Adding Fricatives to the Portuguese Articulatory Synthesizer

António Teixeira

Luís Jesus

Roberto Martínez

Universidade de Aveiro - Portugal

signal processing laboratory

ieeta instituto de engenharia electrónica e telemática de aveiro

Overview

- First attempts at incorporating models of frication into an articulatory synthesizer, with a modular and flexible design, are presented
  - Fricative Modeling
  - SAPWindows
  - Acoustic Model and Noise Sources
  - Results
  - Conclusions

Articulatory Synthesis ?!

"Speech synthesis has achieved a high level of quality and usefulness, but the flexibility of diphone and other acoustically based schemes is limited.

Articulatory synthesis holds promise for overcoming some of the limitations and for sharpening our understanding of the production/perception link."

D. H. Whalen, "articulatory synthesis: advances and prospects**", ICPhS,** Barcelona**, 2003.**

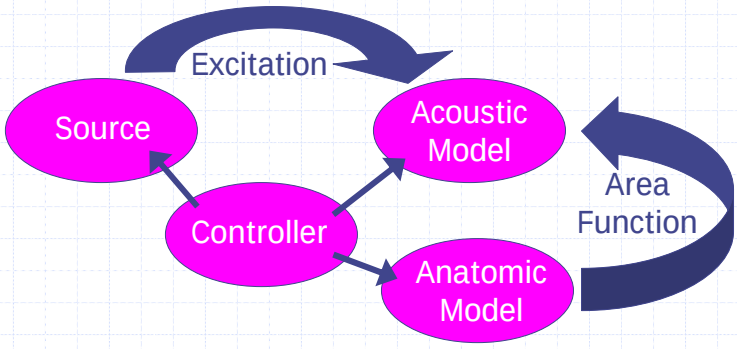
Fricative Production Mechanisms and Models

- The acoustic mechanism for production of fricatives is not as well understood as for vowels.
- These difficulties have been reflected in the relatively poor quality of fricative and affricate synthesis.
- Sondhi and Schroeter (1987) generated frication using only one series pressure source at the point of maximum constriction, or alternatively using a parallel volume velocity source downstream of the main constriction.
- Narayanan and Alwan (2000)'s hybrid source model used a combination of acoustic monopole and dipole sources and a voiced source in the case of voiced fricatives.

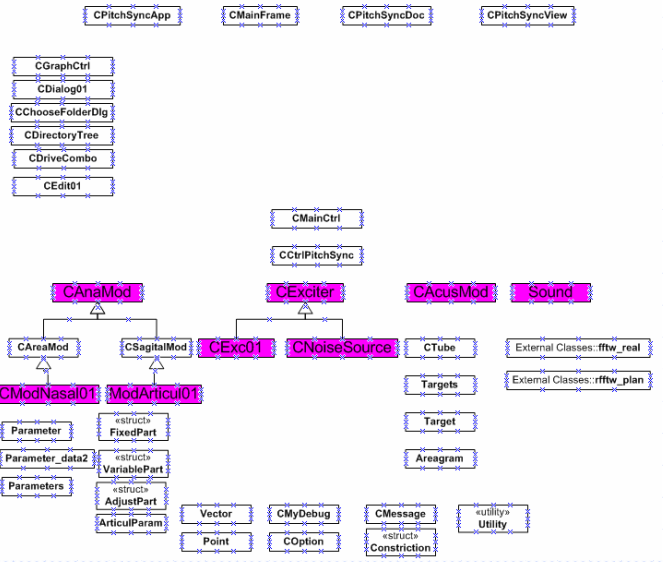
SAPWindows

- SAPWindows is the name given to the university of Aveiro's articulatory synthesizer
  - It stands for "**s**intetizador **a**rticulatório de **P**ortuguês" for **windows**
- It consists of articulatory, source, and acoustic models
- Different sounds are produced when the acoustic model is excited by various sources

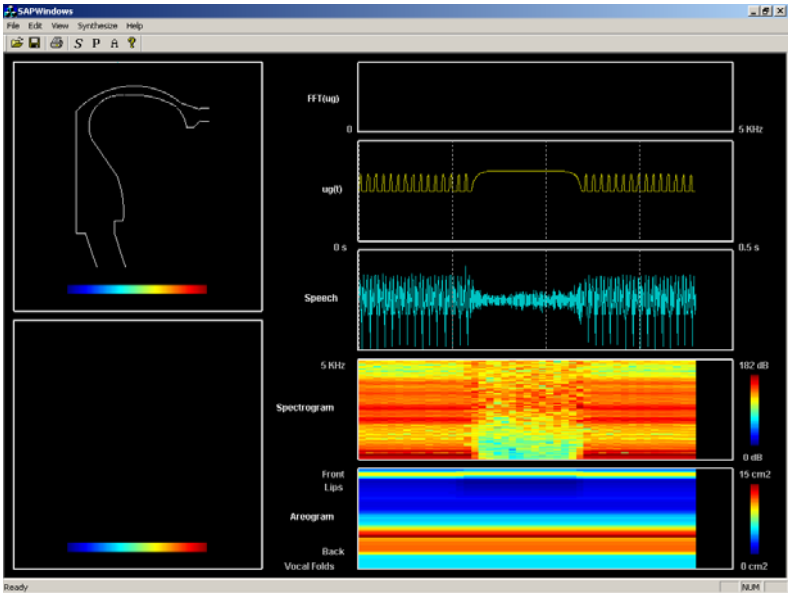
Main synthesis base classes



More detail ...



SAPWindows GUI



Vowel Modeling

- SAPWindows was used in previous experiments (Teixeira et al. 2002) to synthesize voiced sounds.
- The value of F_O was directly controlled by changing the parameterized glottal area of a two-mass vocal fold model.

Fricative Modeling

NEW

- A model of frication was added to the synthesizer **maintaining most of the existing modules,** control processes and parameters
- Noise sources are **part of** the acoustic module
  - Turbulence generated inside the tract

Dipole Sources

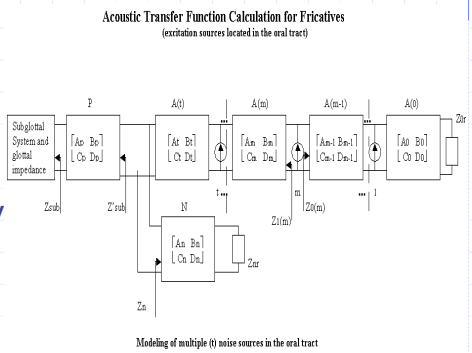
- Since **dipole sources** have been shown to be the most influential in the fricative spectra (Narayanan and Alwan 2000)
- The noise source of the fricatives has only been approximated by equivalent pressure voltage (dipole) sources in the transmission-line model

Monopoles

- Nevertheless, it is also possible to insert the appropriate **monopole sources**, which contribute to the low - frequency amplitude and can be modeled by an equivalent current volume velocity source.

Model adopted

- Frication noise is generated at the vocal tract according to the suggestions of Flanagan (1972), and Sondhi and Schroeter (1987).
- A noise source can be introduced automatically at any t-section of the vocal tract network, between the velum and the lips.



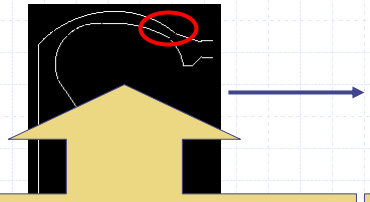
Noise sources activation

1. Articulatory Model

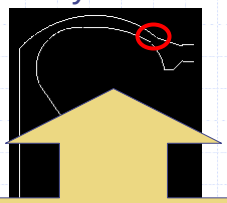
Checks Constrictions

2. Acoustic Model

Checks Reynolds Number



The synthesizer's articulatory module registers which vocal tract tube cross sectional areas are bellow a certain threshold ($A < 1\ cm^2$), producing a list of tube sections that might be part of an oral constriction that generates turbulence.



The acoustic module calculates the Reynolds number at the sections selected by the articulatory module and activates noise sources at tube sections where the Reynolds number is above a critical value.

Noise Sources

- Noise sources can also be inserted at any location in the vocal tract
  - Based on additional information about the distribution and characteristics of sources (Shadle 1990; Narayanan and Alwan 2000)
- This is a different source placement strategy from that usually used in articulatory synthesis where the sources are primarily located in the vicinity of the constriction

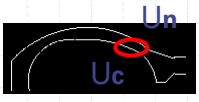
Noise Sources

- The distributed nature of some noise sources can be modeled by inserting several sources located in consecutive vocal tract sections.
- This will allow us to try combinations of the canonical source types (monopole, dipole and quadrupole).

Noise sources amplitude

- Pressure sources have an amplitude proportional to the squared Reynolds number (if activated)
  - $P_n = 2\times10^{-6} \times \text{random}(re^2 - re^2_{crit}),\ re > Re_{crit}$
  - $P_n = 0,\ re \leq Re_{crit}$
- Based in (Flanagan 1972; Sondhi and Schroeter 1987)

Equivalent parallel noise source flow Un

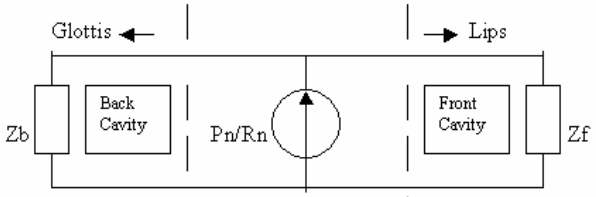


Ug

Un = Pn/Rn,

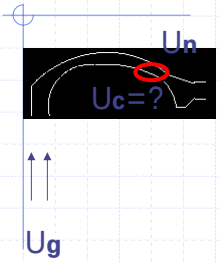
Rn is the noise resistance

Rn = ρ|Uc|/(2*Ac*Ac) , where Uc is low pass filtered, cutoff freq=2000 Hz



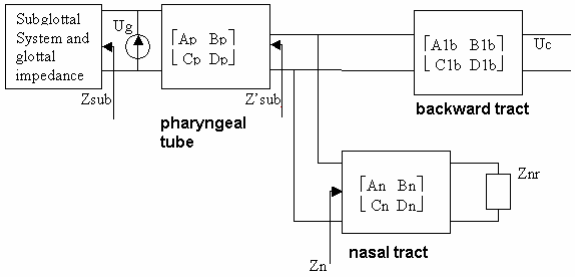
Sondhi and Schroeter (1986-1987) parallel flow source

Flow at the constriction Uc

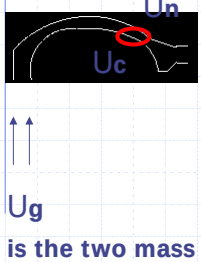


Uc = Ug*IFFT(Hgn)

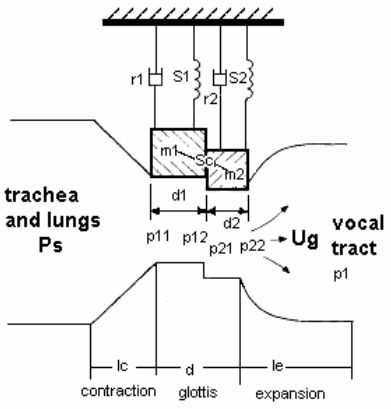
Hgn is the transfer function from the glottis to the constriction



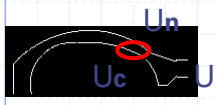
Flow at the glottis Ug



is the two mass model of Ishizaka and Flanagan 1972, 1977



Flow at the lips (Ulips)



Ug

- Psound ~ dU_{lips} / dt
- Each noise source “m” contribute independently
- $H_{nl(m)}$ is the transfer function from the noise source “m” to the lips

$$Psound \sim \frac{d}{dt}\left[U_g * IFFT(H_{g1}) + \sum_{m} (U_m * IFFT(H_{nl}^{m})) \right]$$

glottal constribution

constribution from noise sources

Results

Static

And

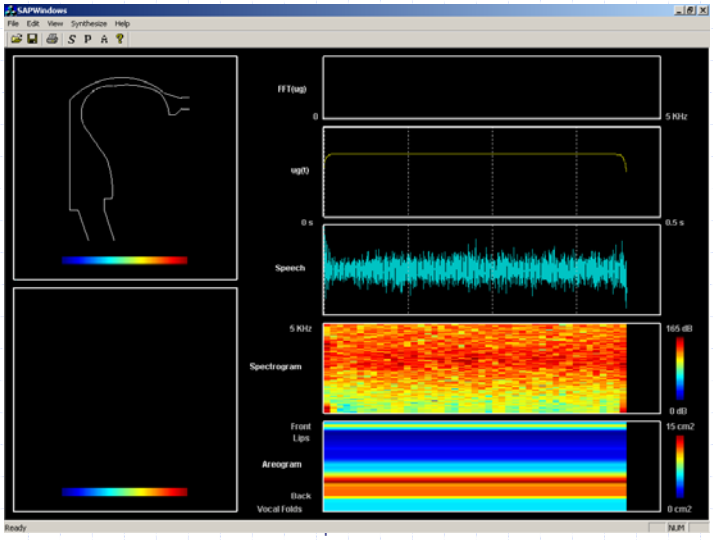
VCV sequences

Experiment I

- In a **first** experiment the synthesizer was used to produce **sustained unvoiced** fricatives.
  - The vocal **tract configuration** derived from a natural high vowel was adjusted by raising the tongue tip in order to produce a sequence of reduced vocal tract cross - sectional areas.
  - The **lung pressure** was linearly increased and decreased at the beginning and end of the utterance, to produce a gradual onset and offset of the glottal flow.
  - The synthesizer **activated two sources** during the steady state of the fricative.

Results: static configuration

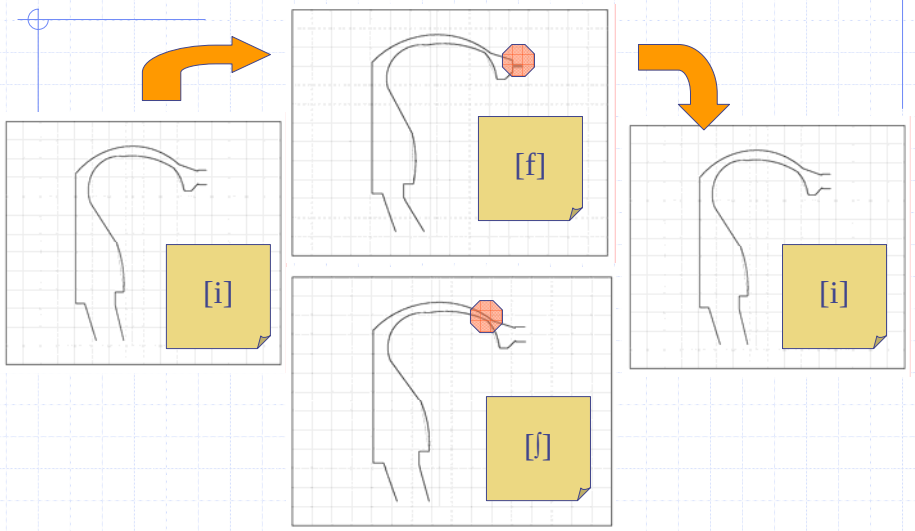
Glottis volume velocity waveform, speech waveform and spectrogram of a synthesized unvoiced fricative /ʃ/.



Experiment II

- In a **second** experiment we synthesized fricatives in **VFV sequences**.
  - Articulatory configurations for vowels obtained by an inversion method were used.
  - During the fricative interval the essentially the tongue tip articulatory parameter was adjusted to the fricative configuration.
  - A F_0 value of 100hz and a maximum glottal opening (Agmax) of 0.3 cm^2 were used to synthesize the vowels.

Tract configurations

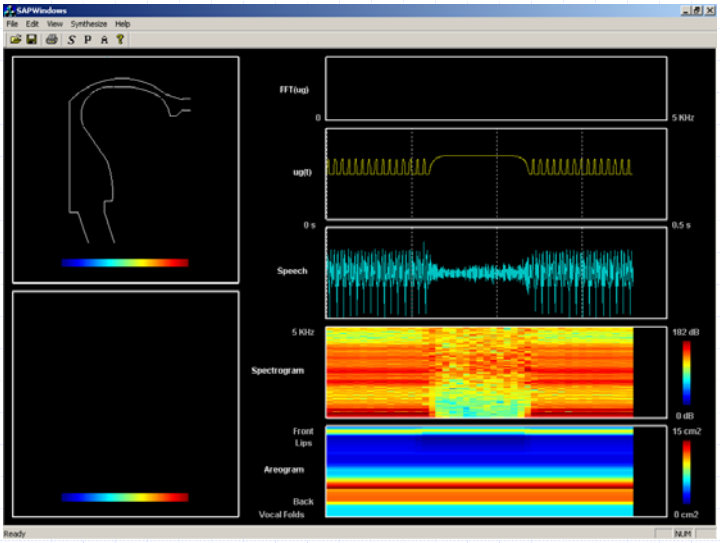


- During the fricative interval the time trajectory of Agmax
  - Starts at 1.5 cm^2
  - Rises to 2 cm^2 at the fricative middle point

  And
  - Returns to 1.5 cm^2 near the end, before assuming the value used during vowel production

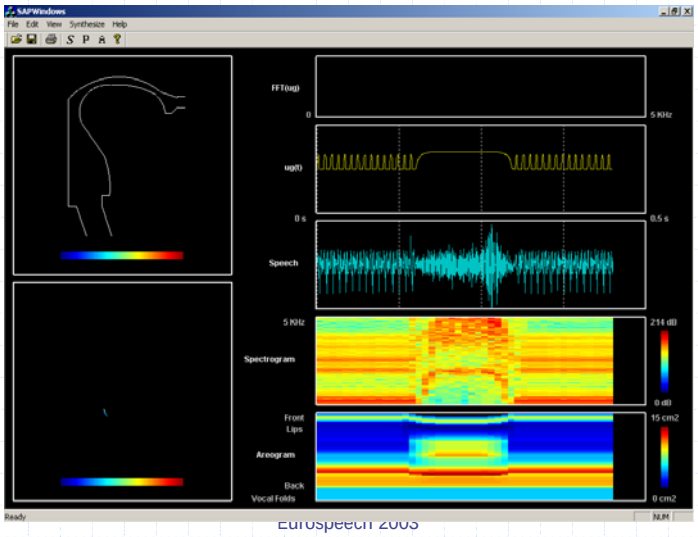
Results /iʃi/

Glottis volume velocity waveform, speech waveform and spectrogram of a synthesized unvoiced fricative VFV /iʃi/.



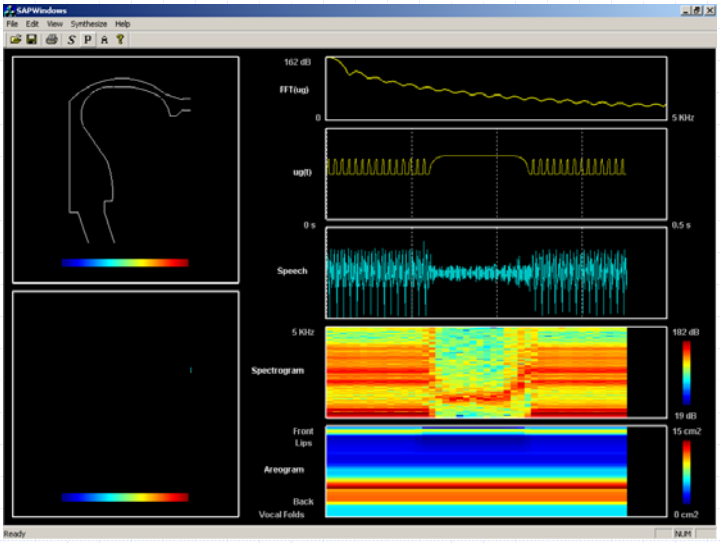
Results /isi/

Glottis volume velocity waveform, speech waveform and spectrogram of a synthesized unvoiced fricative VFV /isi/.



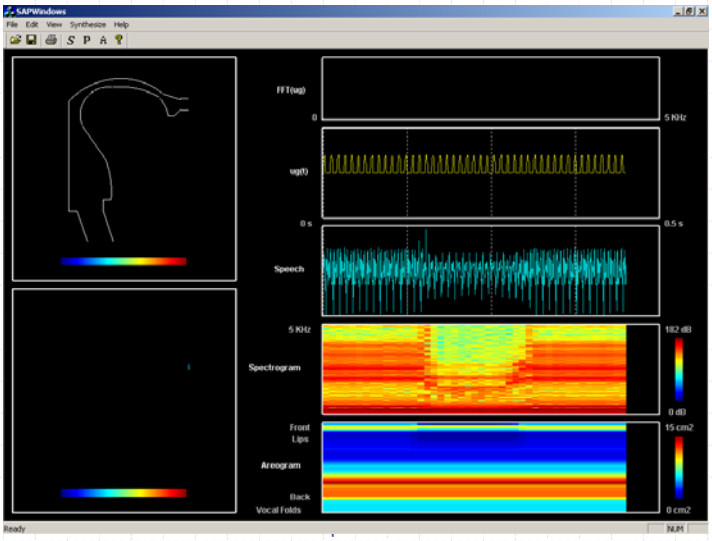
Results - /ifi/

Glottis volume velocity waveform, speech waveform and spectrogram of a synthesized unvoiced fricative VFV /ifi/.



Results /ivi/

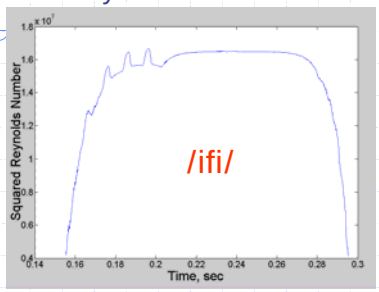
Glottis volume velocity waveform, speech waveform and spectrogram of a synthesized voiced fricative VFV /ivi/.

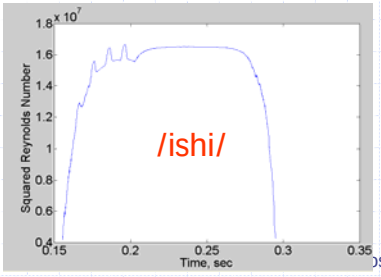


31

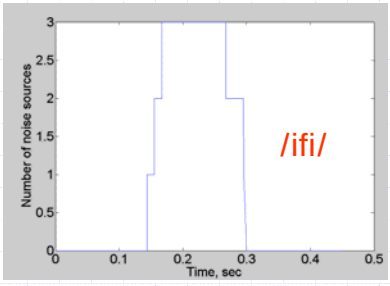
For synthesized VFV /ishi/ and /ifi/

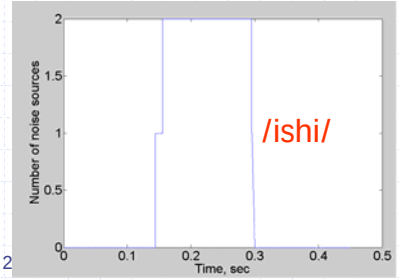
Reynolds number





Number of noise sources





ospeech 2

Conclusions

- With the addition of noise source models and modifications to the acoustic model, our articulatory synthesizer is capable of producing sustained fricatives and fricatives in VCV sequences.
- First results were judged in informal listening tests as being highly intelligible.
- Further validation and checking of the models is still required.
- Nevertheless, this is an important new step towards a complete articulatory synthesizer for Portuguese.

Conclusions

- SAPWindows is a valuable tool for trying out new or improved source models, and running production and perceptual studies of European Portuguese fricatives.
- The possibility of automatically inserting and removing noise sources along the oral tract is a feature we regard as having great potential.

THANK YOU FOR YOUR ATTENTION

António Teixeira, Luís Jesus and Roberto Martínez 1/9/2003