Accurate detection and segmentation of plosives

DyNaVoiceR project meeting

November 9th, 2019

João Pereira da Silva

joaomiguelppsilva@gmail.com

FEUP, UA, FMUP

Cofinanciado por:

COMPETE 2020 PROGRAMA OPERACIONAL COMPETITIVIDADE E INTERNACIONALIZAÇÃO

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

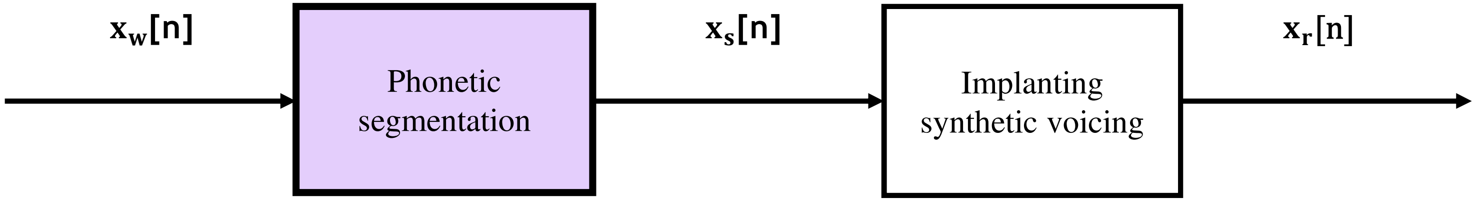
Outline

- Challenges
- Importance of silence
- Plosive detection
- Conclusion and future work

Challenges: an overview

- **Phonetic segmentation:** how to identify the regions in whispered speech that should be converted to voiced regions.
  - Detect plosives (aka stop consonants)

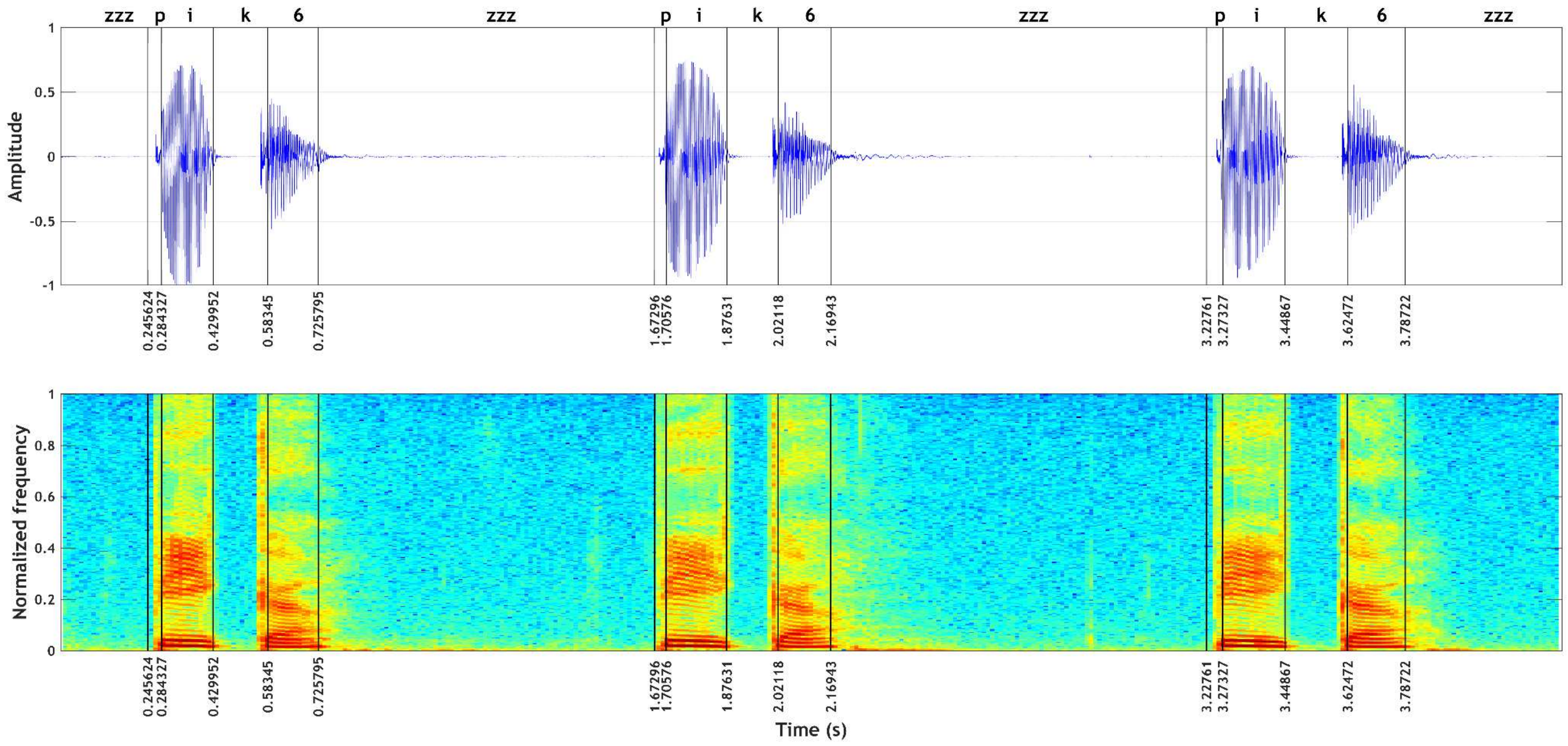
- **Implanting synthetic voicing:** **i)** how to convert whispered regions into voiced regions while preserving and at the same time enhancing the linguistic message; and **ii)** how to convey elements of the acoustic signature of the speaker.



- $x_w[n]$ is the whispered speech;
- $x_s[n]$ is the segmented signal;
- $x_r[n]$ is the desired reconstructed voice.

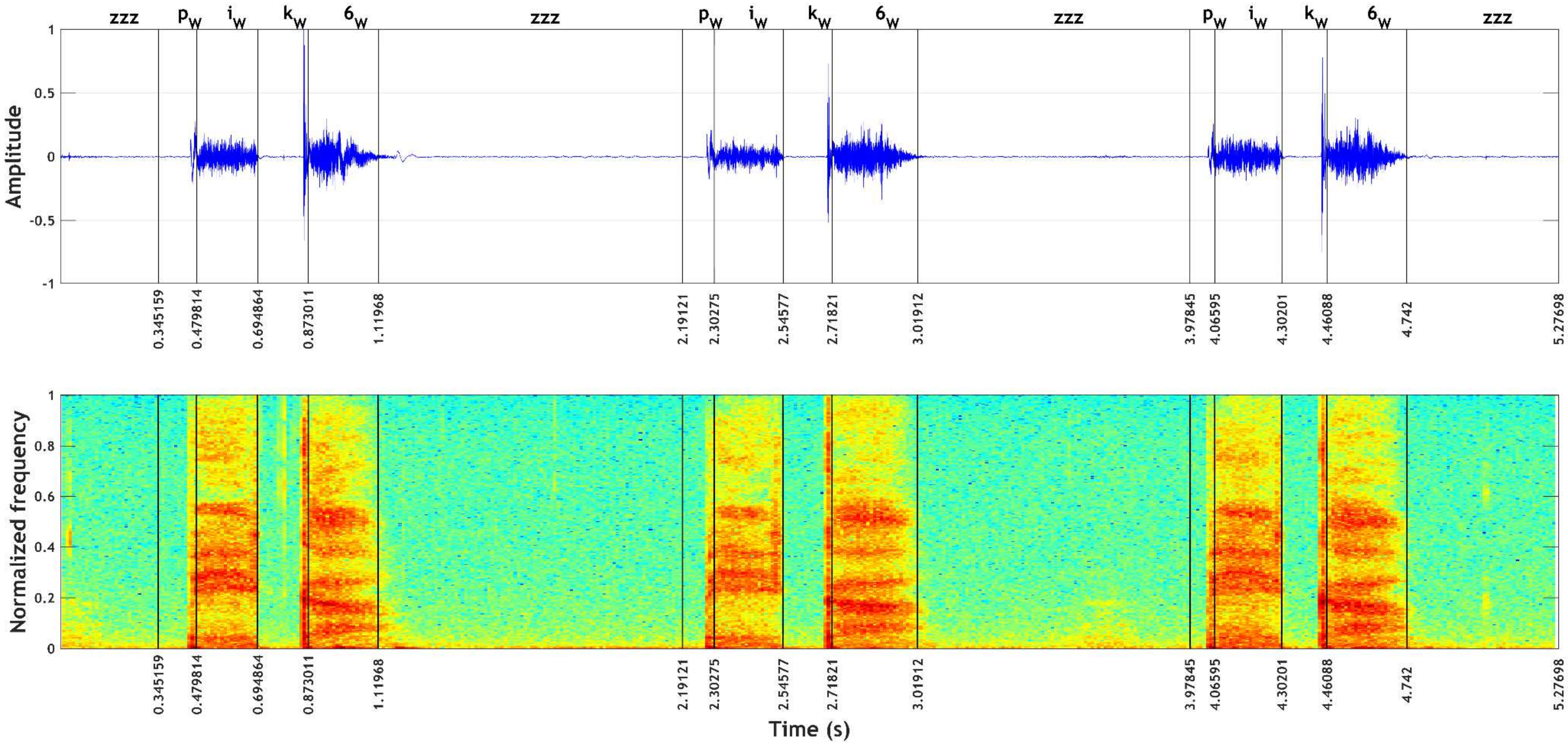
Challenges: natural speech

Illustrative example of normal speech using the Portuguese word “*pica*”



Challenges: whispered speech

Illustrative example of whispered speech using the Portuguese word “*pica*”.



Importance of silence

Interesting experiences with silence

Word 1

Word 2

Word 3

Importance of silence

Interesting experiences with silence

Ri/p/a -> Ri/b/a 🔊

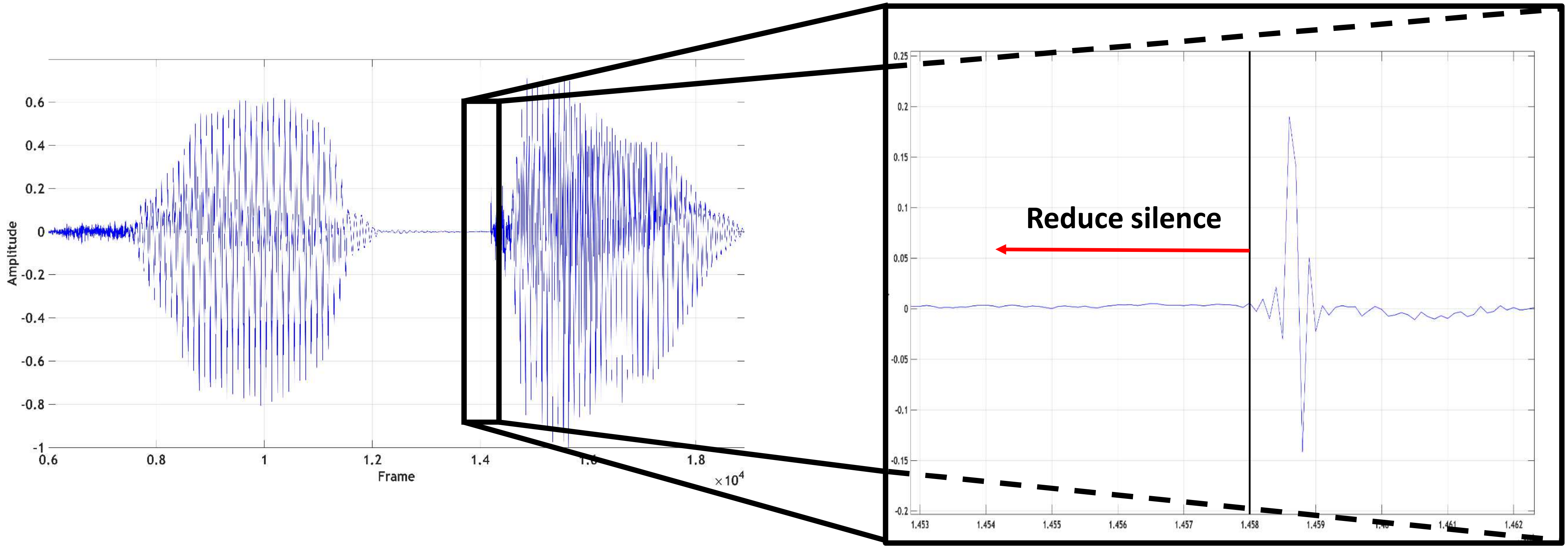
Ri/t/a -> Ri/d/a 🔊

Pi/c/a -> Pi/g/a 🔊

Importance of silence

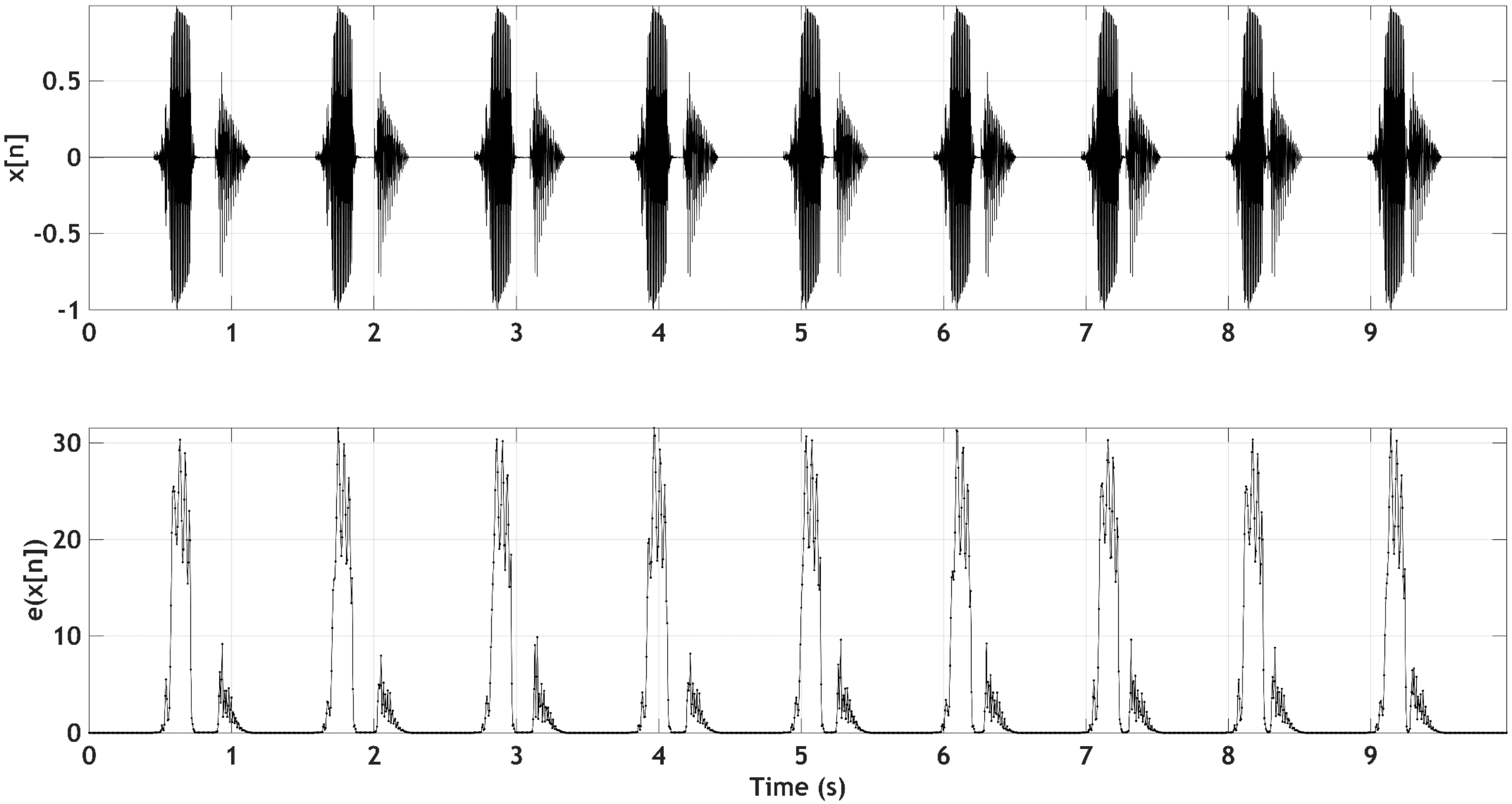
Procedure: Decreasing silence between the moment right before the plosive explosion and previous phoneme (vowel).

Illustrative example using the Portuguese word “***ripa***”, which becomes “***riba***” (word 1).



Importance of silence

Illustrative example using the Portuguese word “***ripa***”, which becomes “***riba***” (word 1).



Importance of silence: summary

[/p/, /t/, /k/] becomes [/b/, /d/, /g/], respectively.

14) nu**c**a -> nu**g**a

15) lu**p**a -> lu**b**a

19) ri**p**a -> ri**b**a

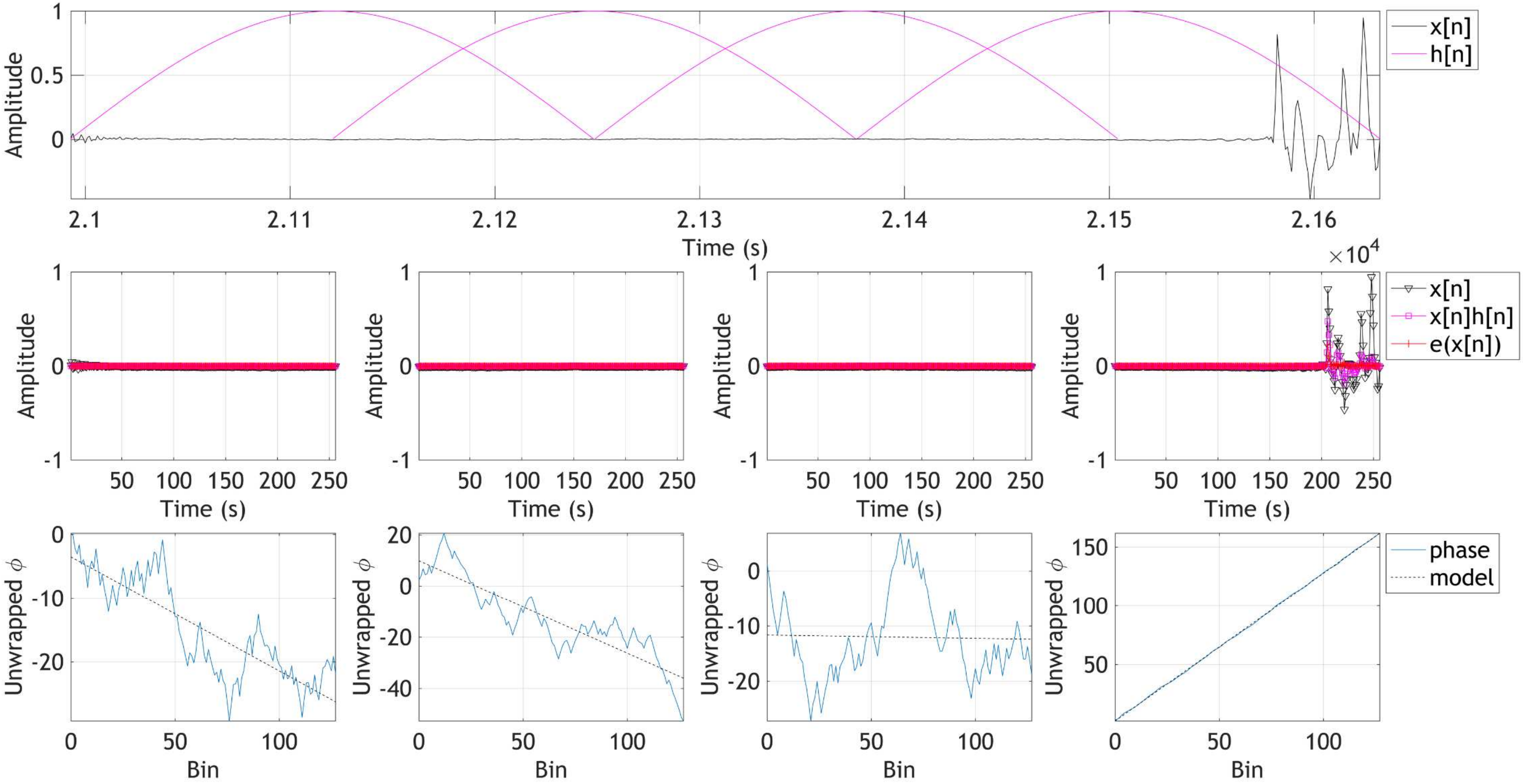
22) pi**c**a -> pi**g**a

31) lu**t**a -> lu**d**a

35) ri**t**a -> ri**d**a

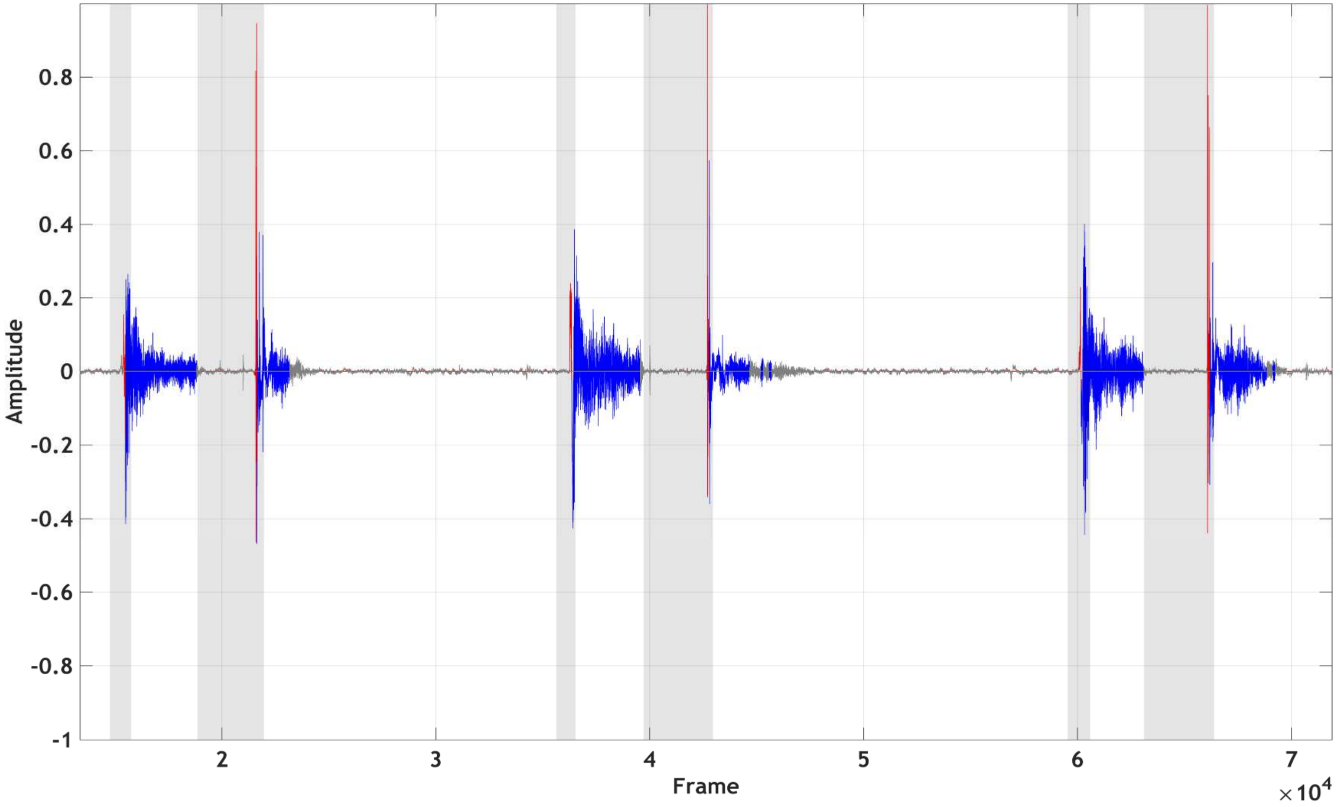
Plosives detection: framework

N=256, 50% overlap



Plosives detection: rules

1. Silence: $e[k-1]\ \&\ e[k-2]\ \&\ e[k-3] < threshold_1$
2. Energy: $e[k] > threshold_2$
3. ODFT phase: $phase[k]_{model} - phase[k] < threshold_3$

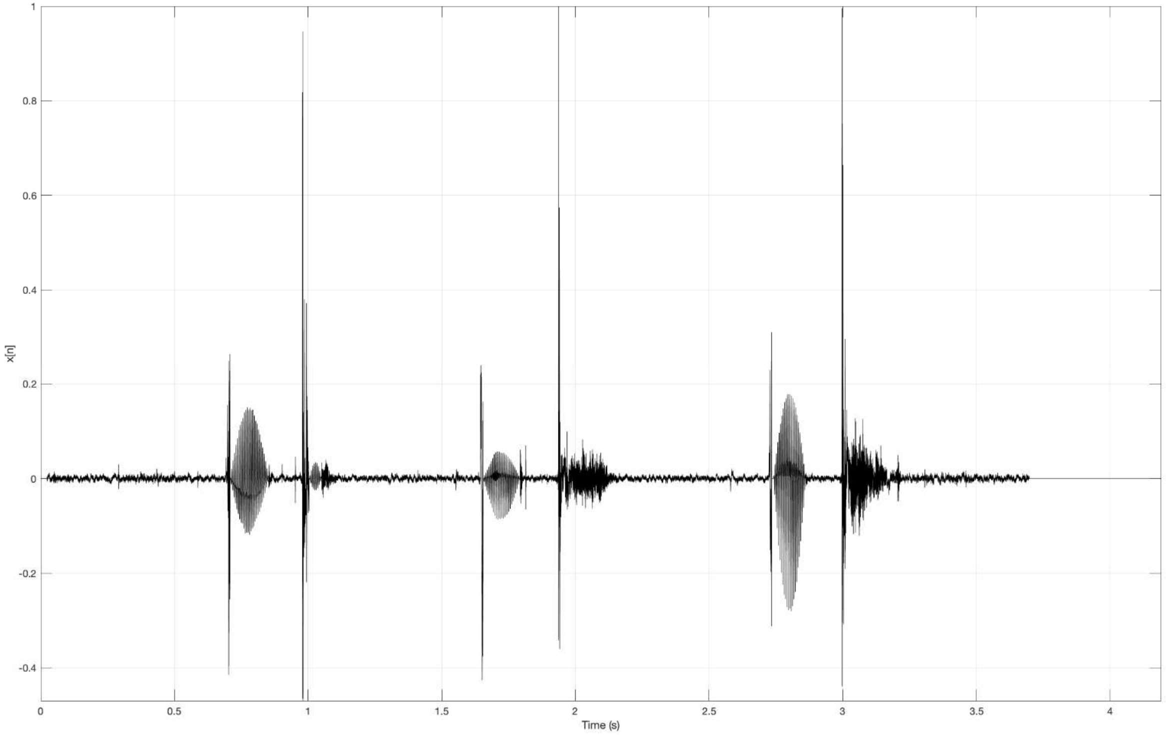


- Red-plosives
- Gray-silence
- Shadow-ground truth
- Blue-the rest (vowels)

Segmentation and “implantation”

Whispered version

New version



Conclusion

- Silence right before the plosive can change its meaning
- Right before a plosive there is always silence
- Plosives have a peak of energy
- Plosive are impulse-like signal: phase structure close to a model
- The combination of 3 rules improves the plosives detection
- Objective evaluation: plosives are within the ground truth area
- Subjective evaluation: simple implantation works

Future work

- Validate algorithm using all words in the database
- Analyze the performance of this algorithm using words without plosives
- Fricatives
- Input to HMM (emissions)

End

Q&A