

ICE-IPAC: An update on the protocol

Audrey Bonfiglioli
Nadine Herry-Bénit
Takeki Kamiyama
Véronique Lacoste
Ioana Trifu

Journées FLORAL-PFC 2015-IPFC
8 Décembre 2015

PLAN

ICE-IPAC

- Protocol
- Segmental and suprasegmental features
- Conventions for orthographic transcriptions
- Analysis of two French-speaking learners

General goals

- describe and analyze inter-phonological phenomena in different learning and language contact situations, i.e. EFL, ESL, and at different stages of the learning process
- consider L1 variation
- determine the status of the phonological system of IL learners/users, and consider variation as an integral part of the learning process



Data collection and protocol

- largely inspired by IPFC and PAC.

TASKS:

- repetition task: L1 specific list (78 words)
- reading of a word list (72+92 words)
- Reading the L1 specific word list (78 words)
- reading of a text
- formal interview and informal conversation



TASK: text reading

“Christmas interview” PAC text for all learners of English.



TASK: Formal interview (15 mins)

- Two different sets of questions based on the level: a) A1-B1; b) B2-C2.
- A list of closed questions identical for both levels.
- A different list of open questions for each level.
- The questions do not necessarily have to be all processed.
- The paper version must be used.

TASK: Informal conversation (20-25 mins)

- Both topics for the informal conversation must be determined and sent to the fieldworker before they perform the task.
- A1-B1 learners have 10 mins to get ready whereas B2-C1 learners have only 5 mins. They are allowed to take notes but they mustn't read their notes during the actual conversation.
- The first learner presents his/her topic then a discussion may unfold.
- The second learner presents his/her topic then a discussion may unfold.

Recording procedure

Two sessions of 1 hour each

SESSION 1

Presentation of the project (bilingual instruction sheet)

- Sign a consent form
- Fill the sociolinguistic questionnaire online OR on a sheet of paper (choice of the model accent)
- Task 1: repetition of L1 specific word list
- Task 2: reading list of words
- Task 3: reading of L1 specific word list
- Task 4: reading text

Recording procedure

Two sessions of 1 hour each

SESSION 2

- Task 5: formal interview
- Task 6: informal conversation

Data sample

- a minimum of 12 learners in each ICE-IPAC corpus.

Segmental features to study (common to ALL learners)

- Rhoticity
- Intrusive r, linking r
- H dropping/insertion
- Plosives: initial / medial/ final position
- /θ, ð/
- Intervocalic /t/ flap or stop
- lettER/CommA in unstressed position
- Vocalic contrasts:
 - BATH/TRAP/DANCE
 - KID/FLEECE
- Word stress placement
- Disyllabic words
- graphonemics

Non-linguistic features: sociolinguistic questionnaire

- Language learning history, extent of L1/L2 use
- Learning context (teacher and learner speech)
- Age
- Gender
- Training in L2 pronunciation
- Access to L2 communities and L2 use opportunities

Conventions for orthographic transcriptions

- If we wish to have comparable data we need to have uniform transcriptions.
- Transcribe under PRAAT using 3 tiers
- Name of tier 1: actual orthography
- Name of tier 2: target
- Name of tier 3: comments

Transcription of word lists

- One sound file for each list
- Create a single textgrid under PRAAT
- Tier 1: put a boundary tone at the beginning and end of each word
- Tier 1: Orthographic transcription of each word + what is said and not what we expected (SPPAS conventions)
- Tier 2: the target words
- Tier 3: comments for you and for the other transcriber. SAMPA symbols can be used.

Transcription of the text

- Create a single textgrid under PRAAT
- Tier 1: put a boundary tone at the beginning and end of each word
- Tier 1: Orthographic transcription of the way they read the text, putting boundaries when pauses are produced (SPPAS conventions).
- Tier 2: the target words
- Tier 3: comments for you and for the other transcriber. SAMPA symbols can be used.

Transcription of the two conversations

- For each conversation, at least 5 minutes conversation per learner must be transcribed orthographically.
- Create a single textgrid under PRAAT
- Tier 1: transcribe the conversation
- Tier 2: comments for you and for the other transcriber. SAMPA symbols can be used.



Conventions for coding

- The transcripts and phonetic / phonological coding should be developed according to the selected focus.
- We intend to start working on /r/ at the segmental and suprasegmental level



ICE-IPAC: French Learner's Speech Productions in English: Towards a classification of deviant forms or variations?



Case study of two French learners

- Two female French learners of EFL both enrolled at the University of Lyon
 - FR001 aged 22 (score TOEIC 920/990) (B2-C1)
 - FR002 aged 23 (not tested)

Corpus transcription

- Total duration of our corpus: 68 minutes
 - FR001: 29 min
 - FR002: 27 min
 - + 12 min of Informal conversation between the two speakers

Previous work on deviant forms

- The deviant forms in IPFC (Racine and al.2011)
 - Orthographic transcription of deviant forms
 - Limit of the IPA
 - Subjectivity of the transcription hence several transcriptors
- Classified according to:
 - Phonetic / phonological variations
 - Lexical transfers and code-switching
 - Morphological variations
 - Slip and performance errors
 - Onomatopoeia, abbreviations and hesitations



Methodology

- Acoustic and perceptual analysis of the 192 words (PAC List) in the word list task reading using PRAAT (Boersma & Weenink 2014)

Our results (word lists)

- Phonetic/phonological variations
 - Based on graphophonemics: Consonants Vs Vowels (Carr, P. 2014, Deschamps et al. 2004, Léon et al 2009)
 - According to their position in the word (Ginésy, 1995)
- Lexical transfers and code-switching
- Slip and performance errors

Monographic and Digraphic consonants

Monographic consonants:

- <p>
 - Initial position : <put>, <pause>, <pose>
 - Final position: <bishop>, <lap>
- <ng>
 - Medial position: <singer>, <stronger>
 - Final position: <rung>, <betting>
- <l>
 - Medial position: <middle>, <little>

- Final position: <peril>, <fell>

- <r>
 - Initial position: <row>, <rose>
 - Medial position: <bury>
 - Final position: <stir>, <err>

Digraphic consonants:

- <th>
 - Initial position: <three>
- <gh>
 - final position: <caught>

Monographic and Digraphic consonants

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FR001	FR002
LACK OF PHONEMIC CONTRAST		
REDUCTION IN UNSTRESSED POSITION		
WORD STRESS PLACEMENT		
PHONETIC TRANSFERS	<ng> pronounced [ŋg] <rung>	
	<ph> <shepherd> [f]	
	<r> final position <war, <far> [χ]	<r> initial position <rose> <r> final position stir <r> roulé créole
CONSONANT CLUSTER REDUCTION		

PHONETIC/ PHONOLOGICAL VARIATIONS

	FR001	FR002
REDUCTION OR LACK OF USE OF FEATURES FOUND IN THE TARGET L2 LANGUAGE	<p> initial position: pet, pat, put	
EXAGGERATION OF PHONOLOGICAL/PHONETIC PROPERTIES (LENGTH RULE)		
PHONE DELETION, SUBSTITUTION OR INSERTION OF EPENTHETIC VOWELS	<l> medial position: little, middle, meddle [æɪ] <l> final position peril [ɪ]	
CREATION OF NEW, TEMPORARY FEATURES, TO BE ABANDONED LATER	<r> medial position bury [w]	

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FR001	FR002
OVERGENERALISATION OF L1/L2 RULES ≈ HYPERCORRECTION INTERMEDIATE FORM BETWEEN L1-LIKE AND L2- LIKE PHONETIC REALISATIONS (VOT, FORMANTS)		<th> initial position <three> in between [θ]&[ʃ]
	<gh> medial position <caught> [f]/[θ]	
CONFORMITY TO NATIVE GRAPHOPHONEMICS	<ng> stronger	<ng> betting, stronger
	<p> initial position: paw, pause, pearl <p> medial position <spirit, sport>	

Conclusion on the pronunciation of monographic and digraphic consonants

- <p> non aspirated
- <r> no stability: rhoticity Vs non-rhoticity
- <gh> produced [f]/[θ]
- <ng> always produced [ŋg]

Monographic vowels

- <a>
 - Initial position: <afterwards>
 - Medial position: <master>, <father>, <marry>, <pant>
- <e>
 - Initial position : <err>
 - Medial position: <pet>, <bored>
 - Final position: <sue>, <file>
- <i>, <y>
 - Medial position: <pit>, <stir>
 - Final position <marry, Mary>
- <o>
 - Initial position: <one>
 - Medial position: <pose>, <rose>, <stronger>
 - Final position: <two>
- <u>
 - Medial position: <duck>, <bury>

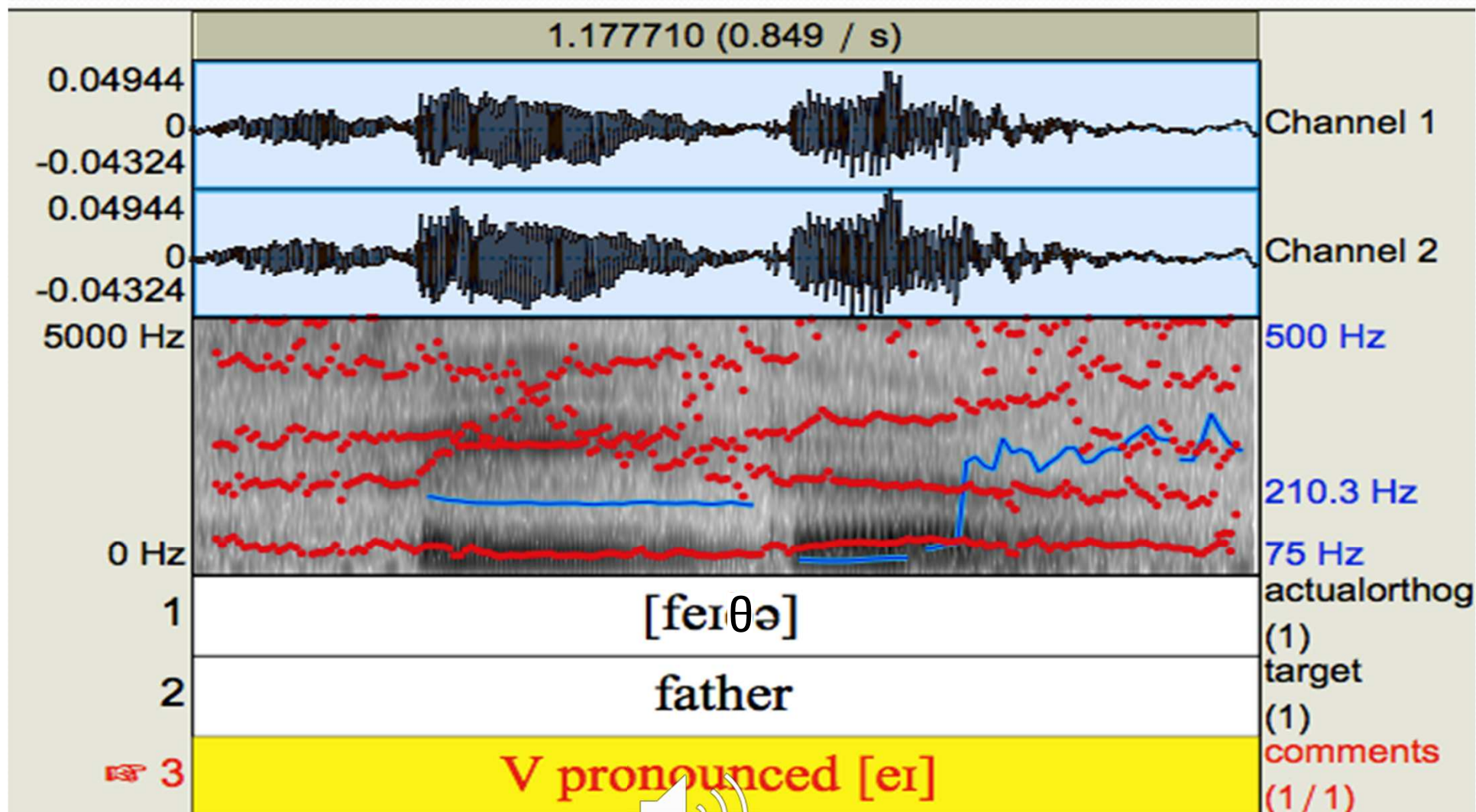
Preliminary results on Monographic vowels: <a>

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FRoo1	FRoo2
LACK OF PHONEMIC CONTRAST		
REDUCTION IN UNSTRESSED POSITION		
WORD STRESS PLACEMENT		
PHONETIC TRANSFERS	< a> initial position <afterwards> [a]	
		<a> medial position <master, bad, pat, lap>
CONSONANT CLUSTER REDUCTION		

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FR001	FR002
REDUCTION OR LACK OF USE OF FEATURES FOUND IN THE TARGET L2 LANGUAGE		
EXAGGERATION OF PHONOLOGICAL/PHONETIC PROPERTIES (LENGTH RULE)		
PHONE DELETION, SUBSTITUTION OR INSERTION OF EPENTHETIC VOWELS		
CREATION OF NEW, TEMPORARY FEATURES, TO BE ABANDONED LATER	<a> <fad, sack, lap, lab> from a French [a] to a native like [æ]	

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FR001	FR002
OVERGENERALISATION OF L1/L2 RULES ≈ HYPERCORRECTION INTERMEDIATE FORM BETWEEN L1-LIKE AND L2- LIKE PHONETIC REALISATIONS (VOT, FORMANTS)		<an> initial position <ants> [ãnts] <an> medial position <pant, plant> [ãnt],
	<an> medial position <pant, plant, dance> [en]	<a> medial position <badge, batch> [ei] [ej]
CONFORMITY TO NATIVE GRAPHOPHONEMICS	<a> initial position <agreed> <a> medial position <farther>	
	<an> medial position	<dance> GA pronunciation

Figure 1 <father> pronounced by FR001



Conclusion on the pronunciation of <a> monograph

- In initial and medial position, in a context followed by <n>, the vowel is nasalized by FRoo2.
- In a medial position and in a context other than near <n>, there's an over-generalization of the L2 rules for the learner, therefore there's a miss-pronunciation of the monograph by FRoo2.
- In the medial position in a context followed by <n> the vowel is closed by the FRoo1.

Vowel digraphs

- <ea>
 - Initial position: <earth> <earthy>
 - Medial position: <seal>, <beard>, <leaven>
- <ee>
 - Medial position: <feel>, <meet>
 - Final position: <degree>, <decree>
- <oo> medial position: <moor>, <poor>, <room>
- <au>, <aw>
 - Medial position: <caught>, <naught>, <pause>
 - Final position: <paw>
- <ou>, <ow>
 - Medial position: <foul>, <gourd>, <four>
 - Final position: <row>, <anyhow>
- <eo>: medial position : <leopard>
- <ee> medial position : <feel>
- <ai> medial position: <fail>
- <oa>: medial position: <foal>
- <oi>: medial position: <foil>

Preliminary results on vowel digraph: <ea>

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FRoo1	FRoo2
LACK OF PHONEMIC CONTRAST		
REDUCTION IN UNSTRESSED POSITION		
WORD STRESS PLACEMENT		
PHONETIC TRANSFERS		
CONSONANT CLUSTER REDUCTION		

PHONETIC/ PHONOLOGICAL VARIATIONS

	FRoo1	FRoo2
REDUCTION OR LACK OF USE OF FEATURES FOUND IN THE TARGET L2 LANGUAGE		
EXAGGERATION OF PHONOLOGICAL/PHONETI C PROPERTIES (LENGTH RULE)		
PHONE DELETION, SUBSTITUTION OR INSERTION OF EPENTHETIC VOWELS		
CREATION OF NEW, TEMPORARY FEATURES, TO BE ABANDONED LATER		

	PHONETIC/ PHONOLOGICAL VARIATIONS	
	FR001	FR002
OVERGENERALISATION OF L1/L2 RULES ≈ HYPERCORRECTION INTERMEDIATE FORM BETWEEN L1-LIKE AND L2- LIKE PHONETIC REALISATIONS (VOT, FORMANTS)		medial position <heaven> <leaven> pronounced /i/
	<seal> <bead> pronounced with a French /i/	
	<sea> French /i/ lengthened	
	<heart> pronounced as a short version of the English /ɜ: / sounds almost like <hurt>	
CONFORMITY TO NATIVE GRAPHOPHONEMICS	<earthy> <zeal>	<beard>

Conclusion on the digraph $\langle ea \rangle$

Realization of :

- Either an intermediate form
 - creation of a new system with properties from both L_1 and L_2
- Or a correct realization

Lexical transfers

- <bid> read like the French word <bide>
- <fad> read like the French word <fade>
- <fall> read like the French word <folle>
- <far> read like the French word <far>
- <full> read like the French word <foule>
- <gourd> read like the French word <gourde>
- <lack> read like the French word <lac>
- <meat> read like the French word <mythe>
- <metal> read like the French word <métal>
- <plant> read like the French word <plant>
- <pot> read like the French word <pot>
- <pour> read like the French word <pour>
- <rose> read like the French word <rose>
- <sack> read like the French word <sac>

Figure 2 <pot> pronounced by FR002

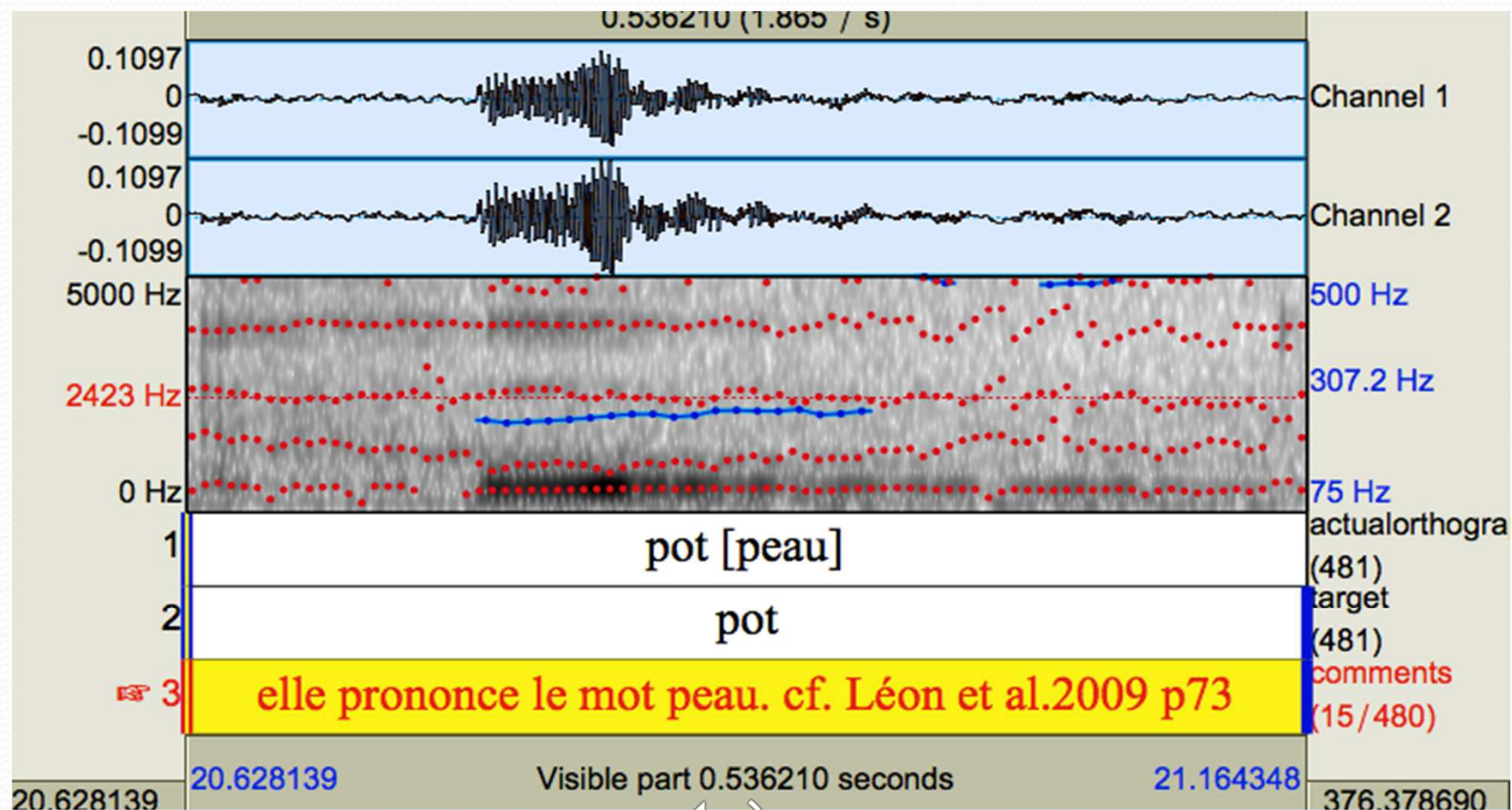


Figure 3 <rose> pronounced by FR002

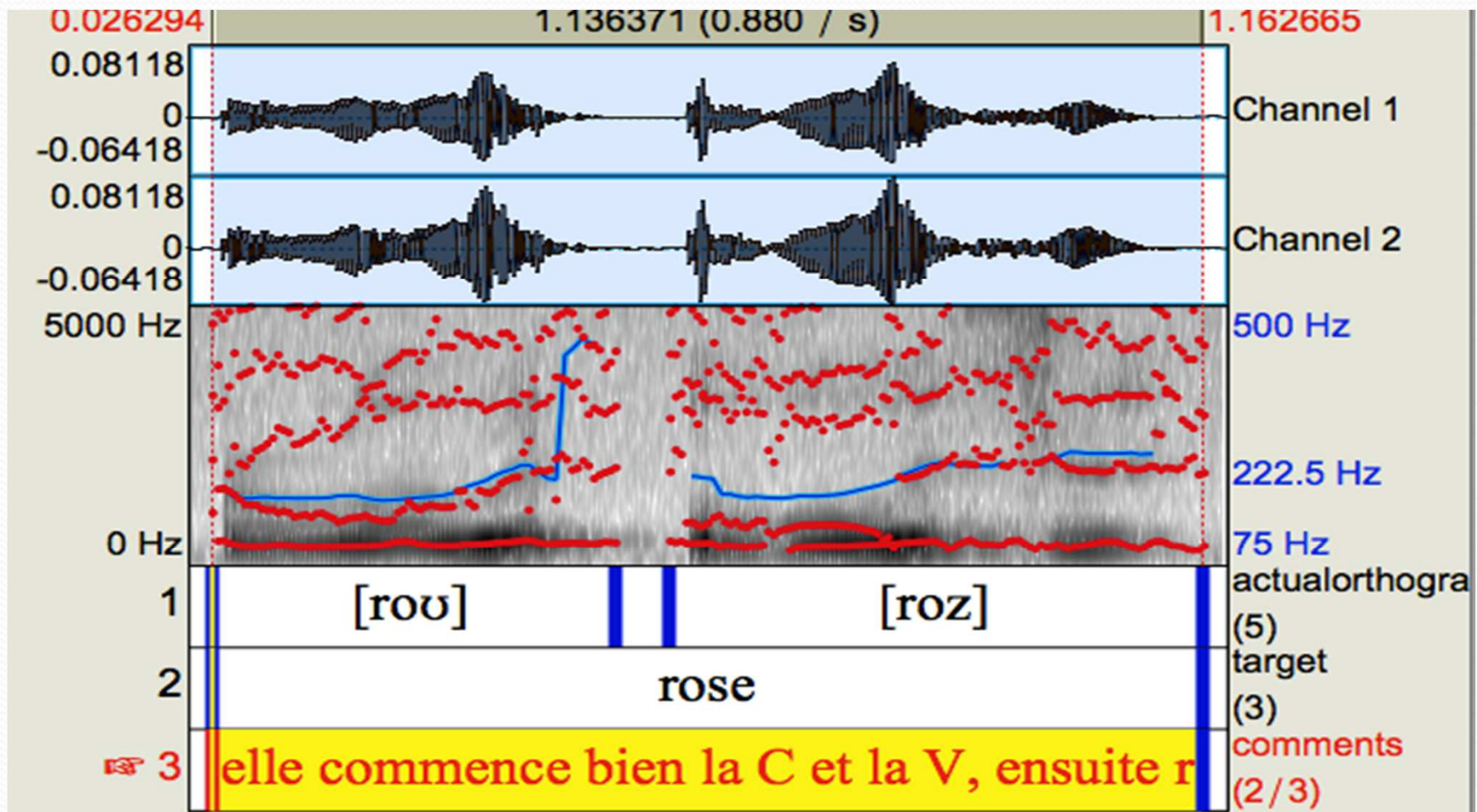
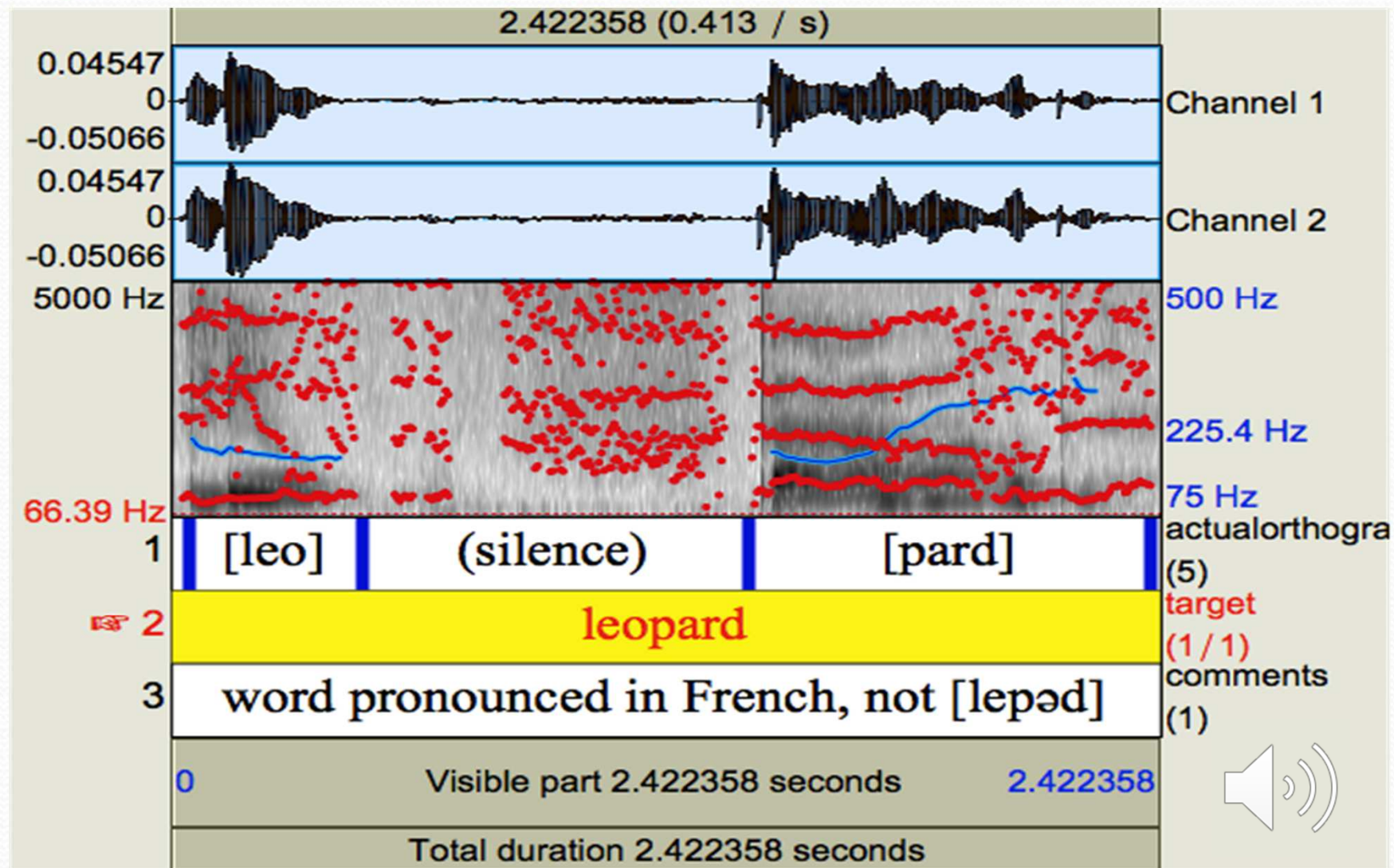


Figure 4 <leopard> pronounced by FR002



Slip and performance variation

- <furl> read like the French word <faux>
- <one hundred and eight> read <one hundred and hate>
- <purr> read almost like <poor>

Conclusion

On a total of 192 words analysed for each speaker:

- 6 consonant graphemes (4 monographic and 2 digraphic)
- 15 vowel graphemes (10 monographic and 5 digraphic)
- Deviant forms are mainly phonetic & phonological

Further discussion

- Correlation tests / oral productions
- Classification to be extended to
 - All graphemes
 - All tasks: so far 1 out of 5
 - More speakers with French L1
 - More speakers with different L1s
- Develop a similar research on suprasegmental features
- Project: PhD research on the /r/ both on segmental and suprasegmental

Bibliography

- **Ballier, N. & Martin, Ph.** to appear in 2015. 'Speech Annotation' in S. Granger, G. Gilquin & F. Meuniers (eds) *The Cambridge Handbook of Learner Corpus Research*, Cambridge: Cambridge University Press.
- **Boersma, P. , & Weenink, D.** (2014). *Praat: doing phonetics by computer* (Version 5.4). [Logiciel, En Ligne]. <http://www.fon.hum.uva.nl/praat/>.
 - **Detey, S., Kawaguchi, Y.** (2008). *Interphonologie du Français Contemporain (IPFC): récolte automatisée des données et apprenants japonais*. Journées PFC: Phonologie du français contemporain: variation, interfaces, cognition. Paris, Dec. 11-13, 2008.
 - **Durand, Jacques, Bernard Laks & Chantal Lyche** (2002). La phonologie du français contemporain: usages, variétés et structure. In: C. Pusch & W. Raible (eds.) *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language*. Tübingen: Gunter Narr Verlag, pp. 93-106.
 - **Durand, Jacques, Bernard Laks & Chantal Lyche** (2009). Le projet PFC: une source de données primaires structurées. In J. Durand, B. Laks et C. Lyche (eds)(2009) *Phonologie, variation et accents du français*. Paris: Hermès. pp. 19-61.
 - **Grabe & Coleman** (2004): The IViE Corpus http://www.phon.ox.ac.uk/files/apps/old_IViE/
 - **Granger, S.** (1993). The International Corpus of Learner English. *The European English Messenger* 2(1), 34. <http://www.essenglish.org/mess/mestoc21.html>.
 - **Gut U., Voormann, H.** (2014): " Corpus Design" in *The Oxford Handbook of Corpus Phonology* edited by Durand, J., Gut, U., Kristoffersen, G , Oxford, Oxford University Press pp. 13-26
 - **Kondo, M.** 2012. Design and analysis of Asian English speech corpus —How to elicit L1 phonology in L2 English data—. In: Tono, Y., Kawaguchi, Y., Minegishi, M. (eds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research*. Amsterdam & Philadelphia: John Benjamins, 251-278.
 - **Léon, P. & al.**(2009), *Phonétique du FLE*, Paris: Armand Colin
 - **Meunier** 2008 : Longitudinal Database of Learner English, LONGDALE www.uclouvain.be/en-cecl-longdale.html
 - **Racine, I., Detey, S., Zay, F., Y. Kawaguchi** (2012). Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2: l'exemple du projet « Interphonologie du français contemporain » (IPFC). In: Kamber, A., Skupiens, C. (éds). *Recherches récentes en FLE*. Berne: Peter Lang, pp. 1-19.
 - **Racine, I., Zay, F., Detey, S. & Kawaguchi, Y.** (2011). De la transcription de corpus à l'analyse interphonologique : enjeux méthodologiques en FLE. In Col, G. & Osu, S. N. (eds), *Transcrire, écrire, Formaliser (1)*. Rennes: PUR. Travaux Linguistiques du CerLiCO 24. 13-30.
 - **Tortel, A.** (2008) : « ANGLISH. Une base de données comparatives de l'anglais lu, répété et parlé en L1 & L2 », *TIPA Travaux Interdisciplinaires du Laboratoire Parole et Langage*, no. 27, 111-122.



Thank you for your attention



