# Studying vowel variation in French-Algerian Arabic code-switched speech

*J. Wottawa, D. Amazouz, M. Adda-Decker, L. Lamel*

LIMSI, CNRS, Université Paris-Saclay, France
LPP, UMR 7018 CNRS - U. Paris 3 / Sorbonne Nouvelle, France





# Introduction




- Bilingual speakers' speech presents more acoustic variation than that of monolinguals [Bullock, 2012, Auer, 2013]
- Bilinguals access more than one phonemic inventory which may lead to potential interferences [Fricke et al., 2016, Grosjean, 1995]
- Focus on French-Algerian code-switched speech in highly proficient bilinguals
- Vowel inventories of different sizes (French is richer)
- Research question: To what extent do bilinguals adapt their vowel productions to the linguistic context?

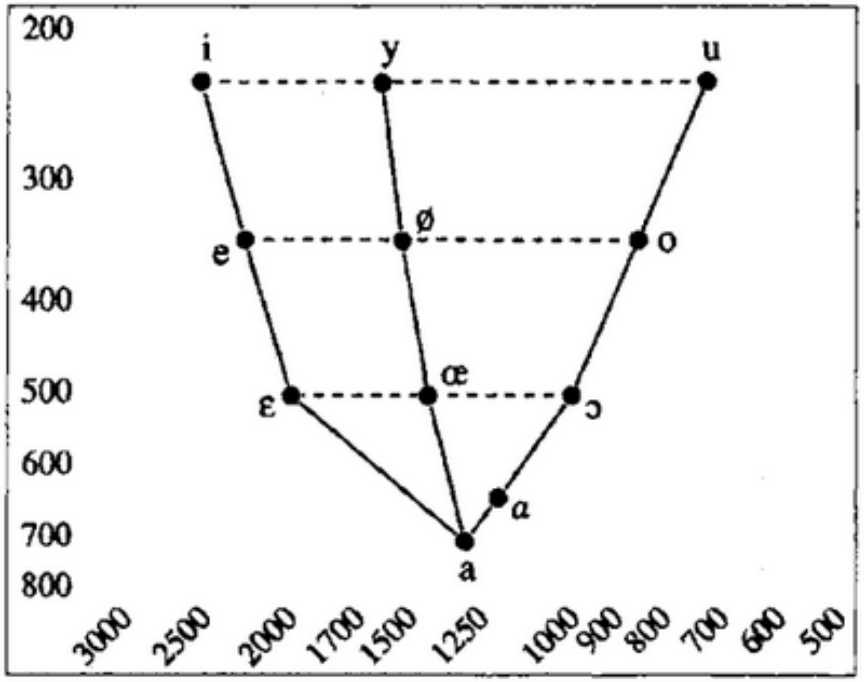
# Introduction



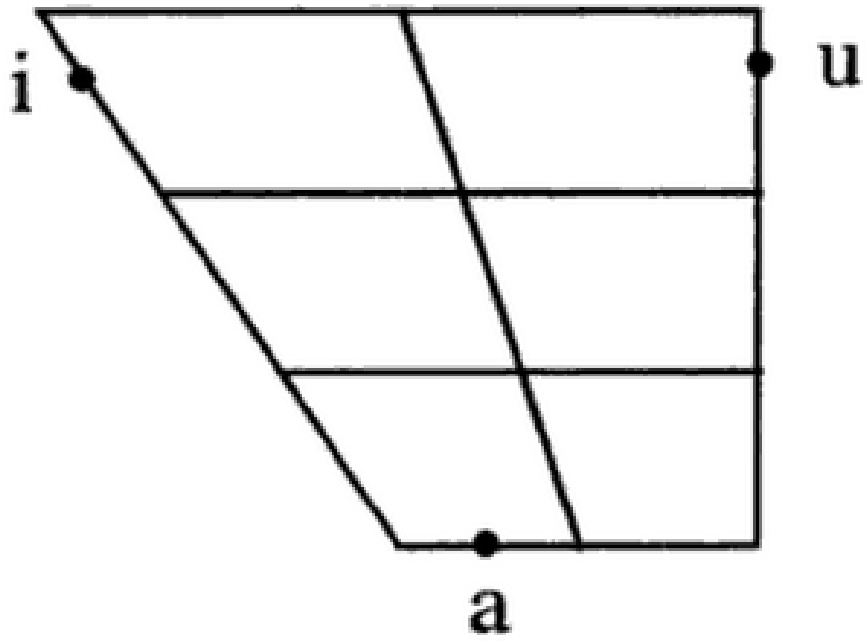

- Methodology: use automatic speech alignment to study vowel variants
- Focus on parallel variants (only substitutions, no deletions or insertions)
- Frequent replacements of the target vowel by competing vowels are considered an indicator of variation
- Three experiments targeting vowel variation:
  - Experiment 1: Vowel variants in French (French natives vs. bilinguals)
  - Experiment 2: Vowel variants in code-switched speech (bilinguals French vs. Arabic)
  - Experiment 3: Vowel centralization in French and Arabic (French natives and bilinguals)

# Outline

# Vowels in French and Arabic




(Delattre, 1966)




(Thelwall & Akram Sa'adeddin, 1999)

- Standard French: 11 oral vowels, 4 nasal vowels, 1 schwa
- Classic Arabic: 3 oral vowels
- How does this difference influence speech production in bilingual speakers?

# Speech material

- Conversational speech
- Languages: French & Arabic
- FACST-corpus [Amazouz et al., 2018]
  - French & Algerian Arabic from 20 bilingual speakers (10 female)
  - 7.5 hours of monolingual read and stimulated spontaneous speech with code-switching
  - Study based on approx. 3 hours of speech from 11 speakers
- NCCFr [Torreira et al., 2010]
  - 36h of conversational French, 46 speakers (24 female)

# Methodology

## Automatic alignment

- Forced automatic alignment with pronunciation variants
- Position-independent monophone French acoustic model
- Parallel variants (only substitutions, no deletions, no insertions)
- Productions in code-switched conversational speech

# Methodology

## Automatic alignment

- Forced automatic alignment with pronunciation variants
- Position-independent monophone French acoustic model
- Parallel variants (only substitutions, no deletions, no insertions)
- Productions in code-switched conversational speech

## Production variation quantification

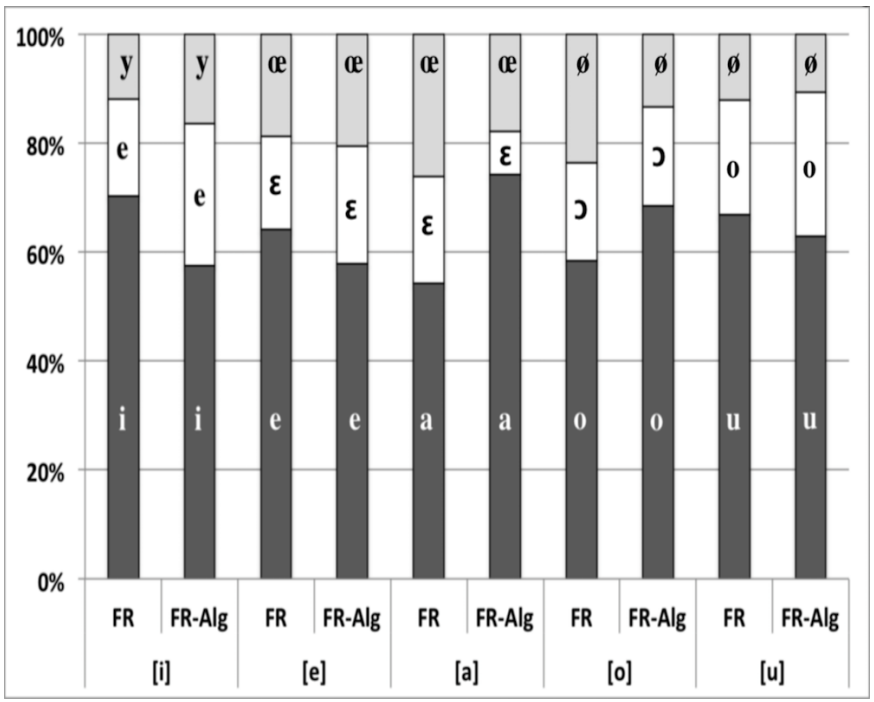
- Quantified for each target vowel
- Count occurrences of competing vowels
  (selected during automatic alignment)

# Exp. 1: Vowel variants in French

- Populations: French natives vs. bilinguals (French-Algerian Arabic)
- Language: French
- Production variation for five target vowels
- Two production variants for each target vowel
- Specific pronunciation dictionary for each condition

| Vowel | Variants | Examples |
|---|---|---|
| i | [e, y] | *lit* (bed): li, le, ly |
| e | [ɛ, œ] | *nez* (nose): ne, nɛ nœ |
| a | [ɛ, œ] | *chat* (cat): ʃa, ʃɛ, ʃœ (anterior) |
| a | [ɔ, œ] | ʃa, ʃɔ, ʃœ (posterior) |
| o | [ɔ, ø] | *chaud* (hot): ʃo, ʃɔ, ʃø |
| u | [o, ø] | *loup* (wolf): lu, lo, lø |

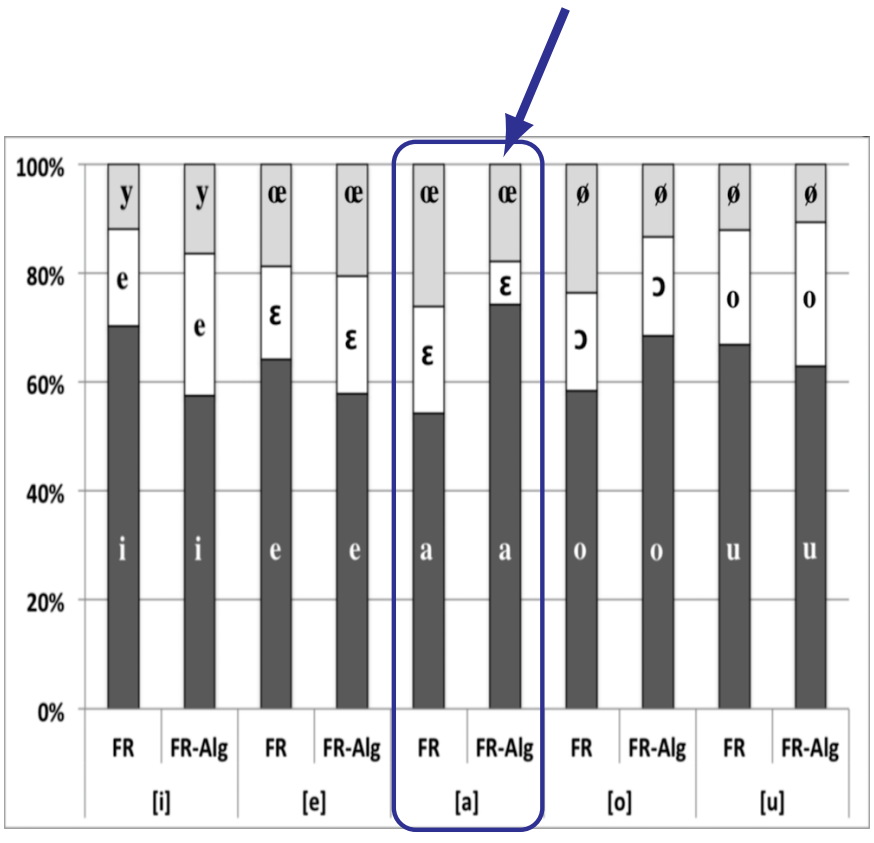
# Exp. 1 - Results (French: natives vs. bilinguals)




- Observed variation is vowel independent
- Comparable amount of variation in both groups (French natives, bilinguals)

- The plot does not show the posterior variants for [a]

# Exp. 1 - Results (French: natives vs. bilinguals)




- Observed variation is vowel independent
- Comparable amount of variation in both groups (French natives, bilinguals)
- One exception : for [a] with anterior variants, bilinguals show considerably less variation than French natives
- The plot does not show the posterior variants for [a]

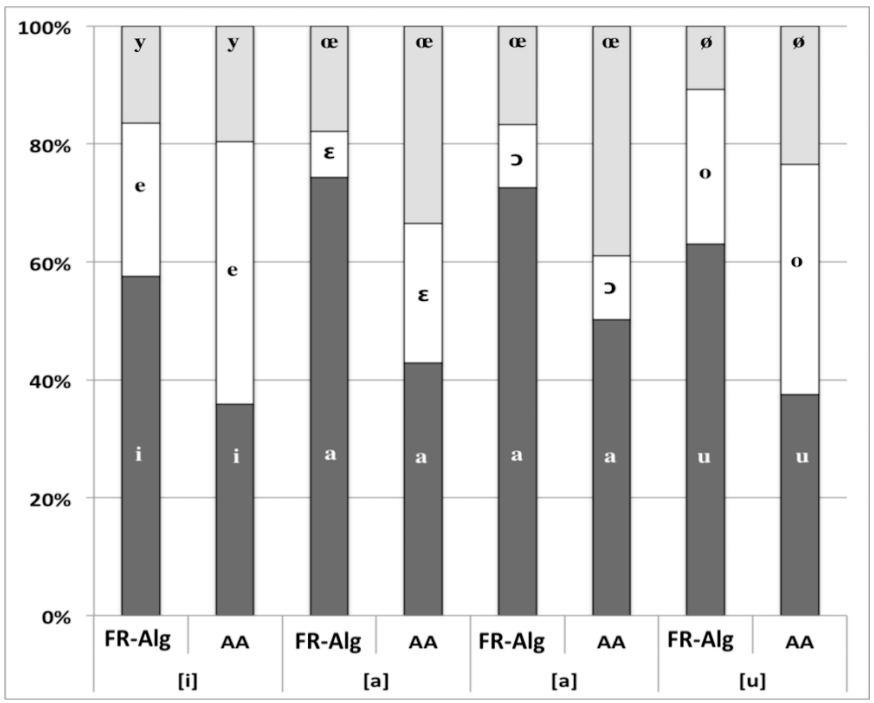
# Exp. 2: Vowel variants in code-switching

- Population: bilinguals (French-Algerian Arabic)
- Languages: French, Algerian Arabic
- Production variation for three target vowels, each with two variants
- Vowel production variation in bilinguals as a function of language
- French acoustic model

Are the realizations of Arabic vowels acoustically close to French vowels?

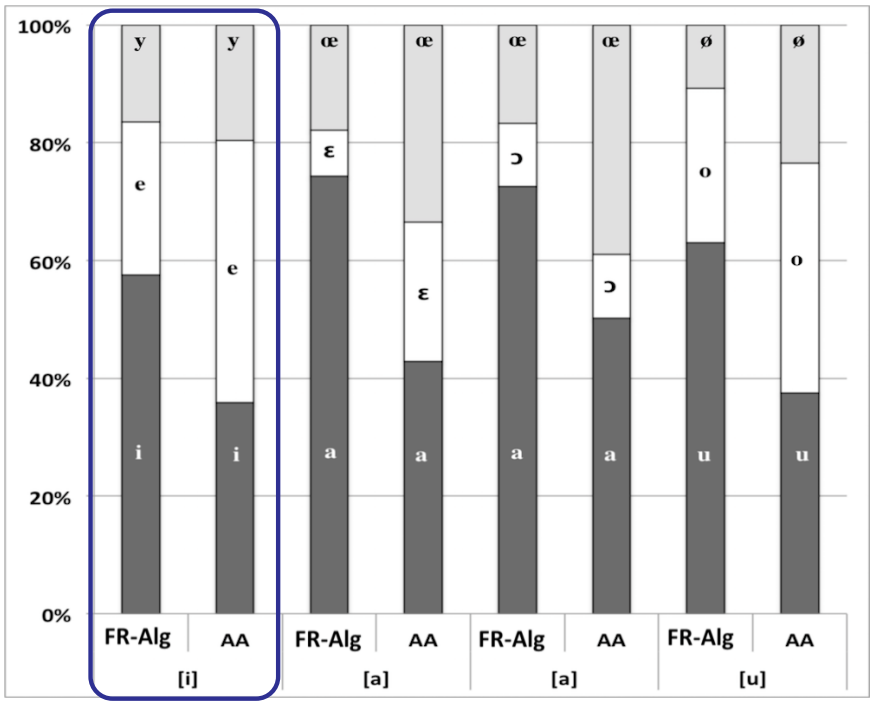
| Vowel | Variants |
|---|---|
| i | [e, y] |
| a | [ɛ, œ] (anterior) |
| a | [ɔ, œ] (posterior) |
| u | [o, ø] |

# Exp. 2: Results (bilinguals: code-switching)



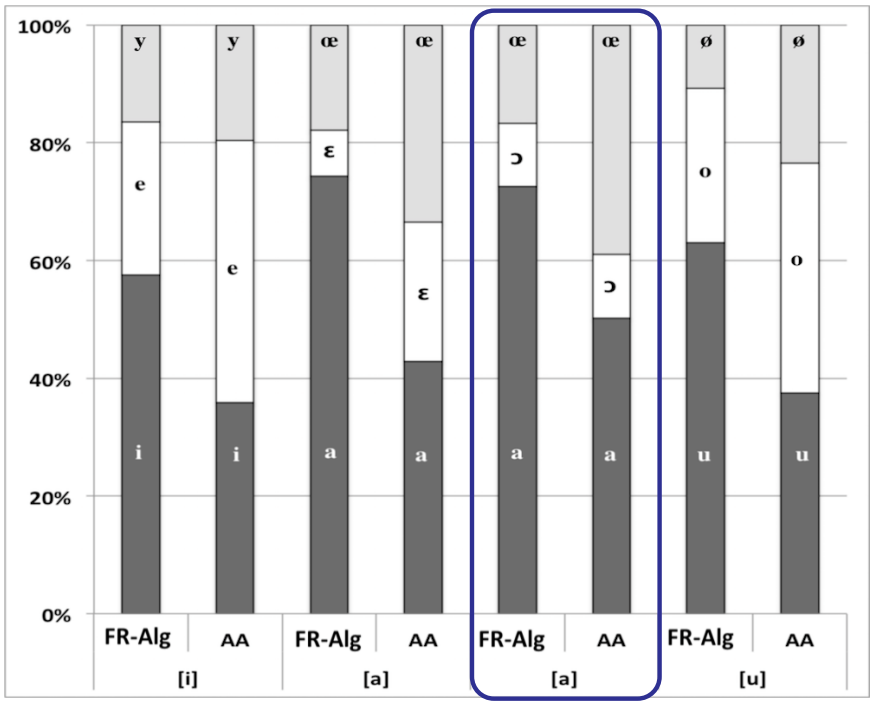

- The observed variation is vowel dependent

# Exp. 2: Results (bilinguals: code-switching)



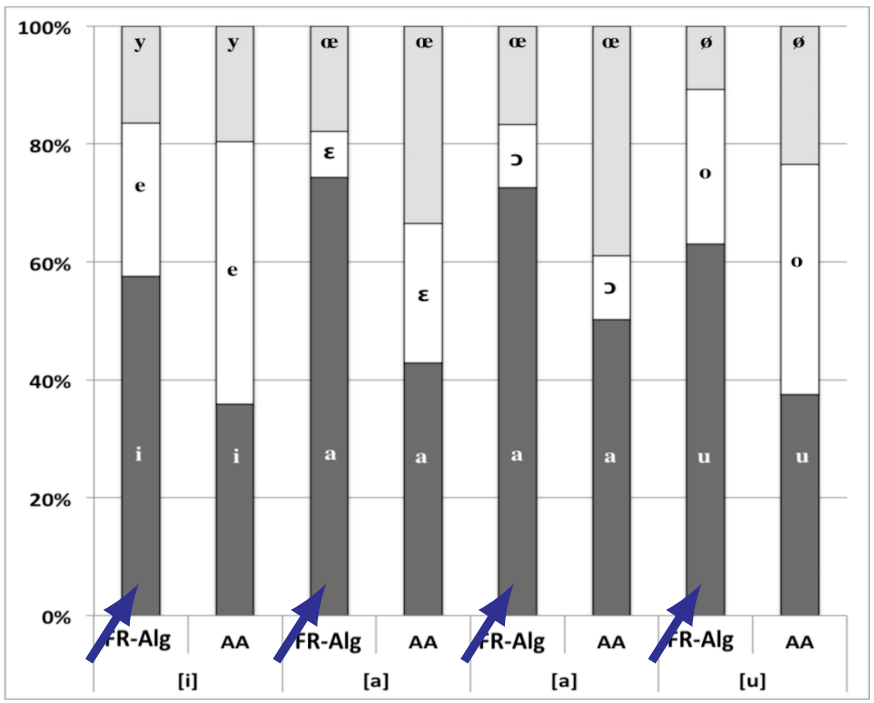

- The observed variation is vowel dependent
  - [i] is substituted more often than the other vowels ([a], [u])

# Exp. 2: Results (bilinguals: code-switching)




- The observed variation is vowel dependent
  - [i] is substituted more often than the other vowels ([a], [u])
  - [a] (post) is least often substituted

# Exp. 2: Results (bilinguals: code-switching)




- The observed variation is vowel dependent
  - [i] is substituted more often than the other vowels ([a], [u])
  - [a] (post) is least often substituted
- Language also has an impact on vowel variation
  - in French, the target vowel is more often produced than in Algerian Arabic
  - this pattern is observed for all target vowels

# Exp. 3 (a & b): Vowel centralization

- **Exp. 3a**
  - French from natives vs. bilinguals
  - Target vowels: [i, e, ɛ, a, ɔ, o, u, ɛ̃, ã, ɔ̃]
- **Exp. 3b**
  - Algerian Arabic from bilinguals (read vs spontaneous code-switching)
  - Target vowels: [i, iː, a, aː, u, uː]
- One production variant for each target vowel: schwa [ə]

Quantify movement of peripheral vowels towards the center of the vowel triangle

# Exp. 3a: Results (French: vowel centralization)

| Vowel | FR | FR-Alg |
|---|---|---|
| i | 14.1 | 12.8 |
| e | 20.9 | 24.4 |
| ɛ | 34.1 | 15.9 |
| a | 34.0 | 15.9 |
| ɔ | 39.4 | 20.2 |
| o | 33.5 | 21.6 |
| u | 25.0 | 16.2 |
| ɛ̃ | 13.6 | 7.7 |
| ɑ̃ | 17.5 | 8.7 |
| ɔ̃ | 17.7 | 6.5 |

Schwa variant rates (%)

- In French, vowel centralization is vowel dependent

# Exp. 3a: Results (French: vowel centralization)

| Vowel | FR | FR-Alg |
|---|---|---|
| i | 14.1 | 12.8 |
| e | 20.9 | 24.4 |
| ɛ | 34.1 | 15.9 |
| a | 34.0 | 15.9 |
| ɔ | 39.4 | 20.2 |
| o | 33.5 | 21.6 |
| u | 25.0 | 16.2 |
| ɛ̃ | 13.6 | 7.7 |
| ã | 17.5 | 8.7 |
| ɔ̃ | 17.7 | 6.5 |

Schwa variant rates (%)

- In French, vowel centralization is vowel dependent
  - [ɔ] is most affected by vowel centralization (29.8 %)

# Exp. 3a: Results (French: vowel centralization)

| Vowel | FR | FR-Alg |
|---|---|---|
| i | 14.1 | 12.8 |
| e | 20.9 | 24.4 |
| ɛ | 34.1 | 15.9 |
| a | 34.0 | 15.9 |
| ɔ | 39.4 | 20.2 |
| o | 33.5 | 21.6 |
| u | 25.0 | 16.2 |
| ɛ̃ | 13.6 | 7.7 |
| ɑ̃ | 17.5 | 8.7 |
| ɔ̃ | 17.7 | 6.5 |

Schwa variant rates (%)

- In French, vowel centralization is vowel dependent
  - [ɔ] is most affected by vowel centralization (29.8 %)
  - [ɛ̃] is least affected by centralization (10.7 %)

# Exp. 3b: Results (Arabic: vowel centralization)

| Vowel | Reading | CS |
|---|---|---|
| i | 56.5 | 37.9 |
| iː | 15.0 | 19.7 |
| a | 42.4 | 49.0 |
| aː | 26.8 | 36.4 |
| u | 44.7 | 41.1 |
| uː | 24.0 | 33.0 |

Schwa variant rates (%)

- In Algerian Arabic, vowel centralization is also vowel dependent

# Exp. 3b: Results (Arabic: vowel centralization)

| Vowel | Reading | CS |
|---|---|---|
| i | 56.5 | 37.9 |
| iː | 15.0 | 19.7 |
| a | 42.4 | 49.0 |
| aː | 26.8 | 36.4 |
| u | 44.7 | 41.1 |
| uː | 24.0 | 33.0 |

Schwa variant rates (%)

- In Algerian Arabic, vowel centralization is also vowel dependent
  - [iː] is less often centralized (17.4 %) than the other vowels (39.2 %)

# Exp. 3b: Results (Arabic: vowel centralization)

| Vowel | Reading | CS |
|---|---|---|
| i | 56.5 | 37.9 |
| iː | 15.0 | 19.7 |
| a | 42.4 | 49.0 |
| aː | 26.8 | 36.4 |
| u | 44.7 | 41.1 |
| uː | 24.0 | 33.0 |

Schwa variant rates (%)

- In Algerian Arabic, vowel centralization is also vowel dependent
  - [iː] is less often centralized (17.4 %) than the other vowels (39.2 %)
- Speech style i.e. read vs spontaneous CS does not have much impact on vowel centralization in Algerian Arabic

# Discussion – Experiments 1 & 2

## Parallel vowel variants

- In French (French natives, bilinguals):
  - comparable amount of variation in both populations
  - bilinguals seem to vary less on low vowels than high vowels
    $\rightarrow$ different pattern of variation compared to French natives
- Productions in bilinguals (French, Algerian Arabic):
  - more vowel variation in Algerian Arabic than in French
  - different production strategies according to the language
  - in our study, bilingual speakers adapt to the language and vary their vowels accordingly
- Our data suggest that in Algerian Arabic, the phonetic variant [e:] seems rather a variant of [i] than [a]

# Discussion – Experiment 3a & b

## Vowel centralization

- In French (French natives, bilinguals):
  - [ɔ] is more often centralized compared to the other target vowels
  - conform to the findings in [Boula de Mareüil et al., 2008]
  - bilinguals centralize vowels in the same way as do French natives
- In Algerian Arabic (reading, code-switching):
  - speech style does not have an impact on vowel reduction rate
  - [iː] is less often centralized than the other vowels
  - possible reason: extreme position of [iː] in the vowel triangle
  - in order to investigate this hypothesis, further acoustic analyses are needed

# Conclusions

- French-Algerian Arabic bilinguals are able to vary their vowel productions according to the language they speak
- The bilinguals from our study adjust their productions to the respective vowel systems

# References I


Amazouz, D., Adda-Decker, M., and Lamel, L. (2018).

The French-Algerian Code-Switching Triggered audio corpus (FACST).

In *Proc. Eleventh International Conference on Language Resources and Evaluation LREC 2018*, pages 1468–1473.

Auer, P. (2013).

*Code-switching in conversation: Language, interaction and identity*.

Routledge.

Boula de Mareüil, P., Vieru-Dimulescu, B., Woehrling, C., and Adda-Decker, M. (2008).

Accents étrangers et régionaux en français.

*Traitement Autom. Lang*, 49(3):135–163.

Bullock, B. E. (2012).

*Phonetic reflexes of code-switching*, chapter 10, pages 163–181.

Cambridge University Press, Cambridge.

# References II


Fricke, M., Kroll, J. F., and Dussias, P. E. (2016).

Phonetic variation in bilingual speech: A lens for studying the production–comprehension link.

*Journal of memory and language*, 89:110–137.

Grosjean, F. (1995).

A psycholinguistic approach to code-switching: The recognition of guest words by bilinguals.

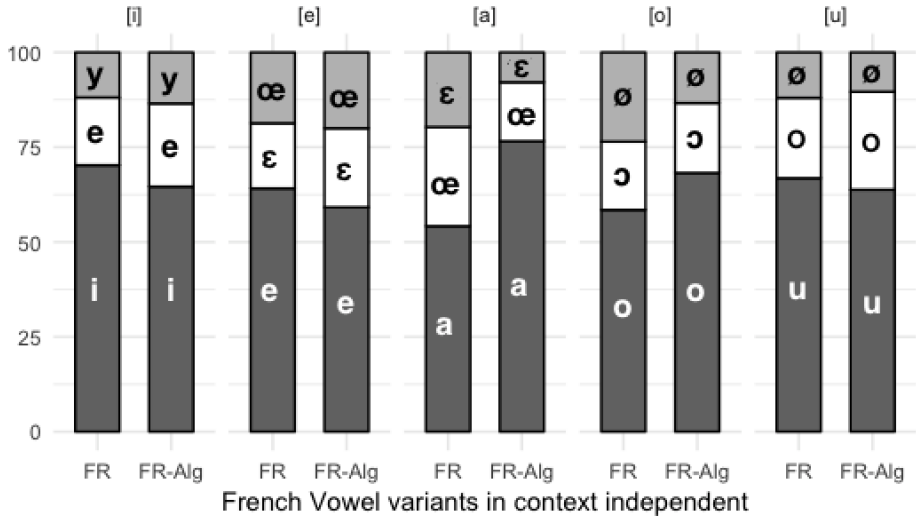
*One speaker, two languages: Cross-disciplinary perspectives on code-switching*, pages 259–275.

Torreira, F., Adda-Decker, M., and Ernestus, M. (2010).

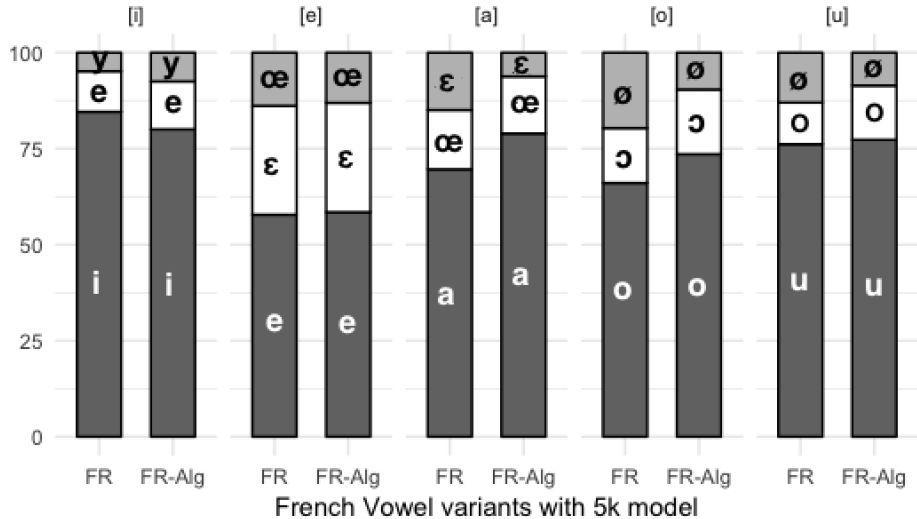
The Nijmegen corpus of casual French.

*Speech Communication*, 52(3):201–212.

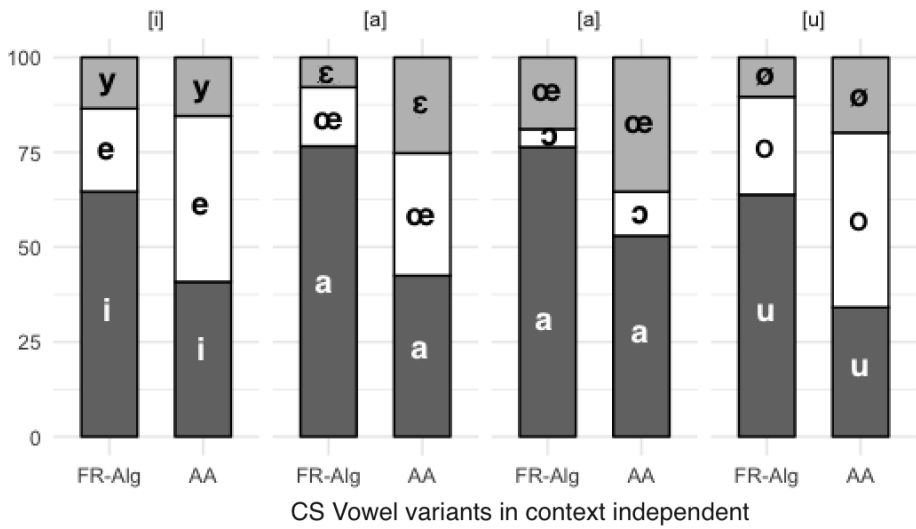
# Acoustic models: CI vs CD



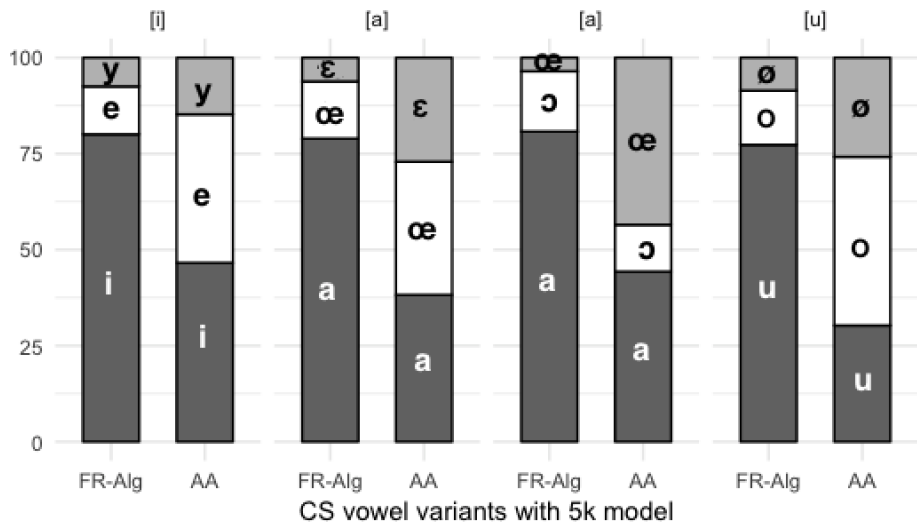
French Vowel variants in context independent



French Vowel variants with 5k model



CS Vowel variants in context independent



CS vowel variants with 5k model