Automatic identification of vocalic intervals in speech signal

Jesus Garcia
Antonio Galves

Phonological Support:

Flaviane Fernandes
Janaisa Viscardi
Ulrike Gut

Many thanks to:

Dafydd Gibbon
Emmanuel Dupoux
Franck Ramus
Problem

We have:

\[
\begin{align*}
0.179 & \quad 0.301 & \quad v \\
0.301 & \quad 0.390 & \quad c \\
0.390 & \quad 0.440 & \quad v \\
0.440 & \quad 0.498 & \quad c \\
0.498 & \quad \ldots \\
\end{align*}
\]

and we want:

\[
\begin{align*}
0.179 & \quad 0.301 & \quad v \\
0.301 & \quad 0.390 & \quad c \\
0.390 & \quad 0.440 & \quad v \\
0.440 & \quad 0.498 & \quad c \\
0.498 & \quad \ldots \\
\end{align*}
\]
Identification

Cues to the segmentation:

- acoustic signal plot
- spectrogram plot
- listening
Example:

“A autoridade do governador diminuiu”

Acoustic Signal

Spectrogram
vowels are recurrent
“t” signal:
\( c(i,t) = \text{energy for the frequency } i \text{ in time } t \)

\[
p \Phi_i, t \equiv c \Phi_i, t \prod_{k=1}^{N} \bar{a} \quad c \Phi_k, t \prod_{k=1}^{N} \bar{a}
\]
Relative entropy of the column $t$ respect to the column $s$.

$$h_{\Phi,s} = \sum_i \ln \frac{p_{\Phi,t}(i)}{p_{\Phi,s}(i)}$$
we use \( h(t-1,t) \), \( h(t-2,t) \) e \( h(t-3,t) \).

\[
h(t) = h(t-1,t) + h(t-2,t) + h(t-3,t)
\]

plot for \( h(t) \):
Distance between consecutives columns
\[ d \Phi, t A 1 \mathcal{E} = \sqrt{\mathcal{a}} \Phi P \Phi, t \mathcal{B} p \Phi, t A 1 \mathcal{E}^2 \]
Spectrogram for the sentence:
'A hurricane was announced this afternoon on the TV.'
Signal

Relative Entropy
800-3000 hz

Relative Entropy
0-800 hz
Spectrogram and segmentation for
'A hurricane was announced this afternoon on the TV.'
Box Plots for the integral of the relative entropy for each language
Spectrogram and R. entropy for the sentence with the lower integral
Ramus classification (1999).

\[ \text{%V} \times \Delta C \]

\[ \text{%V} = \text{vowel percent} \]

\[ \Delta C = \text{standard deviation of consonantal intervals} \]