Aerodynamics of Fricative Production in European Portuguese

Cátia M. R. Pinho¹, Luis M.T. Jesus², Anna Barney³

¹ IEETA, Universidade de Aveiro, Portugal; ² IEETA and ESSUA, Universidade de Aveiro, Portugal; ³ ISVR, University of Southampton, UK

catiap@ua.pt, lmtj@ua.pt, ab3@soton.ac.uk

Abstract

The characteristics of steady state fricative production, and those of the phone preceding and following the fricative, were investigated. Aerodynamic and electroglotographic (EGG) recordings of four normal adult speakers (two females and two males), producing a speech corpus of 9 isolated words with the European Portuguese (EP) voiced fricatives /v, z, Z/ in initial, medial and final word position, and the same 9 words embedded in 42 different real EP carrier sentences, were analysed. Multimodal data allowed the characterisation of fricatives in terms of their voicing mechanisms, based on the amplitude of oral flow, F1 excitation and fundamental frequency (F0).

Index Terms: Speech aerodynamics, speech signals, speech production, fricatives.

1. Introduction

Our study investigates the phenomenon of devoicing in European Portuguese (EP) fricatives using aerodynamic as well as acoustic measures. Our long term goal is to understand the mechanisms by which voicing is initiated and maintained as a guide to improving strategies for initiating and maintaining voicing in patients with laryngeal impairment such as unilateral vocal fold paralysis.

It is the interaction between the moving airstream and the vocal tract geometry which generates speech sounds, so it is important to understand the various aerodynamic effects as well as the acoustic factors when we measure or analyse any aspect of speech production [1]. The study of fricatives is made harder by the relatively high flow velocities and constricted geometry in the acoustic source region, however Rothenberg mask [2] recordings provide a well-established means to estimate the flow velocity at the oral constriction and to identify the acoustic source location. They can be obtained under both normal and pathological speech conditions and have, for example, recently been recognised as a valuable velopharyngeal incompetence evaluation technique [3].

Jesus and Shadle [4] showed that voicing is often maintained over only part of a voiced fricative and that the devoicing rate was very high for EP fricatives, especially when compared with studies of other languages [5, 6]. This makes these sounds ideal for the study of onset and maintenance of voicing, a subject not investigated in great detail to date.

When speech is impaired due to organic lesions of the vocal folds, such as in unilateral vocal fold paralysis (UVFP), phonation clearly changes and patients find it hard to maintain voicing. Understanding when/why normal speakers devoice (e.g., during voiced fricative production) will help us develop new strategies/treatments to aid UVFP patients, and should permit identification of a baseline of

knowledge from which, ultimately, the relationship between pathological signals and medical diagnosis of voice pathologies can be developed.

In this paper we focus on the characteristics of steady state voiced fricative production relative to those of the preceding and following phone. Our goal is to establish by which aerodynamic and acoustic parameters the relatively weak voicing during the fricative may be differentiated from the stronger voicing of the contextual vowel.

2. Method

2.1. Speakers, Recording and Annotation

Data were collected from two adult female (JG and HV) and two adult male (LJ and RS) speakers of EP. None had reported speech, language or hearing disorders, and had normal vocal qualities. All were assessed by an experienced Speech and Language Therapist (SLT) using a standardised evaluation protocol.

Speakers were recorded producing 51 utterances, including 9 isolated words containing the EP voiced fricatives /v, z, Z/ in initial, medial and final word position, and the same 9 words embedded in 42 different carrier sentences, that represented a variety of consonantal (taps, laterals, stops and nasals) and vocalic (close, open front and back vowels) sounds in real EP sentence contexts.

Recordings were made in a quiet room using a Rothenberg mask [2] and a PTW-1 pressure transducer (Glottal Enterprises, USA) for measuring the airflow at the mouth. An EGG signal was also collected using an EGG processor (model EG2-PCX, Glottal Enterprises, USA). The oral airflow and EGG signals were recorded with a MS-110 electronics unit (Glottal Enterprises, USA), connected via an external sound card to a laptop computer running WaveviewPro Version 2.2.6 (16 bits, 44.1 kHz sampling frequency). Airflow calibration and zero-setting of signals were ensured before each recording session using Glottal Enterprises's FC-1 flow calibrator and WaveviewPro Version 2.2.6 standard procedures.

The time waveforms of all the corpus words were manually annotated using Praat Version 5.0.43 [7] to detect the start of Phone1 (the phone or silence preceding the fricative), the start and end of the fricative (Phone2) and the end of Phone3 (the phone or silence following the fricative).

2.2. Devoicing and Weak Voicing

Devoicing phenomena were studied, by Jesus and Shadle [4], using criteria that used both the acoustic and EGG signals to determine if a fricative was devoiced or not. Initially we applied their [4] criteria for manually labelling the voicing category of the fricative. However, this proved to be insufficiently precise for our purposes (see Section 3.1) and we developed an improved method for

classification when the voicing is relatively weak such as is found in a voiced fricative.

During fricative production the amplitude of the formant oscillations in the flow waveform (see Figure 6) is reduced (relative to that in the adjacent phones), i.e., there is more damping of the resonances due to an incomplete closure of the glottis. We refer to voicing without a strong formant excitation as *weak voicing*. However the relationship between weak voicing and the aerodynamic events within the larynx is not well defined currently.

It is important to define weak voicing (or devoicing) in terms of events in the flow waveform only, because we do not have a clear, well defined relationship between ripple in the EGG and the nature of physical activity in the larynx.

The ripple in the time domain flow signal is mainly at F1, so if the amplitude of F1 is <x% (see Section 3.2 for the choice of threshold criterion) of its value in a vowel (considering that there is strong voicing in a vowel), and we can detect a valid F0 in the flow wave, we define the voicing as weak. This definition requires careful detection of the presence or absence of voicing.

F1 was tracked and its intensity measured. A drop of x dB % (see Section 3.2) below the vowel level signalled the onset of weak voicing (providing F0 also existed). Similar schemes for voicing decisions based on the acoustic signal have been used successfully in the past [3].

We also used a threshold that could signal the onset of *weak voicing*, based only in the oral flow signal (% decrease of oral flow amplitude in the fricative production relative to adjacent vowel production).

2.3. New Measures

To characterise fricatives in terms of their production mechanisms, we analysed the oral flow signal, F0 and Flexcitation strength during the steady state of fricative and adjacent vowel production.

The strategy used to correlate and extract information on these different stages of speech production was based on average values calculated from 20 ms windows centred within Phone1, Phone2 and Phone3 (see Figure 1), in order to characterise the steady state of fricative production, relative to adjacent phones.



Figure 1: Normalized oral flow signal.

Absolute mean oral flow values were extracted at the middle of Phone1, Phone2 and Phone3, from all the recordings and for every speaker.

Information regarding the amplitude of oscillations was extracted from the oral flow waveform, at three 20ms windows, as shown in Figure 1. Matlab scripts were used to extract the amplitude values shown in Figure 2.

The relative vowel-fricative (Phone1-Phone2) amplitude values (A) and the relative fricative-vowel (Phone2-Phone3) amplitude values (B), were also calculated using:

$$A(\%) = \frac{\left[Mean(Phone1)_{W_{20mr}} - Mean(Phone2)_{W_{20mr}}\right]^* 100}{Mean(Phone1)_{W_{20mr}}}$$
(1)

$$B(\%) = \frac{\left|Mean(Phone3)_{W_{20ms}} - Mean(Phone2)_{W_{20ms}}\right| * 100}{Mean(Phone3)_{W_{20ms}}}$$
(2)



Figure 2: Criteria used to calculate the flow amplitudes.

F0 (Hz) and amplitude (dB/Hz) of F1 were measured, from a 20ms window for the three different stages of speech production shown in Figure 1. Data was extracted using the following open source Praat Version 5.0.43 functions:

- (F0): To Pitch ... 0 75 600; [Time step(s), Pitch floor(Hz), Pitch ceiling(Hz)] autocorrelation method.
- (F1): To Formant (burg)... 0.01 5 5500 0.025 50; [Time step(s), Max. number of formants, Maximum formant(Hz), Window length(s), Pre-emphasis from (Hz)] - split Levinson algorithm.

Absolute values of F0 and F1 amplitude and relative Phone1/Phone2 and Phone2/Phone3 F1 amplitudes were then calculated from the outputs of the two functions.

3. Results

3.1. Devoicing

Initially, a classification of the voicing category of fricatives produced by the four speakers, was produced using the criteria proposed by Jesus and Shadle [4]. Results showed that 36% (/v/ - 12%; /z/ - 18%; /Z/ - 6%) of recorded files were devoiced, much less than was reported (/v/ -48%; /z/ -77%; /Z/ - 78%) in [4]. Jesus and Shadle [4] used an acoustic speech signal, recorded with a microphone located 1m in front of the subject's mouth, and an EGG signal. The signals we based our decisions on were the airflow (recorded at the mouth) and the EGG. Applying the criteria of Jesus and Shadle [4] is not the ideal solution, because the very low amplitude oscillations that we frequently observed in the flow signal (probably resulting from minute mucosal oscillations for an open glottis), and which lead to a categorisation of voiced under their criteria, are not always related to periodicity in the acoustic signal. We therefore devised a new strategy to characterise voicing of the phonemes in our Corpus, and defined (see section 2.2) a new term: weak voicing.

3.2. Oral Flow

A decrease in the amplitude of the oral flow waveforms, during the production of a fricative, when compared to the amplitude of the previous and following phone (see Figure 6), was observed for each of the four speakers. The absolute amplitude values for female speakers were lower than for male speakers, as shown in Table 1. The mean values presented were calculated using more than 30 different speech samples.

Oral Flow (cm ³ /s)	് (LJ & RS)			♀ (JG & HV)		
Phone	1	2	3	1	2	3
/v/	379±97	107±35	316±90	229±79	44±27	180±73
/z/	352±76	100±72	306±76	220±87	33±19	168±70
/Z/	348±15	87±57	309±88	188±67	51±26	167±80

Table 1. Mean \pm std absolute values of oral flow.

Typical relative average values (see equations 1 and 2), such as those shown in Figure 3, and a thorough examination of the data by comparing the vowel and fricative oral flow values and the phoneme boundaries set during the annotation phase, led us to set a threshold (that could be used to define the onset of weak voicing) at x = 60%. The numbering in the x axis of Figure 3, Figure 4 and Figure 5, corresponds to the numbering of the files in the Corpus. These numbers vary from graph to graph because we are analysing different fricatives (that correspond to different files in the Corpus).



Figure 3: Flow relative values for HV's /v/.

Results of classifying all the recordings of the four speakers with this criterion are present in the Table 2. There were very few examples with no voicing (reported in Table 2 as having weak voicing).

Table 2.	Fricatives classified (based on flow) as	
having	weak and strong voicing respectively.	

Fricatives	/v/ (%)	/z/ (%)	/Z/ (%)
LJ 👌:			
Weak voicing	59	82	82
Strong voicing	41	18	18
RS ∂:			
Weak voicing	88	88	82
Strong voicing	12	12	18
JG ♀:			
Weak voicing	76	88	47
Strong voicing	24	12	53
HV ♀:			
Weak voicing	94	71	76
Strong voicing	6	29	24

The mean absolute values of F0 decreased during fricative production: $117 \text{ Hz} \le \text{Phone1} \le 127 \text{ Hz}$, $110 \text{ Hz} \le \text{Phone2} \le 119 \text{ Hz}$ and $118 \text{ Hz} \le \text{Phone3} \le 123 \text{ Hz}$ for male speakers; $203 \text{ Hz} \le \text{Phone1} \le 210 \text{ Hz}$, $191 \text{ Hz} \le \text{Phone2} \le 194 \text{ Hz}$ and $204 \text{ Hz} \le \text{Phone3} \le 207 \text{ Hz}$ for female speakers.

The relative Phone1/Phone2 and Phone2/Phone3 F0 values (from equations 1 and 2) varied from 0 to 16%, as shown in Figure 4.



Figure 4: F0 relative values for LJ's /Z/.

Analysis of F1 amplitude revealed that for a steady state of 20 ms, the difference between Phone1 and Phone2, and the difference between Phone3 and Phone2, presented values in the range of 16-42 dB/Hz for male speakers and in the range of 19-36 dB/Hz for female speakers.

The relative average values of the decrease of the Phone1/Phone2 F1 amplitude and increase in the Phone2/Phone3 F1 amplitude varied between 50 and 200%, (as shown in Figure 5). These relative values were also calculated using equations 1 and 2.



Figure 5: F1 amplitude relative values for LJ's /z/.

Classification results based on a 60% threshold, showed the following percentages of weak voicing: $|v| \ge 88\%$, $|z| \ge 71\%$ and $|Z| \ge 59\%$ for male speakers; $|v| \ge 71\%$, $|z| \ge 82\%$ and $|Z| \ge 53\%$ for female speakers.

4. Conclusions

The aerodynamic conditions under which voicing can be initiated and maintained when the voicing source is relatively weak, as in a voiced fricative, are little explored or understood at present. Nevertheless these correspond to the prevalent conditions in pathological speakers and their study will contribute to our ability to design improved therapy for voice rehabilitation.

We have characterised the changes in three aerodynamic parameters, oral airflow, F0 and F1 amplitude, during voiced fricative production compared to their values in contextual vowels and have shown on average reductions in these parameters during fricative production for all VFV tokens in our corpus for both male and female speakers. These parameters were chosen for analysis in part due to the relative ease with which they can be measured for both normal and pathological speakers.

In addition to the usual voiced/unvoiced classification of signals, a new and more general criterion for the definition of *weak voicing*, based on the analysis of the flow signal alone, was also presented. This will allow us in the future to objectively identify portions of voicing in normal speech that correspond most closely to the mode and manner of voicing generally found in UVFP pathological speech.

Classification results, using a specific threshold that defines the onset of weak voicing based on the relative amplitude of the flow signal, and based on the amplitude of F1, showed a comparable number of fricatives as having weak voicing, and the number of weakly voiced tokens was also similar to the number designated by Jesus and Shadle [4] as being devoiced or partially devoiced (/v/ – 61%; /z/ – 89%; /Z/ – 91%). Our definition has the advantage that it is based on the flow signal alone and therefore does not rely on subjective interpretation of the relationship between the EGG signal and laryngeal activity.

Our future work will focus on defining the laryngeal mechanisms and the vibration mode pertaining during weak voicing in normal subjects and in drawing parallels with the corresponding conditions in UVFP patients to develop a model of UVFP voicing that may be used to enhance therapy for vocal rehabilitation.

5. Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia, Portugal (PTDC/SAU-BEB/67384/2006).

6. References

- C. H. Shadle, "The Aerodynamics of Speech," in The Handbook of Phonetic Sciences W. J.Hardcastle and J. Laver, Eds.: Blackwell 1997.
- [2] M. Rothenberg, "A new inverse-ffitering technique for deriving the glottal air flow waveform during voicing," The Journal of the Acoustical Society of America, vol. 53, pp. 1632-1645, 1973.
- [3] J. P. Dworkin, M. T. Marunick, and R. J. Stachler, "Neurogenic Velopharyngeal Incompetence: Multidisciplinary Diagnosis and Treatment," The Journal of Applied Research in Clinical Dentistry, vol. 1, pp. 35-46, 2004.
- [4] L. M. T. Jesus and C. H. Shadle, "A Parametric Study of the Spectral Characteristics of European Portuguese Fricatives," Journal of Phonetics, vol. 30, pp. 437-464, Jully 2002.
- [5] K. Pirello, S. E. Blumstein, and K. Kurowski, "The characteristics of voicing in syllable-initial fricatives in American English," Acoustical Society of America, pp. 3754-3765, 1997.
- [6] C. L. Smith, "The devoicing of /z/ in American English: effects of local and prosodic context," Journal of Phonetics, vol. 25, pp. 471-500, 1997.
- [7] P. Boersma, "Praat, a system for doing phonetics by computer," in Glot International, 2001, pp. 341-345.



Figure 6: Oral flow waveforms from words with the fricative /v/ (17 files), recorded by male speaker LJ. The vertical lines represent the Phone1-Phone2 and Phone2-Phone3 boundaries where phone2 is the fricative. The x and y axes were normalised.