The Speed-Vel Project: a Corpus of Acoustic and Aerodynamic Data to Measure Droplets Emission During Speech Interaction

Carbone[1] F., Bouchet[2] G., Ghio[1] A., Legou[1] T., André[1] C., Kadri[1] S., Lalain[1] M., Petrone[1] C., Procino[3] F., Giovanni[1,4] A.

(1) Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France
(2) Aix Marseille Univ, CNRS, IUSTI, Marseille, France
(3) Università degli Studi di Napoli Federico II, Naples, Italy
(4) Aix Marseille Univ, ENT-HNS Department, La Conception University Hospital, Marseille, France

LPL, 5 Avenue Pasteur, 13100 Aix-en-Provence, France
{francesca.carbone, gilles.bouchet, alain.ghio, thierry.legou}@univ-amu.fr

Abstract

Conversations (normal speech) or professional interactions (e.g., projected speech in the classroom) have been identified as situations with increased risk of exposure to SARS-CoV-2 due to the high production of droplets in the exhaled air. However, it is still unclear to what extent speech properties influence droplets emission during everyday life conversations. Here, we report the experimental protocol of three experiments aiming at measuring the velocity and the direction of the airflow, the number and size of droplets spread during speech interactions in French. We consider different phonetic conditions, potentially leading to a modulation of speech droplets production, such as voice intensity (normal vs. loud voice), articulation manner of phonemes (type of consonants and vowels) and prosody (i.e., the melody of the speech). Findings from these experiments will allow future simulation studies to predict the transport, dispersion and evaporation of droplets emitted under different speech conditions.

Keywords: droplets emission, speech production, aerodynamic data

1. Introduction

Research on droplets emission during speech production has only recently gained attention after the emergence of the pandemic disease caused by SARS-CoV-2. The few lines of studies dealing with this topic have suggested that voice intensity (normal vs. loud speech), articulation manner of phonemes (i.e., type of vowels and consonants) and voicing (voiced vs. voiceless phonemes) are significant factors influencing the emission rate of the droplets. Concerning intensity, a common result is that loudness is positively correlated with the emission rate of droplets (Asadi et al., 2019; Anfirud et al., 2020; Stadnytskyi et al. 2020). For instance, Asadi and co-workers (2019) have shown that the number of droplets produced during loud speech is higher than during "calm" speech or calm breathing from 1 to 50 droplets/s (which corresponds to a rate of 0.06 to 3 particles per cm^3), regardless of the language used (English or Spanish). Voice loudness, however, seems to not determine an increase of the airflow initial velocity which is a key factor influencing droplets trajectory (Bourouiba, 2020; Giovanni et al., 2020). In a preliminary work, Giovanni and co-workers (2020) have analysed the initial velocity of the exhaled air during vocal exercises through a propylene glycol cloud produced by 2 e-cigarettes' users. Their first results showed relatively slow flow rates and unlikely to entail an increased risk of direct propagation of particles emitted in a loud voice (50 to 80 cm/s) compared to simple expiratory breath (80 to 180 cm/s). The authors explained that this result might be determined by the shock between vocal folds that causes a significant drop in the air pressure which goes from 5 to 10 hPa in the subglottic area to slightly above atmospheric pressure in the supra glottic airway (Giovanni et al., 2020). Another factor impacting speech droplets production is the articulation manner of the phonemes. Plosive consonants (such as /p/, /t/) have been shown to generate more droplets than fricative ones (e.g., as /f/, /v/; Jennison, 1942; Asadi et al., 2020). Using an Aerodynamic Particle Sizer (APS), Asadi and co-workers (2020) have investigated how vocalization of specific English vowels and consonants modulates particles emission rates during speech. Participants were asked to read with a normal level of loudness (as in normal conversation) three different types of speech material: single vowels /a/, /i/, /u/; disyllabic words containing plosives, fricatives, and nasals (e.g., *baba, fafa, nana*); and a short excerpt of the Rainbow passage text (i.e., a text frequently used by speech therapists to assess vocal ability, Fairbanks, 1960). Their results showed that the vowel /i/ generate more particles than the vowel /a/ and that plosives produce more particles than fricatives. One possible explanation of these outcomes may lie in the physiological mechanisms underlying the production of the phonemes. Both the production of close front vowels (such as /i/) and plosives involve the use of a large amount of airstream from lungs (Johnson and Morawska, 2009) and the narrowing of the vocal tract. These physiological characteristics may lead also to an acceleration of the airflow. For instance, plosives have been shown to produce enhanced directed transport, including conical jet with average velocities of tens of centimeters per second and over long distances of about 1 m (Abkarian et al., 2020). Furthermore, voiced consonants (e.g., /b/) have been shown to produce a greater number of droplets than voiceless ones (e.g., /p/, Morawska et al., 2009; Asadi et al., 2020). Such an increase in droplets production is supposed to occur at the level of the vocal folds (Wei and Li, 2016) whose vibration can cause droplets generation due to the aerosolization of secretions lubricating the vocal folds (Johnson and Morawska, 2009; Morawska et al., 2009).

1.1 Our proposal to previous limitations

Previous literature has shown several methodological limitations. Research has mostly concentrated their attention on the impact of speech properties on the droplets emission rate (e.g., Asadi et al., 2019; Asadi et al., 2020). Other measures that are crucial for characterising droplets dissemination (e.g., duration of emission, average and total oral airflow during emission, expired air velocity; Bourouiba, 2020) have been overlooked or not adequately considered. For instance, Giovanni and co-workers (2020) have investigated the velocity of the exhaled air using a propylene glycol cloud produced by 2 e-cigarettes' users. This device, though, only delivers 0.3 microns droplets whose permanence in the air is around 11s (Bertholon et al., 2012), not allowing more accurate measurements.

Significant limitations are shown also from a linguistic point of view. Most previous studies have employed speech material consisting of single words, syllables and isolated phonemes (mostly in English), resulting in low ecological validity. In the few cases where utterances and texts were used, they were usually pronounced outside of a communicative or interactional context (Asadi et al., 2019, Abkarian et al., 2020; Asadi et al., 2020), making them unrepresentative of everyday life conversations.

The present paper describes the experimental protocol of an ongoing project called "SpeeD-Vel" (Speech Droplets Velocity) that seeks to overcome both methodological and theoretical limitations of previous literature. We aim at investigating multiple measures of droplets dissemination (velocity and direction of the airflow, number and size of droplets) for the same speech materials and for the same speakers, hence having a more complex picture of the impact of speech on droplets emission. Differently from Giovanni et al. (2020), we provide a more accurate measure of the airflow velocity during speech production using EVA2 workstation (SQLab-LPL, Aix en Provence, France; Teston and Galindo, 1995) and the hot-wire anemometer(s) (one in Experiment 1 and three in Experiment 2; Bruun et al., 2006). EVA2 workstation allows measuring the airflow volume in l/s through a flow meter based on a resistive grid (pneumotachograph principle) with a small dead volume and specific linearization for the exhaled airflow (Ghio and Teston, 2004). The hot-wire anemometer is a thermal transducer that permits instantaneous flow velocity (m/s) to be calculated from electric voltage measurements (Bruun et al., 2006). It presents a wire that is electrically heated and cooled by the passage of the airflow. Such a cooling determines a change in the resistance of the wire and give us measures of the airflow velocity. The advantage of the hot-wire anemometer is associated with its very high temporal resolution and excellent frequency response characteristics (Barratt et al, 2016).

From a linguistic point of view, our study is, to our knowledge, the first to attempt to investigate the emission of droplets during speech production in an interactive setting, here in the ecologically valid situations of a dictate to an interlocutor. Considering the constraints of the measurements, we developed an experimental paradigm using restricted forms of interaction, based on partially pre-recorded and partially live interaction.

Another novelty of our study is that we looked at the effects of prosody and the possible interaction with segmental identity (consonants and vowels), a topic which has been completely neglected by the literature on droplets emission.

In prosody research, universal theories of intonation have proposed biologically determined codes, such as the "Effort Code" (e.g., Gussenhoven, 2016), based on the fact that greater glottal effort is employed "iconically" for prosodic meaning, such as in the case of accentual prominence or focus realizations. For instance, in a sentence such as "MARY will come tomorrow", "MARY" is the most important piece of information of the utterance (i.e., it's the focused constituent) and it is contrasted with another potential referent (e.g., "JOHN" in "JOHN will come tomorrow"). Phonetically, prosodic focus in French is signalled by multiple parameters, such as fundamental frequency (f0, the main acoustic cue of intonation), intensity, segmental duration, voice quality and spatial/temporal extent of articulatory movements. As a consequence, prosodic focus can lead to an increase in physiological effort, potentially increasing droplets emission.

2. Instrumental protocol

We report in the following paragraphs the instrumental protocol of three ongoing experiments aiming at investigating speech droplets production in French. In Experiment 1, we measured the airflow volume (l/s) emitted during speech using EVA2 workstation (Ghio et al., 2012) and the airflow velocity (m/s) using one hot-wire anemometer. In Experiment 2, we measured the airflow velocity (m/s) emitted during speech using three hot wire anemometers disposed at different position in space around the mouth. In Experiment 3, we measured the number and size of droplets emitted during speech using an Aerodynamic Particle Sizer (APS). Phonetic (articulation manner of phonemes, voice intensity) and prosodic conditions (focus) are manipulated via two linguistic tasks.

2.1 Participants

Twenty-three native French speakers between ages eighteen and fifty-five (21 F and 2 M), without respiratory or speaking disorders, have participated so far in the experiments. All participants were asked to complete a brief questionnaire including age, gender, weight, height, general health status and smoking history as these variables could influence breathing and speech production. The study complied with the Declaration of Helsinki. It was approved by the Ethics Committee of Aix-Marseille University. Speakers signed an informed consent form before the experiment.

2.2 Linguistic tasks

For each experiment, participants were asked to accomplish two linguistic tasks that were presented in random order. During the instructions, they were told to be connected online with an interlocutor who transcribed their speech production (pseudowords and sentences) to test a telecommunication system. We expected the speakers to adopt a specific vocal modality, known as "Precise Speech" or "Projected Voice" (Le Huche, 1984), corresponding to an increase of the energy employed throughout the speaker's vocal production and which is related to the importance of the speaker's message. From a phonetic point of view, these vocal modalities are characterized by an increase in the subglottal pressure (SGP) and in the duration of closure of the vocal folds.

Such a fiction was played through several elements. At the beginning of the experiment, participants were shown with a short video in which the fictitious interlocutor (one of the experimenters, a female Native speaker aged 48 y.o.) introduced herself. At the end of each repetition, we showed the participants an identification score of the pseudowords/sentences transcribed correctly by the (fictitious) interlocutor. The experimenter fed the fiction also by talking frequently with the fictitious interlocutor giving her feedback on the identification score. At the end of the experiments, we asked the participants whether they understood that the interaction was not real. None of the currently tested participants understood that the interaction was fake.

Pseudowords task. Speakers were instructed to dictate twelve trisyllabic pseudowords to a fictitious interlocutor. Pseudowords were composed of three syllables contrasted by voicing (voiced vs. unvoiced consonants, e.g., *pataka*, *bagada*), articulation manner (plosives vs. fricatives, e.g., *badaga*, *vazaja*), and vowel type (low vowel /a/, back vowel /u/, high-front vowel /i/, e.g., *pataka*, *poutoukou*, *pitiki;* a complete list of the pseudowords is reported in Table 1). We excluded from the corpus the nasals as they involve the emission of airflow from the nose which may impact the measurement of the airflow's velocity from the mouth.

Each pseudoword was pronounced with a neutral prosody consequently five times with a pause of one second. The complete list of pseudowords was repeated three times in random order. It was produced with two levels of intensity (normal vs. loud) within two randomised blocks (one block for each level of intensity; 12 pseudowords X 5 times X 3 repetitions X 2 levels of intensity = 360 pseudowords). Speakers controlled the intensity level through a Standalone Feedback Vumeter. They were asked to reach the yellow level of the vumeter and to keep the voice intensity constant as best they could. The parameters of the vumeter were changed according to the intensity of the voice (normal vs. loud). Speakers were given a short break at the end of each repetition during which they systematically drunk a sip of water to keep the hydration high. A training session preceded the experiment.

Consonants	Plosives	Fricatives
unvoiced	pataka	fassacha
	poutoukou	foussouchou
	pitiki	fissichi
voiced	badaga	vazaja
	boudougou	vouzoujou
	bidigui	viziji

Table 1. A complete list of pseudowords employed in the Pseudowords task.

Prosodic task. This task considers the effects of prosody by manipulating the position of the most prominent word in the sentence, as prominence leads to stronger degrees of effort at the word level (Gussenhoven, 2016). An elicitation procedure was designed to obtain variation in sentence level prominence naturalistically using a question-answer paradigm. Speakers were instructed to dictate to the fictitious interlocutor their answers to questions pre-recorded by one of the experimenters (who played the fictitious interlocutor). Questions induced the participants

to focus on the target words (indicated in uppercase on the screen) that were constituted by the pseudowords used in Pseudowords task and embedded in the carrier sentences ("focus" condition). Sentences were composed prevalently by voiced consonants with the same manner of articulation (either plosives or fricatives) to limit the influence of consonants type on the release of droplets. In addition to the "focus" condition, we used as baseline questions not triggering focus on any sentence constituents [e.g., "focus" condition: *Est-ce que Christian donne la dague? (Non). MONSIEUR BADAGA donne la dague,* 'Is Christian giving the dagger? (Non), Mr. BADAGA gives the dagger'; e.g., "no focus" condition: *Qu'est-ce qu'il se passe? Monsieur Badaga donne la dague*, 'What is going on? Mr. Badaga gives the dagger; a complete list of the questions and answers is given in Table 2]. Speakers repeated three times (in random order) five questions/answers for each consonant group (plosives, fricatives) for the "focus" and "no focus" conditions. As for the Pseudowords task, they pronounced the sentences with two level of intensity (normal vs. loud) within two randomised blocks [5 sentences X 2 consonant type (fricatives, plosives) X 2 focus conditions (focus vs. no) X 3 repetitions X 2 intensity level = 120 sentences]. They controlled the intensity level by themselves through the vumeter. They were given a short break after each repetition for drinking a sip of water. The experiment was preceded by a training session.

No focus condition		
Consonant group	Question	Answer
Plosives	*Qu'est-ce qu'il se passe ?* 'What is going on?'	*Monsieur Badaga donne la dague* 'Mr Badaga gives the dagger'
		Monsieur Badaga déduit la dette 'Mr Badaga deducts the debt'
		Monsieur Badaga goûte le gâteau 'Mr Badaga tastes the cake'
		Monsieur Badaga gobe les bonbons 'Mr Badaga gobbles up the sweets'
		Monsieur Badaga boit la bibine 'Mr Badaga drinks the booze'
Fricatives		*Monsieur Vazaja vise le voleur* 'Mr. Vazaja targets the thief''
		Monsieur Vazaja venge le voisin 'Mr Vazaja avenges the neighbour'
		Monsieur Vazaja jette la valise 'Mr Vazaja throws away the suitcase'

		Monsieur Vazaja vole le vélo 'Mr Vazaja steals the bike'
		Monsieur Vazaja joue le joker 'Mr Vazaja plays the joker'
Focus condition		
Plosives	*Est-ce que Robert donne la dague?* 'Is Robert giving the dagger?'	*(Non). MONSIEUR BADAGA donne la dague* '(No). MR BADAGA gives the dagger''
	Est-ce que Michel déduit la dette ? 'Is Michel deducting the debt?'	*(Non). MONSIEUR BADAGA déduit la dette.* '(No). MR BADAGA deducts the debt'
	Est-ce que David goûte le gâteau? 'Is David tasting the cake?'	*(Non). MONSIEUR BADAGA goûte le gâteau.* '(No). 'MR BADAGA tastes the cake'
	Est-ce que Xavier gobe les bonbons? 'Is Xavier eating the candies?'	*(Non). MONSIEUR BADAGA gobe les bonbons.* '(No). 'MR BADAGA gobbles up the sweets'
	Est-ce que Christian boit la bibine? 'Is Christian drinking the booze?''	*(Non). MONSIEUR BADAGA boit la bibine.* '(No). MR BADAGA drinks the booze''
Fricatives	*Est-ce que Robert vise le voleur?* 'Is Robert targeting the thief?'	*(Non). MONSIEUR VAZAJA vise le voleur.'* '(No). MR. VAZAJA targets the thief'
	Est-ce que Michel venge le voisin ? 'Is Michel avenging the neighbour?'	*(Non). MONSIEUR VAZAJA venge le voisin.* '(No). MR VAZAJA avenges the neighbour'
	Est-ce que Xavier jette la valise ? 'Is Xavier throwing away the suitcase?'	*(Non). MONSIEUR VAZAJA jette la valise.* '(No). MR VAZAJA throws away the suitcase'
	Est-ce que David vole le vélo ? 'Is David stealing the bike?'	*(Non). MONSIEUR VAZAJA vole le vélo* '(No). MR VAZAJA steals the bike'

Est-ce que Christian joue le joker ? 'Is Christian playing the joker?'	*(Non). MONSIEUR VAZAJA joue le joker* '(No). MR VAZAJA plays the joker'

Table 2. A complete list of questions/answers employed in the Prosodic task.

2.3 Experimental set-up

The general set-up of the experiments involved the use of several workstations. An Interactional Workstation (Figure 1) was used for the presentation of the linguistic material (pseudowords and questions/answers) to the participants. The material was presented by means of PERCEVAL software (André et al., 2003) and displayed to the participants via a screen connected by a VGA cable to the computer. A Standalone Feedback Vumeter (Android Application called "Audio Level ", Trajkovski Labs) was placed on a Samsung Tablet next to the screen to allow participants to check the level of voice intensity and at the same time to view the linguistic material (Figure 1, see Standalone Feedback Vumeter). Participants had to reach the yellow level of the vumeter. To induce a change in vocal effort, we adjusted the gain of the tablet microphone: low gain to induce loud speech vs high gain to induce normal speech. Two web cameras filmed the mouth of the speakers to ensure that they kept constantly the same position during data acquisition (Figure 1, see Video Controller).

The hot-wire anemometer controller was a StreamLine Pro CTA system, Dantec Dynamics. This device was connected to the probes (dimensional probes Dantec 55P11), to a thermocouple, to a Multifunction I/O Device National Instruments via BNC and to a computer equipped with the StreamLine software by a RS232 protocol (Figure 1, see Anemometer Controller). Simultaneous data acquisition was implemented through the EVA2 workstation (Figure 1, Data Acquisition Workstation). This device (SQLab-LPL, Aix-en-Provence, France, Ghio et al., 2012) could record simultaneously the wave sound and the physiological data (airflow, velocity). Audio recordings were made with a Neumann TLM 102 Cardioid Condenser microphone connected to an acoustic input of EVA2 workstation. The airflow volume (in l/s) was measured with the EVA2 built-in flowmeter. The air velocity was recorded by plugging the anemometer controller outputs to the auxiliary inputs of EVA2. All signals were synchronised as we used a single data acquisition system. Data were recorded, displayed and analysed with the SESANE Phonedit software developed by the LPL (www.lpl-aix.fr/~lpldev/phonedit).

Unlike Experiment 1 and 2, Experiment 3 did not require the use of the EVA2 workstation and the anemometer controller. Data acquisition was made through the Aerodynamic Particle Sizer (APS, TSI model 3221) and TSI Aerosol Instrument Manager Software for APS [Aerodynamic Particle Size Analyzer 3300, Instruction Manual (Serial Number = 145) (1983). TSI Inc., St. Paul,

MN]. Details about the experimental set-up of the single experiments are given in the following paragraphs.
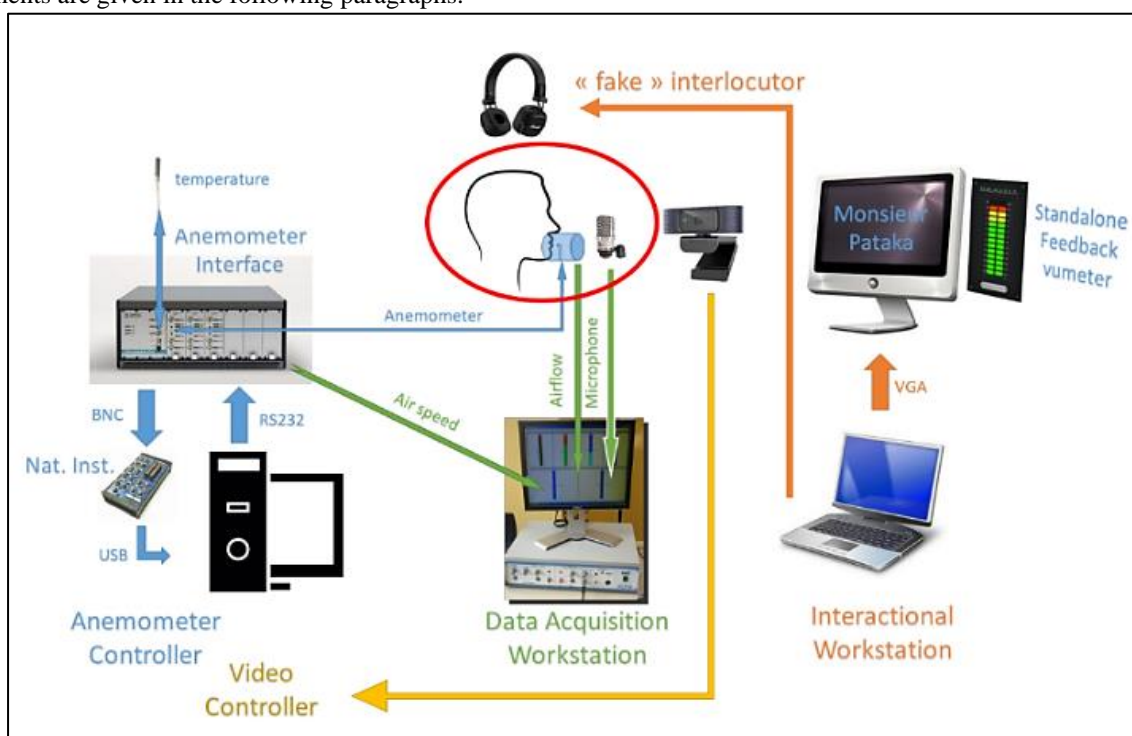

Figure 1.Workstations employed in the experiments

2.4 Experiment 1

The goal of this experiment was to measure simultaneously the expired airflow volume (l/s) and the airflow velocity (m/s) during speech production (Figure 2). By collecting these two measures, we had the advantage of correlating the aerophonometric data (airflow volume) with the airflow velocity, a measure mainly used in the field of fluid dynamics and unusual in the linguistic domain. In addition, a multiparametric correlation would allow cross-validation of the results, ensuring the robustness of the collected measurements.

The airflow volume (l/s) was recorded through a pneumotachograph with a stainless-steel wire (Ghio et al., 2004). It presented a reduced size (three cm diameter and two cm length) to optimize its response time and its linearity in all the articulatory contexts. It can capture a flow of the order of one cm^3/s because of its association with highly sensitive and stable differential pressure transducers (Honeywell DCXL). The resistance of the grid was 10 Pa by dm^3/s, i.e., approximately 1% of the intra oral pressure of a normal subject, which does not disturb the functionality of the vocal tract. The pressure tap was made in eight points of the circumference of the measurement pipe and a grid of tranquillization (negligible in resistance) was laid out in front of the pressure taps. This reduced the non-linearity of measurement caused by aerodynamic turbulences produced during speech production. The pneumotachograph was a built-in sensor of EVA2 workstation and provided directly calibrated data using the SESANE Phonedit software (Ghio et al., 2012). This device required speakers to wear a soft silicone rubber mask (Ghio et al., 2004) during linguistic tasks. The mask prevented air leakage without hindering articulatory movements.

The measure of the airflow velocity at the outlet of the mouth was performed using one hot-wire anemometer (Jorgensen, 2002) that provides good time resolution (Bruun, 1996). This device, driven through the StreamLine Pro CTA system (Dantec Dynamics), was embedded into the pneumotachograph. It presented a miniature wire probe with one sensor of five µm diameter (Dantec 55P11) forming the heating element. It consisted of 1.25 mm long plated tungsten wire sensors suspended between two straight prongs. A measure of the temperature of the exhaled air was carried out simultaneously using a thermocouple as the operating principle is based on cooling by forced convection of a small tungsten wire heated by the passage of an electric current. This device was placed at the outlet of the mouth parallel to the centre of the mouth so that the maximum airflow velocity was recorded. The output of the StreamLine Pro CTA system was an analogue electrical signal connected to an auxiliary input of EVA2 system. As the data were electrical tension, they were transformed as velocities using a calibration curve previously performed.

During the experiment, we also recorded the speech signal and SPL intensity in dB (independent of the recording gain). Speech signal was recorded with a microphone Neumann TLM 102, placed at 15 cm away from the lip corner. It was placed on the side so as not to disturb the airflow measurements (Figure 2). The EVA2 workstation had a built-in sonometer, that allowed us to compare calibrated SPL intensities between different recordings independently from the recording gain level. This system enabled us to compare the speech production in normal vs loud voice (see linguistic tasks).
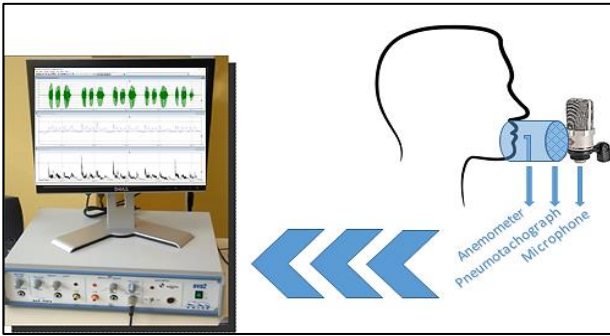
Figure 2. Simultaneous data acquisition with EVA2 workstation of expired airflow volume in l/s (pneumotachograph), airflow velocity in m/s (anemometer) and speech sound.

2.5   Experiment 2

The goal of this experiment was to simultaneously measure the expired airflow velocity (m/s) at different position in space around the mouth during speech production (Figure 3 and 4). This choice was motivated by the fact that the air jet exhaled during speech production is not uniform, neither in space nor in time. Such a non-uniformity results in variable airflow velocities depending on the position in the space.

The measurement of the airflow velocity at the outlet of the mouth was performed using three hot-wire anemometers with one-sensor probe (Dantec 55P11) driven through StreamLine Pro CTA system (Dantec Dynamics, see Experiment 1) and connected to the EVA2 station. They were separately mounted on a Dantec probe holder and arranged vertically and parallel to the speakers' mouth (Figure 3). Anemometer 1 was placed above the lips of the speakers, anemometer 2 parallel to the center of the lips, anemometer 3 below the lips. The vertical distance between anemometer 1 and 2 was 1.5 cm, between anemometer 2 and 3 was 2 cm. The horizontal distance between the three anemometers and the speaker's lips was 5.75 cm. Such an arrangement was thought to cover the whole cone of airflow emitted at the exit of the mouth during speech (Abkarian et al., 2020; Giovanni et al., 2020).
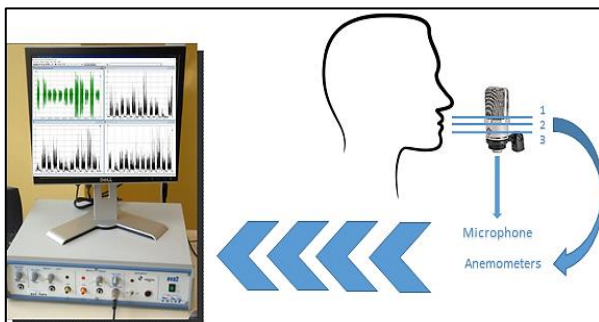

Figure 3. Simultaneous data acquisition with EVA2 workstation of expired airflow velocity on three positions (anemometers) and sound speech

The participant position was controlled visually by the experimenter through two webcams focusing on the mouth in a front and sagittal positions. A headrest helped the speakers to maintain the same head position, limiting their

movements. This experiment did not require speakers to wear a mask (as in Experiment 1) allowing measurements of exhaled airflow velocity without compression of the mouth. Like Experiment 1, we also collected the waveform of speech signal and the SPL intensity in dB (independent of the recording gain).
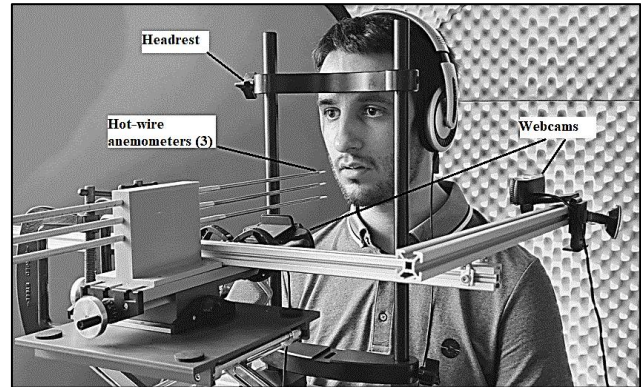

Figure 4. The three hot-wire anemometers, the two web cams and the headrest employed in Experiment 2.

2.6   Experiment 3

The goal of this experiment was to measure the size and number of droplets emitted during speech. Droplets' recordings were made through an Aerodynamic Particle Sizer (APS, TSI model 3321). Data acquisition was made via the TSI Aerosol Instrument Manager Software for APS [Aerodynamic Particle Size Analyzer 3300, Instruction Manual (Serial Number = 145) (1983). TSI Inc., St. Paul, MN]. The APS operates at a total flow rate of 5L/min (sheath flow rate $\cong$ 4L/min, sample flow rate $\cong$ 1L/min) and measures the size distribution of particles larger than 0.5µm, but only detects the presence of particles between 0.37µm and 0.5µm without providing precise size measurements.
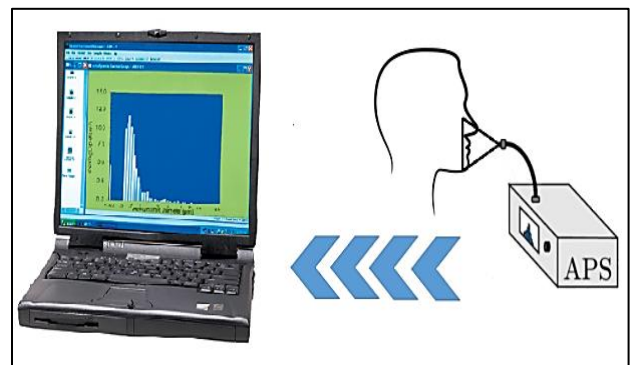

Figure 5. Data acquisition with the Aerosol software, the Aerodynamic Particle Sizer (APS)

During the experiment, participants sat in front of the APS and spoke into a metallic funnel (diameter = 14 cm) that was connected to the APS sampling inlet via a conductive silicon tube (distance between funnel hole to APS inlet=12 cm, tube inner diameter=2 cm). A tennis ball supported by a stand was placed behind the participants' heads to ensure that they kept their heads in the same position during tasks. A screen showing the stimuli was placed behind the APS

to guide participants in accomplishing the tasks. A voice recorder (Zoom H1n) was placed immediately on the side of the APS to record speech production.
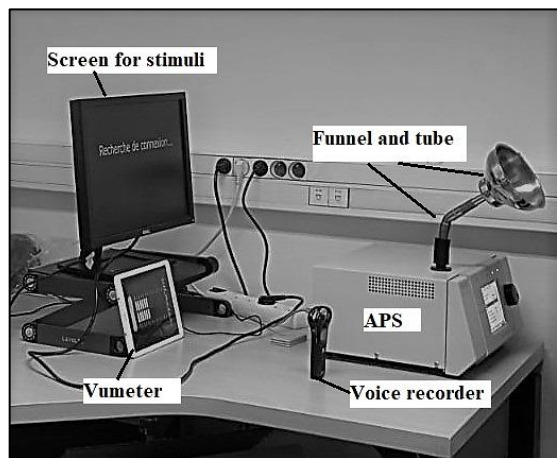

Figure 6. Devices (APS, funnel and tube, voice recorder, vumeter) employed in Experiment 3

We performed the experiments in a controlled environment. The air of room was filtered for reducing the number of particles by means of an air filtering system composed of a filter HEPA attached to a ventilator. Control experiments indicated that this system reduced the particles presented in the room of the 99 %. Due to the necessity of continuously recording data, we had to modify the linguistic tasks. For this experiment, participants were told to accomplish only the Pseudowords task and to pronounce the pseudowords without pauses (avoiding inspirations and expirations outside the pronunciation of the pseudowords) during seven seconds, of which we recorded the five central ones. This modified task allowed us to avoid the influence of expiration on the number and size of speech droplets. Prosodic task was excluded due to the difficulty of pronouncing consecutively sentences without inspirations for the whole duration of the experiment.

3. Data

We report here representative raw data for the three experiments. They are extracted from a 21-year-old female speaker (F7). More accurate analysis of all twenty-three speakers is still in progress.

Fig. 7 reports data from Experiment 1. In this example, F7 spoke the pseudoword /budugu/ with normal (a) vs. loud voice (b). /Budugu/ spoken with loud voice produced higher SPL intensity (max value: 67 dB), increased airflow velocity (max value: 0.188 decavolt, daV, corresponding to 7.23 m/s) and airflow volume (max value: 0.325 dm^3/s = l/m) than the same pseudoword spoken with normal voice (max SPL intensity: 56 dB; max airflow velocity: 0.172 daV / 3.37 m/s; max airflow volume: 0.152 dm^3/s = l/m).

Fig. 8 reports data from Experiment 2. Here, F7 spoke the sentence *MONSIEUR VAZAJA venge le voisin* 'MR VAZAJA avenges the neighbour' prosodically emphasizing the target words ("focus condition") with a normal (a) vs. loud voice (b). The sentence spoken with loud voice presented higher SPL intensity (max value: 81.27 db), increased airflow velocity for the three hot wire anemometers (max value for anemometer 1: 0.205 daV / 14.02 m/s, max value for anemometer 2: 0.172 daV / 3.37 m/s; max value for anemometer 3: 0.163 daV / 2 m/s) than the same sentence spoken with normal intensity (max value for SPL intensity: 75.87 dB; max value for anemometer 1: 0.202 daV / 12.58 m/s; max value for anemometer 2: 0.170 daV / 3.02; max value for anemometer 3: 0.161 daV / 1.76 m/s). In the case of this sentence, for both normal and loud voice, the highest airflow velocities were recorded by the anemometer 1.

Fig. 9 shows data from Experiment 3. Here, F7 spoke the pseudoword /badaga/ with a normal (a) vs. loud voice (b). The graph shows the raw number of droplets and their distribution according to the size (from 0.5 to 20 µm). When /badaga/ is pronounced with loud voice the number (overall mean: 7.74) and size of droplets increased compared to when this pseudoword was spoken with normal voice (overall mean: 5.03).
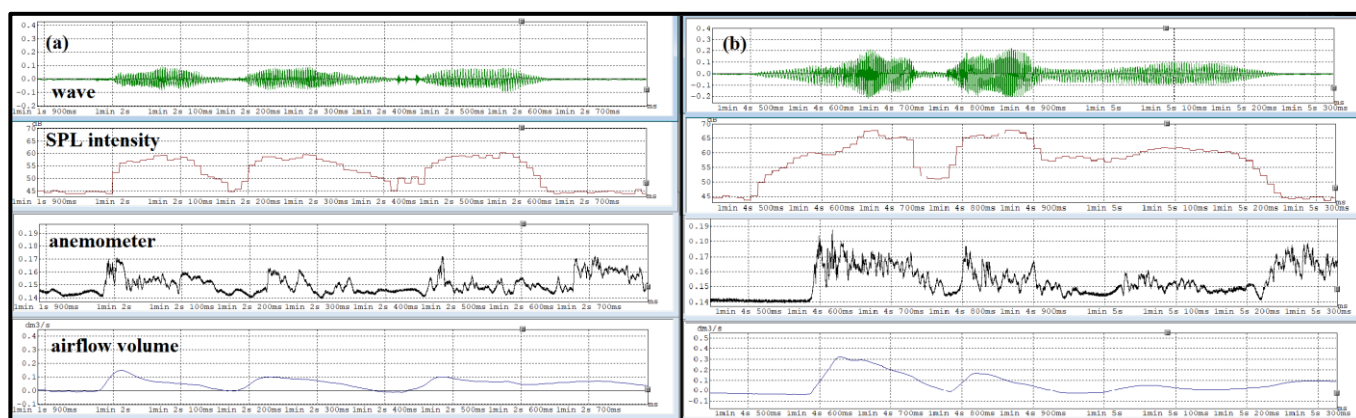

Figure 7 (Experiment 1). The wave signal (first row), SPL intensity (dB, second row), airflow velocity (in decavolt, daV) recorded by the hot wire anemometer (third row) and airflow volume (dm^3/s= l/s) recorded by the pneumotachograph (fourth row) for the pseudoword /budugu/ spoken with normal (a) vs. loud voice (b) by F7.
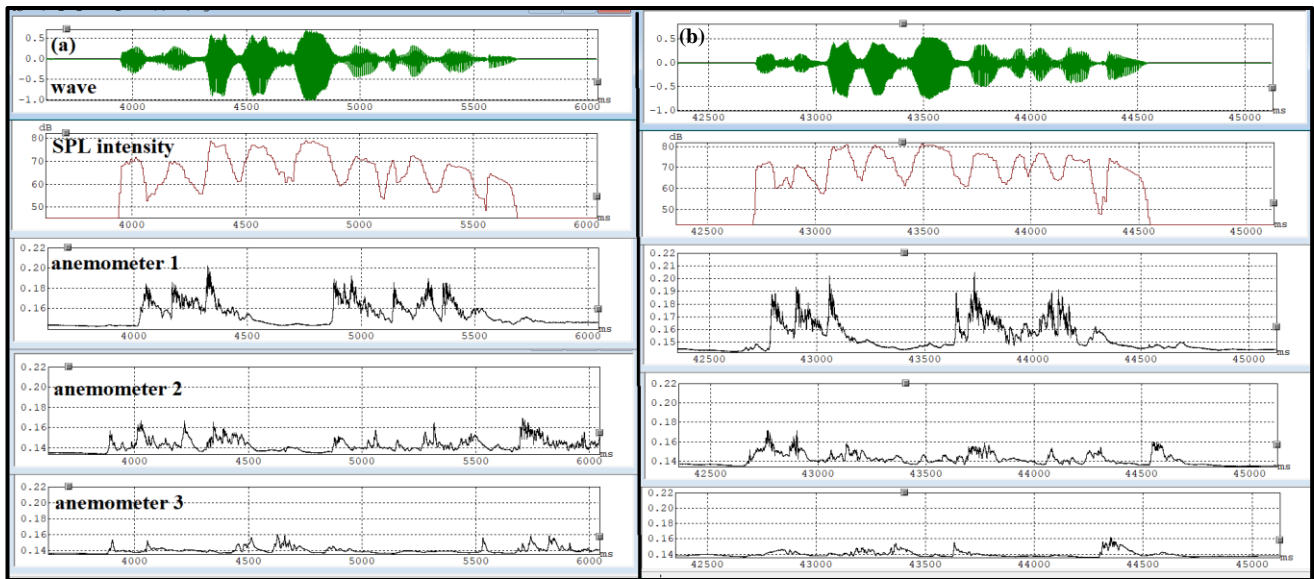
Figure 8 (Experiment 2). The wave signal (first row), SPL intensity (dB, second row), airflow velocity (in decavolt, daV) recorded by the anemometer 1(third row), anemometer 2 (fourth row), anemometer 3 (fifth row) for the sentence *MONSIEUR VAZAJA venge le voisin* 'Mr Vazaja avenges the neighbour'spoken with normal (a) vs. loud voice (b) by F7.
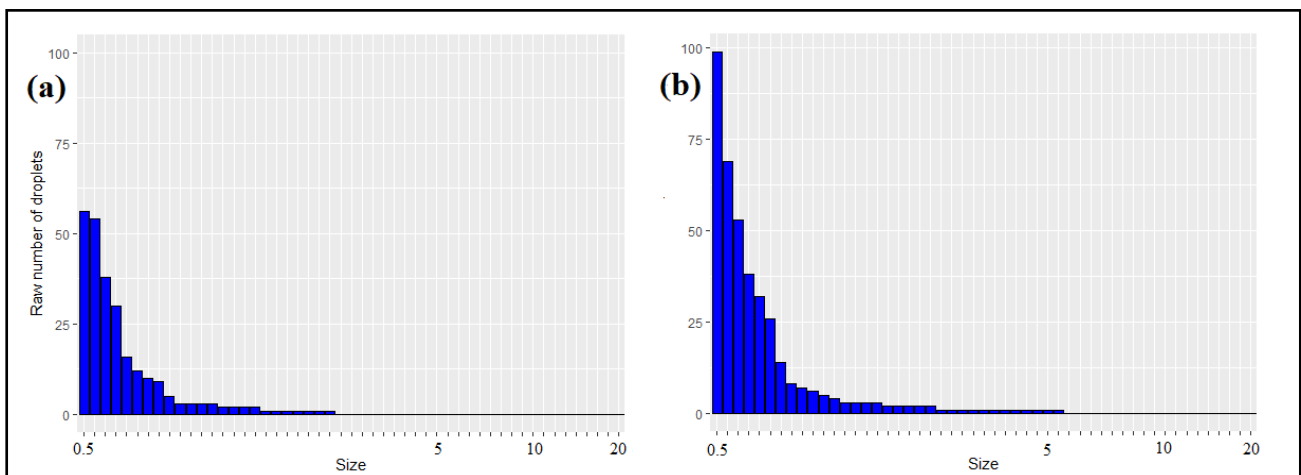

Figure 9 (Experiment 3). Bidimensional representation of the raw number of droplets versus their size (μm) recorded by the APS for the pseudoword /badaga/ spoken with normal voice (a) vs. loud voice (b) by F7.

4. Conclusion

The project Speed-Vel will provide reference data on droplets dissemination during interaction, considering phonemes characteristics and prosodic conditions. These findings will allow future simulation studies to predict the transport, dispersion and evaporation of droplets emitted under different speech conditions. The knowledge of these elements is essential for adapting protective devices (physical distance, masks) to limit the spread of the virus during speech interactions. This work also brings new knowledge on fluid mechanics in speech which could allow inference of speech production mechanisms in a non-invasive way.

5. Acknowledgments

6. Bibliographical references

Abkarian, M., Mendez, S., Xue, N., Yang, F., and Stone, H. A. (2020). Speech can produce jet-like transport relevant to asymptomatic spreading of virus. *In Proceedings of the National Academy of Sciences*, 117(41), pages 25237-25245.

André, Ghio, Cavé, Teston (2003). PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception. In *Proceedings of the*

International Congress of Phonetic Sciences (ICPhS), 2003, pages 1421-1424, Barcelona, Spain.

Anfinrud, P., Stadnytskyi, V., Bax, C. E., and Bax, A. (2020). Visualizing speech-generated oral fluid droplets with laser light scattering. *New England Journal of Medicine*, 382(21), pp. 2061-2063.

Asadi, S., Wexler, A. S., Cappa, C. D., Barreda, S., Bouvier, N. M., and Ristenpart, W. D. (2019). Aerosol emission and superemission during human speech increase with voice loudness. *Scientific reports*, 9(1), pp. 1-10.

Asadi, S., Wexler, A. S., Cappa, C. D., Barreda, S., Bouvier, N. M., and Ristenpart, W. D. (2020). Effect of voicing and articulation manner on aerosol particle emission during human speech. *PloS one*, 15(1), e0227699.

Barratt, D., Nguepnang, M. A., and Atkins, M. D. (2016). Fast-response velocity and shear stress measurements. In K. Tongbeum, J. S. Seung Jin (Eds.) Application of Thermo-Fluidic Measurement Techniques, Butterworth-Heinemann, Oxford, pp. 101-123.

Bertholon J-F, Becquemin MH, Roy M, Roy F, Ledur D, Annesi Maesano I et al (2013). Comparaison de l'aérosol de la cigarette électronique à celui des cigarettes ordinaires et de la chicha. *Revue des Maladies Respiratoires*, 30(9), pp. 752–757

Bourouiba, L., Dehandschoewercker, E., and Bush, J. W. (2014). Violent expiratory events: on coughing and sneezing. *Journal of Fluid Mechanics*, 745, pp. 537-563.

Bruun HH (1995) Hotwire anemometry, principles and signal analysis. Oxford Univ Press, UK.

Fairbanks G. (1960). Voice and articulation drillbook, Harper & Row, New York, NY, 2nd edition.

Ghio A., Pouchoulin G., Teston B., Pinto S., Fredouille C., et al.. (2012). How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers?. *Speech Communication*, 54 (5), pp.664-679.

Ghio A., Teston B. (2004). Evaluation of We used Sesane _ software the acoustic and aerodynamic constraints of a pneumotachograph for speech and voice studies. *In Proceedings of the International Conference on Voice Physiology and Biomechanics*, Univ. Méditerranée, Marseille, France, pages 55-58.

Giovanni, A., Radulesco, T., Bouchet, G., Mattei, A., Révis, J., Bogdanski, E., and Michel, J. (2020). Transmission of droplet-conveyed infectious agents such as SARS-CoV-2 by speech and vocal exercises during speech therapy: preliminary experiment concerning airflow velocity. *European Archives of Oto-Rhino-Laryngology*, 1-6.

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. Topics in Cognitive Science, 8 (2), pp. 425-434.

Jennison MW (1942). Atomizing of mouth and nose secretions into the air as revealed by high speed photography, *Aerobiology*, 17, pp. 106-28.

Johnson GR, Morawska L (2009) The mechanism of breath aerosol formation. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, 22(3), pp. 229–237.

Jorgensen, Finn E (2002). How to Measure Turbulence with Hot-Wire Anemometers, Dantec Dinamics, Denmark, 2002

Le Huche F. (1984), La voix: anatomie et physiologie des organes de la voix et de la parole. Masson, Paris.

Morawska L, Johnson GR, Ristovski ZD, Hargreaves M, Mengersen K, Corbett S, et al (2009). Size distribution and sites of origin of droplets expelled from the human respiratory tract during expiratory activities. *Journal of Aerosol Science*, 40(3), pp. 256-69.

Reetz, H., & Jongman, A. (2009). Phonetics: Transcription, production, acoustics, and perception, Wiley-Blackwell, Hoboken.

Stadnytskyi, V., Bax, C. E., Bax, A., and Anfinrud, P. (2020). The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. In *Proceedings of the National Academy of Sciences*, 117(22), pages 11875-11877.

Teston, B., and Galindo, B. (1995). A diagnostic and rehabilitation aid workstation for speech and voice pathologies. In *Proceedings of the Fourth European Conference on Speech Communication and Technology*.

Wei JJ, Li YG (2016). Airborne spread of infectious agents in the indoor environment. *American Journal of Infection Control,* 44(9), pp. S102-S8.