Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 1933–1938
Marseille, 20-25 June 2022

Speech Aerodynamics Database, Tools and Visualisation

Shi Yu[1], Clara Ponchard[1], Roland Trouville[1], Sergio Hassid[2], Didier Demolin[1]
[1]Laboratoire de Phonétique et Phonologie, CNRS-UMR 7018, Université Sorbonne Nouvelle Paris 3,
19 Rue des Bernardins, 75005, Paris, France
[2]Hôpital Erasme, Université Libre de Bruxelles, Route de Lennik 808, 1070 Bruxelles, Belgium
{shi.yu, clara.ponchard, didier.demolin}@sorbonne-nouvelle.fr
sergio.hassid@ulb.ac.be, roland.trouville@gmail.com

Abstract

Aerodynamic processes underlie the characteristics of the acoustic signal of speech sounds. The aerodynamics of speech give insights on acoustic outcome and help explain the mechanisms of speech production. This database was designed during an ARC project "Dynamique des systèmes phonologiques" in which the study of aerodynamic constraints on speech production was an important target. Data were recorded between 1996 and 1999 at the Erasmus Hospital (Hôpital Erasme) of Université Libre de Bruxelles, Belgium and constitute one of the few datasets available on direct measurement of subglottal pressure and other aerodynamic parameters. 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, Subglottal pressure, Database

1. Introduction

The aerodynamic process is an essential phase of speech production since it constitutes the generation of the source of speech acoustic signals. Empirical data on aerodynamic parameters of speech production remain very limited, mainly due to the difficulties of its acquisition. Aerodynamic data are important for the understanding of sound patterns in languages. To consider fundamental questions in phonetics and to resolve problems in clinical phonetics, it is crucial to understand the underlying aerodynamic processes of speech sound production. There are mainly 4 measurable aerodynamic parameters involved in speech production: subglottal/intraoral pressures and oral/nasal airflows. The difficulties of data acquisition increase when these parameters are to be recorded simultaneously with acoustic signals. The difficulties consist of internally the nature of aerodynamic processes and, externally the requirements of clinical and medical assistance. This database is one of the very few databases of the aerodynamics of speech production available today to the scientific community. Therefore it offers a unique opportunity for speech scientists to study the aerodynamics of speech production.

This database was first recorded to study the production of consonants of French in different vocalic contexts. The production of vowels with controlled fundamental frequency was also undertaken. Later a set of French and English sentences was recorded to study various intonation and stress patterns. Finally the production of Amharic language was recorded to study language-specific aerodynamics characteristics and to make comparison of cross-language differences of consonants. The primary goal of this database is to obtain reference values for the aerodynamic parameters of speech production for different segments and, subsequently, to study the aerodynamic process of intonation and stress. Recordings were made on several subjects with 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 measures: intensity and fundamental frequency (f_o) were computed based on speech acoustic signal. The same recording procedure was applied for all three languages (English, French, Amharic) in this database. Due to the environmental constraints and the goal of studies, there are several differences in the number of aerodynamic parameters recorded across subjects and languages. Although the linguistic material may seem to be heterogeneous across languages, the underlying aerodynamic and physiological principles could rule out surface variations of linguistic forms and thus make cross-language comparison possible. 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 all measurements are detailed in the following subsections, the definition of each measurement is provided.

2.1. Subglottal Pressure (P_s)

Subglottal pressure (P_s) is the pressure below the vocal folds which drives their vibrations. In this database, P_s was measured directly by 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.

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 voiced speech sound 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 level of air pressure beneath the vocal folds is necessary to initiate and maintain phonation. In speech production, the P_s must be maintained at least 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 methods to measure P_s indirectly and less-invasively (Draper et al., 1959), but the most precise and efficient albeit the most difficult method for obvious reasons is the 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 measurement 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 consonantal sound patterns in languages and, particularly, it can be used to infer movements of speech articulators. The measurement is made by inserting a small flexible plastic tube through the nasal cavity into the oropharynx. The advantage of this measurement procedure is that it avoids the interference with movement of articulators, and the risk that tongue or lip movements displace 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. The maximum airflow level is fixed at 500 cm^3/s, except for the nasal airflow of female speakers, which is adjusted to 200 cm^3/s. The zero level is adjusted by the experimenter at the beginning of each recording session and is checked regularly during the session.

2.4. Nasal Airflow (Q_n)

Nasal airflow is 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 is sampled at 16 kHz (12 bits). The intensity is computed from the acoustic signal with the root mean square (RMS) method. The fundamental frequency is computed with Praat's autocorrelation method (Boersma, 2001).

2.6. Data acquisition

All measurements are 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. Phygiologia is a multisensor data acquisition system allowing simultaneous recordings of the speech acoustic signal and various aerodynamic measures, it consists of a PC and an acquisition system equipped with various transducers. All signals were processed with Phonedit software (Ghio, 2002) in previous studies of this database.

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 with native speakers:

- 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 74 years old native British speaker and *en_M2* is a 35 years old native American speaker. The linguistic material of this English corpus was designed by two phonetic experts and is 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 of utterance 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 $/ʔVʔV/$), 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 of utterance is constituted by 15 recordings of subject *en_M2*. The 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]. (from low to high)
- **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: [ʧ], [ʤ]
  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 $[CVC_0{}^{\text{'}}VC]$,
       with C_0 :/p/, /b/, /t/, /d/, /s/, /z/, /m/,
       for example: /p/ in "*repel*".

Materials of this group is detailed in Table 1.

3.1.3. Word Length and Stress Position

This group of utterance is conceived based on 4 monosyllabic words. Word length is modulated by adding suffixes successively. The variation of stress position is carried by lexical information such that the aerodynamics process can be studied in an ecological context.

- ***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 emphasis and informational structure.

3.2. French

3.2.1. Logatome

Acoustic and aerodynamic data were collected from two women (labelled *fr_F1* and *fr_F2*) and three men (labelled *fr_M1*, *fr_M2* and *fr_M8*) native French speakers.

The subjects were instructed to produce a series of logatomes combining different consonants with one of the vowels /a, i, u/. These logatomes are carried with the sentence "$[C_1VC_2V]$ dit $[C_1VC_2V]$ encore", for example "*papa dis papa encore*" (papa say papa again). All utterances were repeated five times 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 goals of the study were:

1. To predict the starting, central, and releasing points of frication 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 of the aerodynamic features of fricatives for further linguistic, clinical, physical and computational modelling research.

For stop consonants, the goals of the study were:

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

3.2.2. Vowel

Acoustic data were collected from two native French speakers, one male subject (labelled *fr_M2*) and one female subject (labelled *fr_F2*). The corpus consists of synchronised measurements of Q_o and P_o. The task is to produce sustained vowels [a], [e], [i], [o], [u], and [ɔ̃] while hearing a tone through headphones connected to a synthesiser, subjects are instructed to produce utterance at an intensity level similar to the sound level of the tone they hear. 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

	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 [ði] thain [θe], then [ðɛ] 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 [edʒ]		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 [ɾ], riddle [ɾ]	
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.

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

represents 9 types of vowel repetitions per speaker. For each tone, three different sound levels were selected. This method allows gathering data within a narrow pitch band and covering a wide range of intensities.

These data were studied by Bucella et al. (2000) to examine variations of different vowels and to investigate if the respiratory effort may be used to distinguish different sound patterns. Results showed that:

1. P_s is lower for [a] than for [u] and for [i]

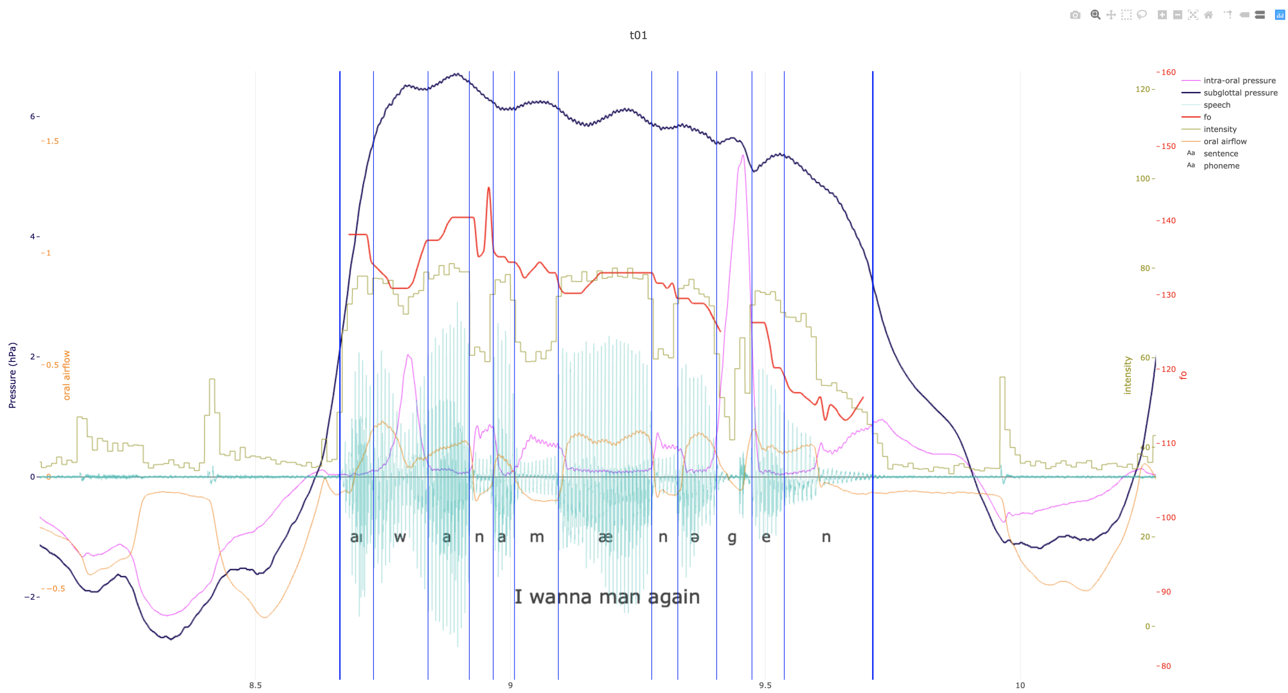

Figure 1: Interface of data visualisation tool

2. There is an effect on the vowels, 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 showed that the relationship varies as a function of the type of vowels and of the pitch.

The effects of oral and nasal vowels on P_s were examined by Demolin et al. (2017). Results show that both speakers effectively produced each of the vowels with a 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 sets 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 is to obtain aerodynamic measurements for singleton and geminated consonants. These consonants consist especially of a set of modal and glottalised consonants. In addition, various consonantal contrasts were recorded: labialised vs. non-labialised consonants; glottalised vs. non-glottalised, and ejectives. A detailed list can be found on the website of this database.

4. Database Website

A website is developed to host the database at the Laboratory of Phonetics and Phonology of the Sorbonne Nouvelle University, Paris 3. The database is designed with the SQL language to exploit relational databases for the ease of data selection according to the characteristics of linguistic materials, and the website is designed in PHP. All the data presented in this article are accessible and downloadable via the database website `https://corpus.ilpga.fr/aerodynamics`.

As the data are originally recorded with Phonedit and its specific data format, a python package is developed to convert data to popular signal format such as .wav and plain text file. This package gives more flexibility to explore the aerodynamic database and allows data processing with modern data analysis software. In addition, a visualisation tool based on plotly (Sievert et al., 2016) and Dash (`https://dash.plotly.com`) is developed to allow inspection of multiple signals simultaneously (Figure 1), an interactive HTML page is rendered with the visualisation tool with the basic functionalities of signal visualisation software. For example, labels of measure on top-right enable show/hide of one or more signals to focus on targeted parameters, synchronised values of multiple parame-

ters can be obtained by clicking directly on the graph. This visualisation tool allows interactive visualisation and exploitation of the database. Each measure is rendered in separate layers so that a focalised view on one specific signal and comparative view of multiple measures can be used to study qualitatively aerodynamic measures related to acoustic signals.
Most of the recordings of this database are segmented and annotated. Annotations can be downloaded from the database website as .TextGrid file and a python package of annotation conversion is also provided for visualisation of both signals and annotations.

5. Conclusion

This speech aerodynamic database provides simultaneous and synchronous recordings of acoustic and aerodynamic measurements of speech production in three languages. Data processing tools and visualisation interface are developed based on this database to enable direct exploitation of this database and allow for conversion to other data formats that can be directly processed with most popular softwares. Previous studies using this database focused mainly on various aspects of language-specific aerodynamic processes and acoustic outcome. A full description of the linguistic material and annotation grants feature extraction and thus allow for studies on cross-linguistic aerodynamic processes. The Speech Aerodynamic Database provides a set of precious aerodynamic measurements on speech production, containing in particular direct P_s measurements by tracheal puncture, which require hospital supervision and are extremely difficult to acquire. This database will allow a substantial development of studies on the aerodynamics of speech production and the modelling research.

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.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, 5(9):341–345.

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.

Draper, M., Ladefoged, P., and Whitteridge, D. (1959). Respiratory muscles in speech. *Journal of Speech and Hearing Research*, 2(1):16–27.

Ghio, A. (2002). Phonedit: Multiparametric speech analysis. *LPL (Aix-en-Provence, France) www. lpl. univ-aix. fr*.

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 *INTERSPEECH*, 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 sousglottique dans la production de la parole et sa contribution à la fréquence fondamental : une étude aérodynamique et acoustique sur l'anglais. Master's thesis, Université Sorbonne Nouvelle.