plot of mask-wearing judgement by recording condition and context. Higher values for higher likelihood of perceiving the voice as masked. Bars: 95% CIs.*

teraction were included. Random intercepts were included for listeners and talker voices, and by-participant random slopes for CONTEXT and RECORDING. Interaction plots and pairwise comparisons were made using the `emmeans` package [26].

As shown in Fig. 2, the likelihood of the presence of a mask increases as follows: ACOUSTIC < FIBER < EVA < ROTH. Isolated vowels (V) are more likely to be judged as produced by someone with a mask on than embedded vowels (pVp) for the ACOUSTIC ($est. = 0.84$, SE $= 0.36$, $z = 2.33$, $p < .05$) and FIBER ($est. = 0.60$, SE $= 0.24$, $z = 2.45$, $p < .05$) conditions, and less likely so for the ROTH condition ($est. = 1.23$, SE $= 0.28$, $z = 4.34$, $p < .0001$).

3. Discussion

Our study provides a perceptual validation of the data recorded with the fiber mask technique, as compared with the EVA™ mask and the Rothenberg mask. Stimuli recorded with the fiber mask gave a high accuracy rate of vowel categorisation, which was similar to the categorisation rate of the acoustic-only recorded data. The rare exception concerned the vowel /u/, which could be due to a talker- or token-specific effect. Compared to isolated vowels, embedded vowels overall decreased the accuracy rate for mid vowels /e, o/, possibly due to phonotactic constraints. Nonetheless, their accuracy rates were similar between acoustic-only recorded data and fiber-mask recorded data. We can thus confidently conclude that the fiber mask has little impact on the intelligibility of oral vowels.

This can be seen as a great improvement compared with rigid masks. The lowering of the formants and/or the spectral attenuation at mid- and/or high-frequency zones led to the miscategorisation of high and front vowels as further back and/or lower. Non-focal vowels /e, ø/ had the highest miscategorisation rate with the Rothenberg mask recorded stimuli, possibly because spectral attenuations affected them to a larger degree for a correct perception than focal vowels, which are perceptually more salient [27].

The accuracy rate for at least some vowels was increased in the embedded context compared to the isolated context, despite similar acoustic distortions as observed in [20]. The complex interaction between perceptual assimilation and normalisation for the contextual information [28] may explain this behaviour.

It should, however, be noted that the stimuli recorded with the fiber mask had a higher score of mask-wearing judgement than acoustic-only recordings. This means that some forms of spectral attenuation or other characteristics in the acoustic signal may give the impression of a masked voice. Nonetheless, stimuli recorded with the EVA2™ mask and the Rothenberg mask were much more likely to be judged as produced with a mask.

Overall, the fiber mask presents non-negligible advantages for the analysis of the acoustic data recorded in this condition. This is valuable when the simultaneous collection of acoustic and aerodynamic data is required for a better-controlled analysis. It is also helpful when the data collection needs to be limited in time in a clinical setting, or simply to avoid participants' fatigue, etc.

However, our present study is restricted to oral vowels. Future research needs to address the impact of aerodynamic devices on other types of sounds, especially fricatives and nasal consonants and vowels, which require fine observations of the aerodynamic mechanisms.

4. Acknowledgements

The present study was funded by the innovative project (University Sorbonne Nouvelle) led by Amélie Elmerich and by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

5. References

[1] B. Teston, "L'oeuvre d'Etienne-Jules Marey et sa contribution à l'émergence de la phonétique dans les sciences du langage," *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, vol. 23, pp. 237–266, 2004, publisher: Laboratoire Parole et Langage.

[2] E.-J. Marey, *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine / par É.-J. Marey,...*, 1878.

[3] J. L. Pech, "La notion du débit respiratoire maximum; mesure pratique de ce débit au moyen du masque manométrique (The effect of the maximum respiratory flow; method of measuring the flow by means of a manometric mask)," *Presse med*, vol. 29, p. 93, 1921.

[4] A. Fleisch, "Der pneumotachograph; ein apparat zur geschwindigkeitsregistrierung der atemluft," *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, vol. 209, pp. 713–722, 1925.

[5] H. R. Gilbert, "Oral airflow during stop consonant production," *Folia Phoniatrica et Logopaedica*, vol. 25, no. 4, pp. 288–301, 1973.

[6] J. F. Lubker and K. L. Moll, "Simultaneous oral-nasal air flow measurements and cinefluorographic observations during speech production," *The Cleft Palate Journal*, vol. 36, pp. 257–272, 1965.

[7] M. Rothenberg, "Measurement of Airflow in Speech," *Journal of Speech and Hearing Research*, vol. 20, no. 1, pp. 155–176, 1977.

[8] B. Teston, "Un système de mesure des paramètres aérodynamiques de la parole: le Polyphonomètre modèle III," p. 9, 1984.

[9] B. Teston and B. Galindo, "A Diagnostic and Rehabilitation Aid Workstation for Speech and Voice Pathologies," *EUROSPEECH-1995*, pp. 1883–1886, 1995.

[10] C. C. McLean, S. W. Kelly, and M. C. Graham Manley, "An instrument for the non-invasive objective assessment of velar function during speech," *Medical Engineering & Physics*, vol. 19, no. 1, pp. 7–14, 1997.

[11] M. Rothenberg, “A new inverse-filtering technique for deriving the glottal air flow waveform during voicing,” *The Journal of the Acoustical Society of America*, vol. 53, no. 6, pp. 1632–1645, 1973.

[12] S. Hertegård and J. Gauffin, “Acoustic properties of the Rothenberg mask,” p. 12, 1992.

[13] P. Badin, S. Hertegård, and I. Karlsson, “Notes on the Rothenberg mask,” p. 9, 1990.

[14] A. Ghio and B. Teston, “Caractéristiques de la dynamique d'un Pneumotachographe pour l'étude de la production de la parole : aspects acoustiques et aérodynamique.” in *Journées d'Etude sur la Parole*, Nancy, France, 2002, pp. 337–340.

[15] K. Honda and S. Maeda, “Glottal-opening and airflow pattern during production of voiceless fricatives: a new non-invasive instrumentation,” *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3738–3738, 2008.

[16] M. Yeou, K. Honda, and S. Maeda, “A palatographic, photoglottographic and airflow study of some coronal geminates in Figuig Berber,” in *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 2011, pp. 2216–2219.

[17] T. Kamiyama, B. Kühnert, and J. Vaissière, “Do French-speaking learners simply omit the English /h/?” in *The 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 2011, pp. 1010–1013.

[18] H. Kim, S. Maeda, K. Honda, and L. Crevier-Buchman, “The mechanism and representation of Korean three-way phonation contrast: External photoglottography, intra-oral air pressure, airflow, and acoustic data,” *Phonetica*, vol. 75, pp. 57–84, 2018.

[19] Y. Chi, K. Honda, and J. Wei, “Glottographic and aerodynamic analysis on consonant aspiration and onset F0 in Mandarin Chinese,” in *ICASSP 2019-2019*, 2019, pp. 6480–6484.

[20] A. Elmerich, A. Amelot, S. Maeda, Y. Laprie, J. F. Papon, and L. Crevier-Buchman, “F1 and F2 measurements for French oral vowel with a new pneumotachograph mask,” in *ISSP 2020-12th International Seminar on Speech Production*, 2020.

[21] G. Stoet, “PsyToolkit: A software package for programming psychological experiments using Linux,” *Behavior Research Methods*, vol. 42, no. 4, pp. 1096–1104, 2010.

[22] ——, “PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments,” *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[23] P. Delattre, “Rapports entre la durée vocalique, le timbre et la structure syllabique en français,” *French Review*, pp. 547–552, 1959.

[24] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest package: Tests in linear mixed effects models,” *Journal of Statistical Software*, vol. 82, no. 13, 2017.

[25] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2022.

[26] R. Lenth, “emmeans: Estimated marginal means, aka Least-squares means,” 2022.

[27] J.-L. Schwartz, C. Abry, L.-J. Boë, L. Ménard, and N. Vallée, “Asymmetries in vowel perception, in the context of the dispersion–focalisation theory,” *Speech Communication*, vol. 45, no. 4, pp. 425–434, 2005.

[28] A. Rysling, A. Jesse, and J. Kingston, “Regressive spectral assimilation bias in speech perception,” *Attention, Perception, & Psychophysics*, vol. 81, pp. 1127–1146, 2019.