



HAL
open science

Speech aerodynamics database

Didier Demolin, Sergio Hassid, Clara Ponchard, Shi Yu, Roland Trouville

► **To cite this version:**

Didier Demolin, Sergio Hassid, Clara Ponchard, Shi Yu, Roland Trouville. Speech aerodynamics database. 2019. hal-02503114

HAL Id: hal-02503114

<https://hal.science/hal-02503114v1>

Submitted on 9 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speech Aerodynamics Database

Didier Demolin¹, Sergio Hassid², Clara Ponchard¹, Shi Yu¹, Roland Trouville¹

¹Laboratoire de Phonétique et Phonologie, CNRS-UMR 7018, Université Sorbonne Nouvelle Paris 3,
19 Rue des Bernardins, 75005, Paris, France

²Hôpital Erasme, Université Libre de Bruxelles, Route de Lennik 808, 1070 Bruxelles, Belgium
{didier.demolin, clara.ponchard, shi.yu}@sorbonne-nouvelle.fr
sergio.hassid@ulb.ac.be, roland.trouville@gmail.com

Abstract

Acoustic signal of speech sounds is underlain by aerodynamic principles. The Aerodynamics of speech contributes to the mechanistic explanations of speech production. This database was designed during an ARC project "Dynamique des systèmes phonologiques" in which the study of aerodynamic constraints on speech was an important target. Data were recorded between 1996 and 1999 at the Erasmus Hospital (Hôpital Erasme) of Université Libre de Bruxelles. The goal was to obtain a substantial amount of data with simultaneous recording, in various context, of the speech acoustic signal, subglottal pressure (P_s), intraoral pressure (P_o), oral airflow (Q_o) and nasal airflow (Q_n). This database contains recordings of 2 English, 1 Amharic, and 7 French speakers and is provided with data conversion and visualisation tools. Another aim of this project was to obtain some reference values of the aerodynamics of speech production for female and male speakers uttering different types of segments and sentences in French.

Keywords: Speech Production, Aerodynamics, Database

1. Introduction

The aerodynamic phase of speech is important since it is at this stage that speech sounds are generated. Empirical data on aerodynamic parameters of speech production remain rare. Aerodynamic data are important for the understanding of sound patterns in languages. To consider fundamental questions in phonetics and to process problems in clinical phonetics, it is crucial to understand the underlying aerodynamic process of sound production. There are 4 parameters involved in speech production: sub-glottal and intra-oral pressures plus oral and nasal airflows. There are some difficulties to gather these parameters simultaneously. They consist in internal factors of aerodynamic processes and in external factors requiring clinical and medical assistance. This database is one of the few aerodynamic databases of speech production available today to the scientific community. This database thus offers an opportunity for speech scientists to advance aerodynamic analyses of speech production.

The database was recorded to study the production of French consonants in similar vocalic environments. The production of vowels with controlled fundamental frequency was also undertaken. A set of French and English sentences was also recorded to study various intonation patterns and stress occurrences. Finally the production Amharic consonants was also done. The primary intention of this database was to obtain some references values for the phenomena under study by recording several subjects for the same protocol when it was possible.

2. Recording Methods

All data were collected by simultaneous and synchronised recordings of 5 types of signal:

- Subglottal Pressure P_s (unit: hPa)
- Intraoral Pressure P_o (unit: hPa)

- Oral Airflow Q_o (unit: dm^3/s)
- Nasal Airflow Q_n (unit: dm^3/s)
- Speech Acoustic Signal

Two additional mesures: intensity and f_o were computed based on speech acoustic signal. The same recording procedure was applied for the three languages (English, French, Amharic) in our database, Different protocols have been applied. Therefore there are differences in the measurements of recordings between subjects and languages. The procedure preserved the rights and welfare of human research subjects, in respect of the ethical committee's rules (<https://www.erasme.ulb.ac.be/fr/ethique>). Recording procedure for each mesure is detailed in the following subsections, the definition of each measure will also be provided.

2.1. Subglottal Pressure (P_s)

Subglottal pressure (P_s) is the pressure below the vocal folds. In this database, P_s was measured by direct tracheal puncture with a 2mm diameter needle inserted in the last ring of the trachea. The needle was placed after local anesthesia with 2% Xylocaine, including the subglottal mucosa. The tip of the needle was inserted right under the cricoid cartilage. A plastic tube of 2mm of diameter linked to a pressure transducer was connected to the needle.

More precisely, P_s represents "the energy immediately available for the creation of acoustic speech signals" (Baken and Orlikoff, 2000). A more or less constant P_s must be maintained during speech production because "inappropriate levels of subglottal pressure or inadequate pressure regulation can cause abnormal levels of speech intensity or sudden changes in the fundamental frequency" (Baken and Orlikoff, 2000). A sufficient amount of air pressure beneath the vocal folds is necessary to initiate phonation. In speech P_s must be about 2 hPa above

atmospheric pressure to make the vocal folds vibrating.

The measurement of subglottal pressure needs to be made accurately both for research and clinical purposes. There are several ways to obtain P_s but the most efficient albeit the most difficult for obvious reasons is by tracheal puncture (Baken and Orlikoff, 2000).

2.2. Intra-oral Pressure (P_o)

Intra-oral pressure (P_o) is the air pressure that is exerted inside the oral cavity. This measure is obtained in our experiment with a small flexible plastic tube of diameter 2mm inserted through the nasal cavity to the oropharynx.

P_o is an extremely robust parameter for the understanding of sound patterns in language and particularly to infer movements from the speech articulators. The measure is made by inserting a small flexible plastic tube through the nasal cavity into the oropharynx. The advantage of this measurement is that there is no interference with the movement of the articulators, or a risk that tongue or lip movements will move the tube.

2.3. Oral Airflow (Q_o)

Oral airflow was measured with a flexible silicone rubber mask covering the mouth. Both the oral and nasal airflow were sampled at 2 kHz (12 bits). The maximum level was fixed at 500 cm^3/s , except for the nasal airflow of female speakers, which was adjusted to 200 cm^3/s . The zero level was adjusted by the experimenter at the beginning of each recording session and checked regularly during the session.

2.4. Nasal Airflow (Q_n)

Nasal airflow was measured through an olive inserted in one or two nostrils. The olive was connected to a 10 cm plastic tube of 0.5 cm internal diameter.

2.5. Acoustic Signal

An AKG C 419 microphone is positioned just behind the oral transducer. The speech signal was sampled at 16 kHz (12 bits). From the acoustic signal, the intensity is computed by means of the root mean square (RMS) method. The pitch is computed by Praat and YAAPT (Kasi, 2002) method.

2.6. Data acquisition

All measurements were connected to the Physiologia workstation and recorded using this workstation (Teston and Galindo, 1990). The experiments took place at the Erasmus Hospital of Université Libre de Bruxelles, Belgium. Physiologia is a multisensor data acquisition system allowing simultaneous recording of the speech signal and various aerodynamic measures, it consists in a PC computer and an acquisition system equipped with various transducers. All signals were processed with software Phonedit (Ghio, 2002).

3. Database Description

This section provides a detailed description of linguistic materials of the Speech Aerodynamics Database. This database contains recordings of utterances of three languages

- English (2 male speakers)
- French (5 male and 2 female speakers)
- Amharic (1 male speakers)

3.1. English

Recordings are collected on March 22, 1999 for subject 1 (labeled *en_M1*) and November 18, 1999 for subject 2 (labeled *en_M2*). *en_M1* is a native British speaker aged 74 years old by the time of data collection, *en_M2* is a native American speaker aged 35 years old. The linguistic material of this English corpus was designed by two phonetic experts and will be detailed below. The English part of this database is previously studied by Yu (2019) and Yu et al. (2019)

3.1.1. Logatome

This group consists of two-syllable logatome, produced with three types of carrier phrase:

1. "say again."
2. "... say ... again."
3. "say" (repetition of logatome until the end of breath)

The form of these logatomes are ($/CVCV/$ or $/?V?V/$), they are constructed by combining three vowels [i], [a], [u] and stop consonants [p], [t], [k], [b], [d], [g].

3.1.2. Variation of vocalic and consonantal features

This group is constituted by 15 recordings of subject *en_M2*. Words in this group mostly consist of $/CVC/$ syllable structure. Variations of features are:

• Vowels

1. Vowel height: [a], [æ], [ʌ], [ɔ], [ɛ], [e], [ɪ], [u], [i].

• Consonants

1. oral vs. nasal
2. voiced vs. unvoiced stops, fricatives, and affricates
 - Stops: [p], [b], [t], [d]
 - Fricatives: [f] [v], [s], [z], [θ], [ð]
 - Affricates: [tʃ], [dʒ]
3. Sonorants: [j], [w], [r], [l]

• Others

1. consonant clusters
 $/\#sp_/, / \#sl_/, / \#sm_/, / \#st\#/, / \#zd\#/, / \#nt\#/, / \#nd\#/, / \#ps\#/, / \#ls\#/, / \#ns\#$
2. two-syllable words
 - nasals + (un)voiced stop
 $/\#_mp_./, / \#_mb_./, / \#_nt_./, / \#_nd_./,$
 for example /mp/ in "wimple".
 - intervocalic pretonic consonants
 $[CVC_0VC]$,
 with $C_0 : /p/, /b/, /t/, /d/, /s/, /z/, /m/,$
 for example: /p/ in "repel".

A summary of this group can be found in Table 1.

	Stops	Fricatives	Affricates	Sonorants	Clusters
initial	peel, bean [i] pin, bin [ɪ] pain, bane [e] pan, ban [æ]	fear, veer [ɪ] fin, vim [ɪ] feign, vain [e] fan, van [æ]		wean [i], yule [u] win [ɪ], year [ɪ], wane [e], yell [ɛ] wan [a], yawl [ɔ]	spall [sp] slam [sl] small [sm]
initial	team, deem [i] tame, dame [æ] tam, dam [æ]	seem, zeem [i] same, zane [e] sam, zam [æ]	cheer, jeer [ɪ] chain, jane [e] chan, jam [æ]	real, lean [i] rain, lame [e] ram, lam [æ]	
initial		thin [θɪ], dheen [ðɪ] thain [θe], then [ðe] thumb [θʌ], dhawn [ðɔ]		mean [i], kneel [i] mint [ɪ] main [e], name [e] man [æ], nom [a]	
final	rape [ep] rabe [eb] late [et] laid [ed]	lafe [ef], laugh [æf], rave [ev], lave [æv], shelf [f], shelve [v] lace [es] raise [ez]	rache [etʃ] rage [eʒ]		rest [st], raised [zd] rent [nt], rend [nd] lance [ns], lapse [ps], false [ls]
intervocalic	ripple [p] ribbon [b]	riffle [f] riven [v] whistle [s] wizen [z]		rimmle [m] whittle [r], riddle [r]	
pretonic	repel [p], rebel [b] entire [t], endure [d] retire [t], reduce [d]	resource [s] resign [z]		wimple [mp], wimble [mb] wintle [nt], windle [nd] remain [m]	

Table 1: List of words in English material. Phonetic transcription indicates the features being varied.

3.1.3. Word Length and Stress Position

This group is conceived based on 4 monosyllabic words. Word length is varied adding suffixes successively. The variation of stress position is carried by lexical information.

- *pit*, *pity*, *pitying*, *pityingly*
- *play*, *playful*, *playfully*
- *photo*, *photographer*, *photographic*
- *man*, *manage*, *manager*, *managerial*, *jeer*

3.1.4. Modality

This group of utterances is produced exclusively by *en_M1*. The sentence "*Jenny's pie is pretty*" is produced by varying the modality (statement vs. question) and by varying the stress position to reflect the informational structure.

3.2. French

3.2.1. Logatome

Acoustic data were collected from native French speakers: two women (labelled *fr_F1* and *fr_F2*) and three men (labelled *fr_M1*, *fr_M2* and *fr_M8*) with normal larynx and no voice problems.

The subjects were recorded pronouncing a series of logatomes combining each of the French consonants with the vowels /a, i, u/. These logatomes were inserted in the sentence "[C₁VC₂V] say [C₁VC₂V] again" (for example "*papa dis papa encore*"), to be repeated five consecutive times by the subjects for each of the consonants in the three intervocalic contexts.

These data were analysed in our previous study Signorello et al. (2017) for fricative consonants French and by Ponchard (2019) for stop consonants.

For fricative consonants, the goal of the study was:

1. To predict the starting, central, and releasing points of friction based on the measurements of P_s , P_o , and Q_o ;
2. To compare voiceless and voiced fricatives and their places of articulation;
3. To provide reference values for the aerodynamic features of fricatives for further linguistic, clinical, physical and computational modelling research.

For stop consonants, the goals of the study was:

C'est une chanson triste.	C'est une chanson qui m'attriste.
C'est une maison grise.	C'est une maison qui me grise.
La démonstration du président de l'assemblée nationale m'a convaincu de la gravité de la situation.	La démonstration du président m'a semblé convaincante.
Cette nouvelle théorie linguistique provoque si j'ai bien compris, une nouvelle polémique.	Tu aurais tout intérêt si tu tiens à conserver son amitié à lui téléphoner plus souvent.
Tu vois cette maison ? C'est la maison que j'aimerais visiter.	Ce n'est pas le château, c'est la maison que j'aimerais visiter.
Mais non tu n'as rien compris, c'est la maison que j'aimerais visiter, pas le château.	Elle est complètement débile cette histoire de passeport volé.
Cette histoire de passeport volé elle est complètement débile	Je n'aime pas les films violents que l'on passe à la télé le samedi après-midi et toi?
Anne-Marie dit lui de venir tout de suite.	Dis-lui de venir tout de suite Anne-Marie.
Elle n'est pas venue Anne-Marie.	Selon moi, elle prendra le train ou elle ne viendra pas.
Elle viendra ou elle ne viendra pas ?	Il a une version écrite ou une version orale de ce document ?

Table 2: French sentences

1. To automate the processing of aerodynamic data;
2. To analyse pressure variations according to voicing, intervocalic context and places of articulation;
3. To find relevant descriptors for the automatic classification of French stop consonants.

3.2.2. Vowel

Acoustic data were collected from two French speaking subjects, a male subject (labelled *fr_M2*) and a female subject (labelled *fr_F2*). The corpus consists of synchronised measurements of Q_o and P_o . The task was to pronounce sustained vowels [a], [e], [i], [o], [u], and [ɔ̃] while hearing a tone through headphones connected to a synthesiser, the sound level of the tone defining the intensity level. Three tones were selected, A-C-E (la, do, mi) for the male subject, and C-F-A (do, fa, la) for the female subject, which represents 9 types of vowel repetitions per speaker. For each tone three different sound levels were selected. This method allowed gathering data within a narrow pitch band and covering a wide range of intensities.

These data were processed by Bucella et al. (2000) to examine variations from different vowels and see if this respiratory effort may be used to distinguish different sounds. Results showed that:

1. P_s is lower for [a] than for [u] and for [i]
2. There is a vowel effect, as well for the P_o and for the Q_o , by repeated measures of analysis of variance.

These data were also used to study the relationship between intensity and P_s with controlled f_o (Lecuit and Demolin, 1998). The results show that the relationship varies when one considers the vowels or the pitch parameter.

Oral and nasal vowels effects on P_s were examined by Demolin et al. (2017). This work investigates the effect of oral and nasal vowels on P_s . Measurements show that both speakers produced each of the vowels with stable f_o (2 Hz of difference variation from the given tone). One interesting observation is that there is a substantial difference in P_s between oral and nasal vowels. Both speakers produced nasal vowels with a lower P_s when compared to oral

vowels. Mean differences between both set of vowels were quantified at 2.15 hPa. Nasal vowels were found to have lower intensity than oral vowels.

3.2.3. Sentences

Acoustic data were collected from three French speaking subjects, a female subject (labelled *fr_F3*) and two male subject (labelled *fr_M6* and *fr_M7*). Detailed linguistic materials can be found in Table 2.

3.3. Amharic

The Amharic data were recorded with one male native speaker. One of the goals was to obtain aerodynamic parameters for singleton and geminated consonants. Those consisted in a set of modal and glottalized consonants. In addition various consonantal contrasts were recorded: labialized and non-labialized consonants; glottalized and non-glottalized and labialized and ejectives. A detailed the list can be found on the website of this database.

4. Database Website

A website is developed to host the database, it is developed at the Laboratory of Phonetics and Phonology of the Sorbonne Nouvelle University, Paris 3, and was designed with the SQL language used to exploit relational databases and the availability site was designed in PHP. All the data presented in this article are accessible and downloadable via the database website <https://corpus.ilpga.fr/aerodynamics>.

In addition to the structure of database, a data conversion package coded in Python and a web visualisation tool were developed. The data conversion package helps to process data with Python. The visualisation tool is based on plotly (Sievert et al., 2016). This tool allows interactive visualisation of the database. Each measure is rendered in a separate layer so that focalised and comparative view of data was made possible. For example, we can keep only the variation of P_s and f_o and hide all other parameters to see their correlation. All data were annotated and segmented, so that transcription made with Praat can also be visualised with this tool.

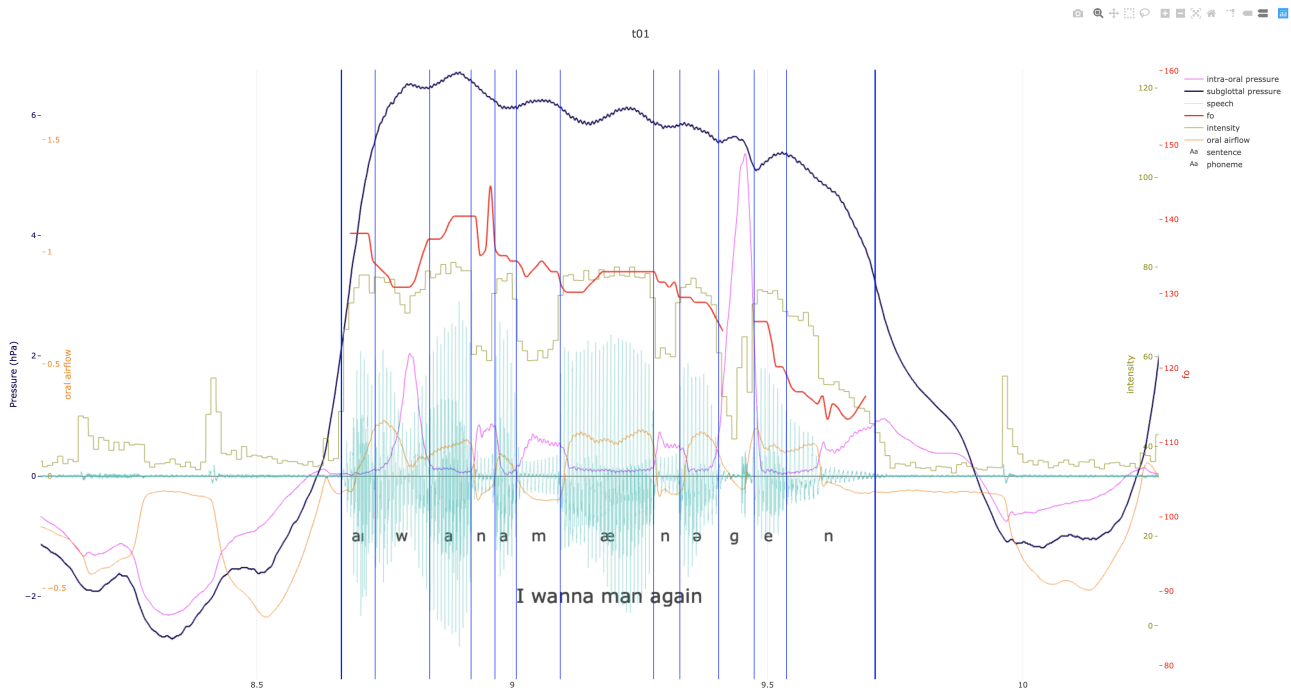


Figure 1: Interface of data visualisation tool

Several functionalities will be made available to enhance this visualisation interface. For example, a selected number of points can be exported to other data formats, this allows us to examine the crucial moments of the variation of these aerodynamic parameters.

5. Conclusion

As part of this study, we contributed to the development of a database for the provision of aerodynamic data to the scientific community. This database is an important contribution because aerodynamic data are extremely difficult to acquire, in particular P_s measurements taken by tracheal puncture, which requires hospital supervision and raise ethical problems because of its invasive nature. This database allows a substantial development of studies in the aerodynamics of speech production and the modelling researches.

6. Acknowledgements

This work is supported by grants ARC 1996-2000 of the Belgian Government and by ArtSpeech (Phonetic Articulatory Synthesis, DS0707–2015) from the French National Research Agency and by the French Investissements d’Avenir - Labex EFL program (ANR-10-LABX-0083).

7. Bibliographical References

- Baken, R. J. and Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.
- Bucella, F., Hassid, S., Beeckmans, R., Soquet, A., and Demolin, D. (2000). Pression sousglottique et débit d’air buccal des voyelles en français. *XXIIIèmes journées d’Etude sur la Parole*, pages 449–452.
- Demolin, D., Trouville, R., Wang, R., and Signorello, R. (2017). Oral and nasal vowels effects on subglottal pressure. *The Journal of the Acoustical Society of America*, 142(4):2582–2582.
- Ghio, A. (2002). Phonedit: Multiparametric speech analysis. *LPL (Aix-en-Provence, France) www.lpl.univ-aix.fr*.
- Kasi, K. (2002). *Yet Another Algorithm for Pitch Tracking:(YAAPT)*. Ph.D. thesis, Old Dominion University.
- Lecuit, V. and Demolin, D. (1998). The relationship between intensity and subglottal pressure with controlled pitch. In *Fifth International Conference on Spoken Language Processing*.
- Ponchard, C. (2019). Classification automatique des lieux d’articulation des consonnes occlusives du français. Master’s thesis, Université Sorbonne Nouvelle.
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., and Despouy, P. (2016). plotly: Create interactive web graphics via plotly’s javascript graphing library [software].
- Signorello, R., Hassid, S., and Demolin, D. (2017). Aerodynamic features of french fricatives. In *INTER-SPEECH*, pages 2267–2271.
- Teston, B. and Galindo, B. (1990). The physiologia system: Description and technical specifications.
- Yu, S., Hassid, S., and Demolin, D. (2019). A phonetic study of subglottal effects on stress and fundamental frequency. In *ICPhS*.
- Yu, S. (2019). La variation de la pression sous-glottique dans la production de la parole et sa contribution à la fréquence fondamentale : une étude aérodynamique et acoustique sur l’anglais. Master’s thesis, Université Sorbonne Nouvelle.