ACOUSTIC ANALYSIS OF A SPEECH CORPUS OF EUROPEAN PORTUGUESE FRICATIVE CONSONANTS

Notes

Introduction

- There have been many studies of the phonetics and phonology of Portuguese. These have shown some interesting features of Portuguese; it is unusually rich in instances of vowel reduction, consonant clusters, and fricatization of plosives (Viana 1984).
- There have been few and limited studies of Portuguese fricatives. These studies included the analysis of duration and acoustic features, and some perceptual experiments (Lacerda 1982; Viana 1984; Andrade 1995; Martins et al. 1995).
- The work to be presented in this talk, is a systematic study of European Portuguese fricatives in a range of contexts, with temporal and spectral analysis. Our long-range goal is synthesis of Portuguese fricatives.
- We have used methodology of previous fricative studies, begun with the EC SCIENCE "Fricative" Project, conducted by Shadle, Scully, Guérin and others at ICP, Grenoble. That study was focussed on characterizing fricatives in general.
- Here, we've adapted that methodology to focus on Portuguese fricatives in particular, and thus use real words and phonology of Portuguese.

Recording Method

Our subject was a male adult Portuguese native speaker.

The Corpus is divided in four parts:

Corpus 1, which I'll call the word corpus, contains 154 words, containing all Portuguese fricatives /f, v, s, z, ʃ, ʒ/ in combination with the Portuguese non-nasal vowels /i, ɨ, e, ɛ, ɐ, a, ɔ, o, u/. The carrier sentence "Diga ..., por favor." was used to record the words in a balanced phonetic context and with a neutral prosody.

Although only fricative /ʃ/ can occur word-finally, we have elicited all the other fricatives word-finally by using Portuguese words with a reduced /ɨ/ as final phoneme.

Corpus 2, the sentence corpus, is made up of 12 sentences that use 61 of the 154 words in **Corpus 1**.

Corpus 3, the nonsense-word corpus, has 54 nonsense words that follow the pattern /pV$_1$FV$_2$/, where V$_i$ is one of the vowels /ɨ, ɐ, u/. These nonsense words follow Portuguese phonological rules.

Corpus 4, the sustained corpus, consists of sustained fricatives at 3 different effort levels.

The acoustic and laryngograph signals were recorded onto a DAT tape at 16 bits and a sampling frequency of 48 kHz, in a sound treated booth.

Temporal Analysis Method

- We started by identifying the VF, F and FV boundaries using the time waveforms.
- During this phase we noticed numerous devoiced examples, which led us to study this phenomena in detail, and classify devoicing using a new criterion based on both the acoustic and laryngograph signals.
- Previous studies have only used one of the two signals.
- We have devised empirical criteria for devoicing:

  devoiced is defined as no periodic structure in the acoustic and/or laryngograph signals;

  partially devoiced – a few steady cycles;

  voiced – steady cycles, even if the amplitude is much lower than in the vowel.
- We have also used a quantitative criterion for devoicing: ...

Quantitative Criterion for Devoicing

We started by computing the variance of the laryngograph signal in the VF transition, and during the fricative.

Then, the ratio of variances was calculated. That is, the variance in the VF transition divided by the variance during the fricative.

Obviously the ratio gets bigger if the laryngograph signal during the fricative gets really small relative to the transition – a good threshold was 15. Bigger than that, the fricative is labelled *devoiced*; less than 15, *voiced*.

Time-averaged power spectrum (Corpus 1, 2 and 4)

For the spectral analysis of fricatives from Corpus 1, 2 and 4, we have used *time-averaged power spectra.*

We computed the power spectra for 9 ms Hamming windows, located as you see here, and averaged at each frequency. Shorter fricatives mean more overlap, but for the word and sentence corpus we have single tokens and no choice.

Ensemble-averaged power spectrum (Corpus 3)

For Corpus 3, the nonsense words, we have many tokens, so we can do ensemble-averaging.

The *ensemble-averaged power spectra* was calculated using 18 ms Hamming windows placed at the beginning, middle or end of the fricative.

Temporal Analysis Results – Overall Devoicing (Corpus 1, 2 and 3)

Overall results from the analysis of devoicing in Corpus 1, 2 and 3, using the empirical criteria, show that more than 50% of the fricatives devoice.

Devoicing by Corpus

Devoicing rate by Corpus differs between the three fricatives, and between Corpus 1, 2 and 3. There is no apparent pattern.

Devoicing by Position in Word (Corpus 1)

If we plot the percentage of devoicing by position in words from Corpus 1, there's a significant increase in devoiced examples from word-initial, through word-medial to word-final positions.

The *empirical* and *quantitative* criteria for devoicing were similar; over 80% of the fricatives were classified the same way by both methods.

Temporal Analysis Results (Duration – Corpus 1)

Here are some duration results. We looked at duration by fricative but wanted to make sure any differences weren't simply due to segmentation decisions.

- The mean duration of the unvoiced fricatives is always greater than the mean duration of the voiced fricatives.
- There is no significant difference by place of articulation.
- The mean duration of the VF transition is greater than the mean duration of the FV transition.
- The average durations of the VF and FV transitions of unvoiced fricatives (shown on the left column) are quite similar to those of their voiced counterparts (shown on the right column). So the unvoiced-voiced difference appears to be real, not just an artifact of segmentation.
- For fricative /s/ in initial position, as the following vowel's place of articulation moves further back, the duration of the fricative diminishes.

Spectral Analysis Results (Corpus 4)

Spectral analysis of sustained fricatives shows substantial differences of voiced and unvoiced, same-place fricatives.

Not only are the voiced spectra lower in amplitude, as expected, but differences in spectral shape occur, particularly for /ʃ–ʒ/.

Spectral Analysis Results (Corpus 4)

- We can also see that the amplitude increases with effort level.
- The amplitude difference between the three effort levels is greater at high frequencies.
- This amplitude difference at high frequencies is larger for unvoiced fricatives than for their voiced counterparts.
- These differences are associated with source type and strength, and are similar to results for American English and French subjects.

Spectral Analysis Results (Corpus 1 and 3)

We were also interested in the following questions:

Does effort level of sustained fricatives correspond with stress of the syllable containing the fricative in the word and nonsense-word Corpora?

Does it correspond to position within the word in Corpus 1?

Spectral Analysis Results (Corpus 1 and 3)

This graph shows the averaged power spectra of the fricative [ʃ] in the word [kɐˈpuʃ] from Corpus 1 (solid line), and the same fricative [ʃ] in the nonsense word [puʃu] from Corpus 3 (dash-dotted line).

We have same fricative, same vowel context, but a difference in syllable stress and position in the word.

You can see here a difference in spectral amplitude, on the order of 5-10 dB from 2 to 20 kHz.

We have looked at seven such pairs-by-vowel-context, and focussed our attention on the differences in the spectral amplitudes, particularly above 2 kHz.

Spectral Amplitude Comparisons

- Four of the seven such pairs-by-vowel-context showed the same pattern, as I just showed you, with the Corpus 3 spectral amplitude consistently higher (by 5-10 dB) than that of Corpus 1 for frequencies above 2 kHz.
- This amplitude difference across the frequency range corresponds strongly to a difference in stress between the two fricatives.
- The spectral shapes and amplitudes are similar to the soft effort level for all destressed /ʃ/ fricatives, and to the medium effort level for stressed /ʃ/ fricatives.
- No fricatives from Corpus 1 or 3 resemble their high-effort-level Corpus 4 counterparts.
- The main peak for /uʃu/ context was at a significantly lower frequency than in, e.g., /i̟ʃi/ context; this is as expected from previous work (Shadle and Scully 1995).

Conclusions

In this talk I've described a Corpus for studying Portuguese fricatives. Our conclusions, based on so far one native speaker, are:

- Voiced fricatives devoice in over one half the cases in both nonsense and real words.
- Spectral analysis revealed a correspondence between the effect of effort level and of syllable stress, and showed some effect of vowel context.
- Corpus 3 is better controlled and easier to analyse than Corpus 1 or 2; validating its use would give an important advantage. But it is important to retain the real word and sentence corpora for our ultimate goals.
- Future/ongoing: analysis of other Portuguese speakers.