Perception of non-standard prosodic features in French via re-synthesis in PRAAT: preliminary findings

Christopher Stewart & Zsuzsanna Fagyal & Peter Golato
University of Illinois at Urbana-Champaign

Between Perceptual Dialectology and Experimental Phonetics

If perceptually identifiable phonetic variation gives rise to dialectal / social differences in speech, can experimental methodologies help us reveal the nature of such differentiation?

Yes!

"Future work in the perception of a variety might focus more specifically on the exact linguistic elements that give rise to perception rather than on the global presentation of varieties," Dennis Preston, *Perceptual dialectology (1999:xxxviii)*

Why perception?

- still not enough labeled, on-line speech corpora
- analysis of production data is costly
- ‘quick-and-dirty’ studies virtually impossible
  - e.g. looking into just one variable or the combined effects of two
- best way to ‘zoom in’ on certain variables

Language features triggering identifications

- Graff, Labov & Harris's (1986) -- instrumental manipulation of /aw/
  - fronting of the vowel alone is sufficient in signaling a speaker's ("white") ethnicity in Philadelphia
- Niedzielski 1999, Plichta & Preston -- 2000 vowel matching / identification tasks
  - stereotypes and cultural expectations influence the perception and identification of dialects in the US
- Stuart-Smith (2003) -- productions of /s/ in Glasgow English
  - Even the most precisely shaped acoustic articulations, e.g. fricatives, can carry social-indexical meaning (i.e. gender)

Speech synthesis

- Relatively little explored
- Tool to create the perfect ‘matched-guise’ test
- ‘Humming of tunes’ won’t get us far enough
- Siebenhaar, Frost & Keller (2004) – prosodic dialectal differences Bernese and Zurich Swiss German intonation

“*in a synthesis system, each prosodic parameter can be modified independently of all others [which] allows us to test these for the naturalness of speech as well as their language-specific or language-independent nature.*” (*p. 220*)

Testing perception via speech RE-synthesis

- same reasons as expressed by Siebenhaar et al. (2004)
  - *"each prosodic parameter can be modified independently of all others"*
- controlled manipulations of single or multiple acoustic correlates
- PRAAT: widely-used computer program
- automatic manipulations made possible by widely diffused and shared **scripts**
- all scripts used in this study are available on-line

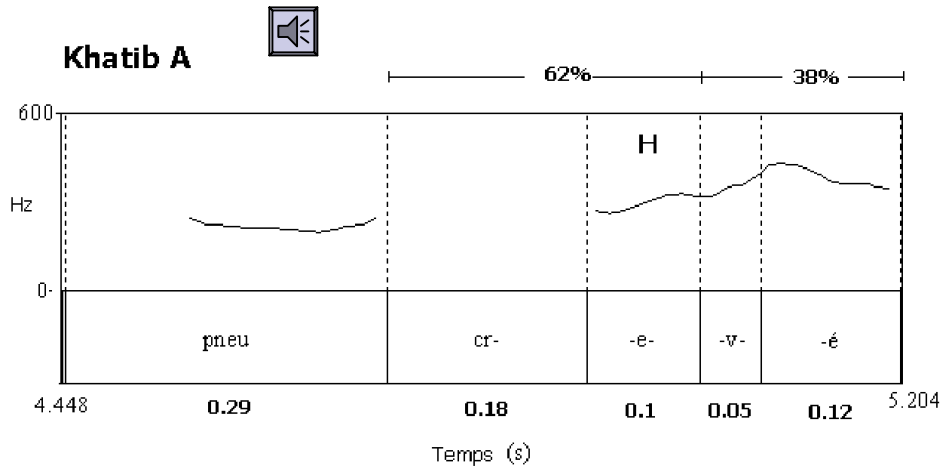
The cost of perceptual investigations…

- tedious
- questions sociolinguists tend to ask about speech are complex => need to be broken down
- typically two components
  - psycho-acoustics → discrimination
  - social attribution → identification / ranking
- different tasks (might) call for different experimental paradigm

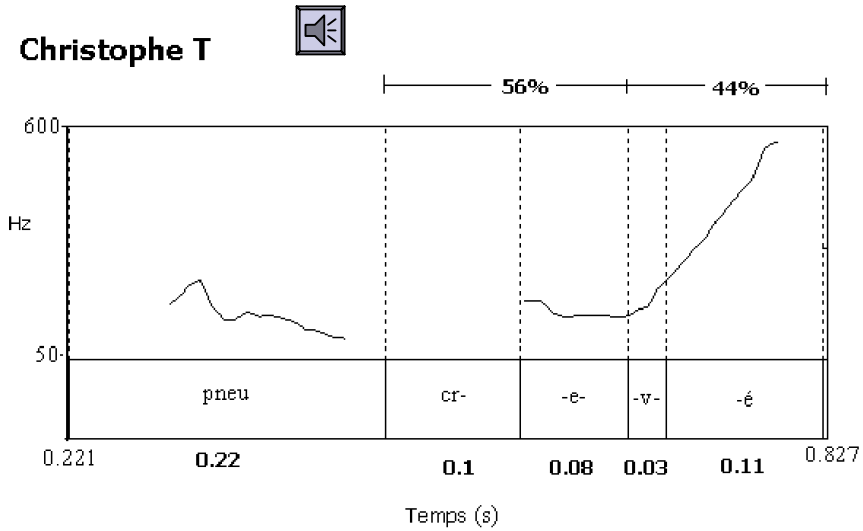
Rationale for this study

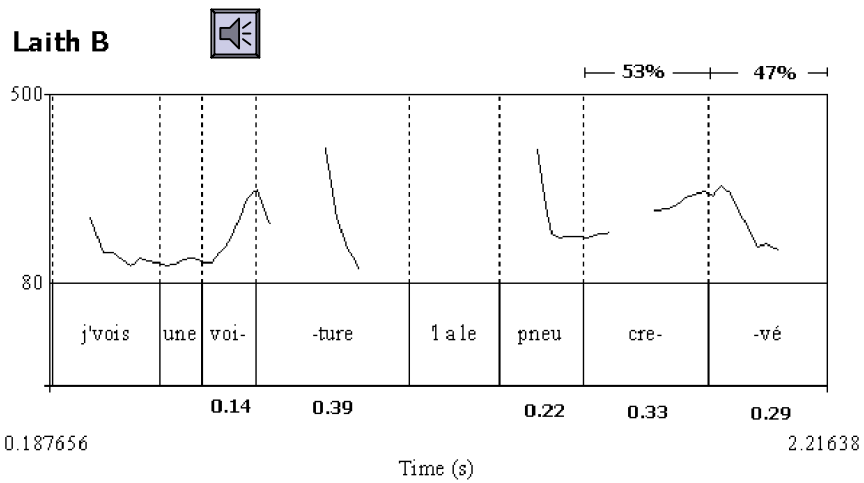
- testing one aspect of an earlier claim from a production study
- two listing contours from a picture-naming task from working-class adolescents (Fagyal 2003a, b, in press)
  - "standard" – phrase-final rise on the final syllable - LH*+H%
  - "non-standard" – "early" phrase-final rise - LH*+HL%
- non-standard contour more typical in speech of adolescents of North African immigrants
  - lengthening of the penultimate and shortening of the final
  - also important: rising pitch on the penultimate

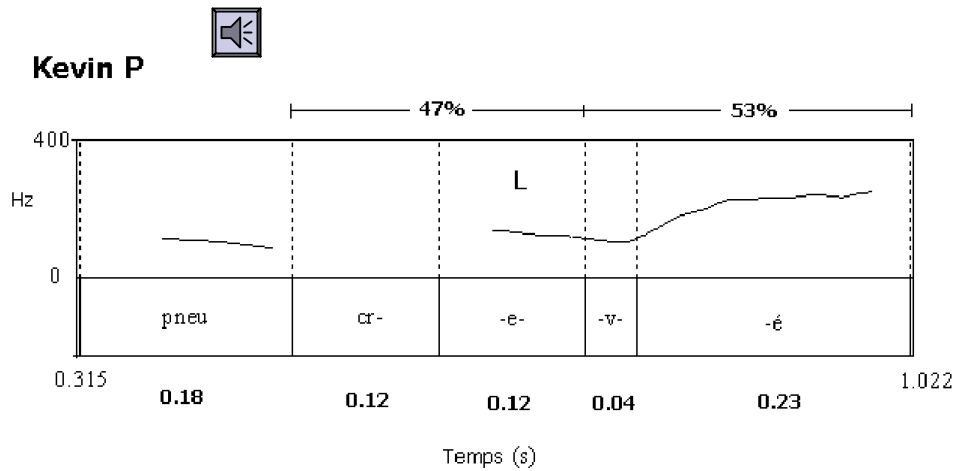
LH*+HL% - “early rise” (on the penultimate)



LH*+H% - “standard rise” (on the final)







What cues « non-standardness » on the penultimate?

- Long penultimate?
- Rising pitch (“early rise”) on the penultimate?
- Early rise coupled with the final fall?
- All of the above?
- None of the above?

Previous studies: undecided

duration:

- *tu t'rends: compte* 'Do you realize'?
- *de::: quoi?* 'Of what'?
- *faut être ba::laize* '[One] has to be tough'

Conein & Gadet (1998:109)

F0:

"a consistent and recurrent use of this 'penultimate high' tonal pattern, **with or without lengthening**, could account for the perceived prosodic distinctiveness of the vernacular of the working-class Parisian youth"

Fagyal (2003:xxx)

Break it down…

- Early rise vs. standard rise durational differences in two locations
  - Penult: long vs. short
  - Final: length can vary
- But → How short is short? How long is long?

  What is the listeners' threshold of discrimination between different durations?
- Testing first whether / how these syllable durations are perceived
- Pitch can come next

Experiment (1) - stimuli

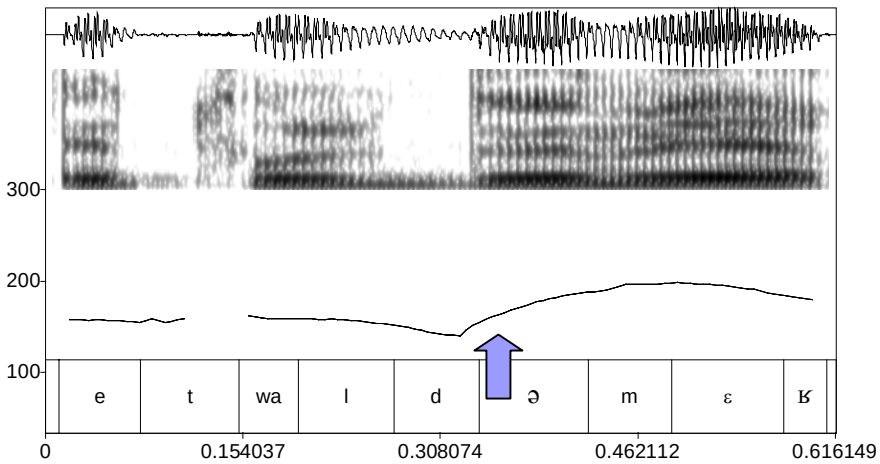
- ABX paradigm
  - A B is X same as A or B?
- systematic manipulations of penultimate and final syllable durations
  - mean syllable durations from non-standard and standard contours
  - shortened and lengthened in steps of
    - 1) 10 ms – 2) 20ms – 3) 40 ms – 4) 80 ms
- steps 1) to 3) yielded unrealistically high N of stimuli
- testing carried out in 8 steps of 80 ms for each syllable
  - Penult: 23 ms – 583 ms
  - Final: 60 ms – 620 ms

Experiment (2) - task

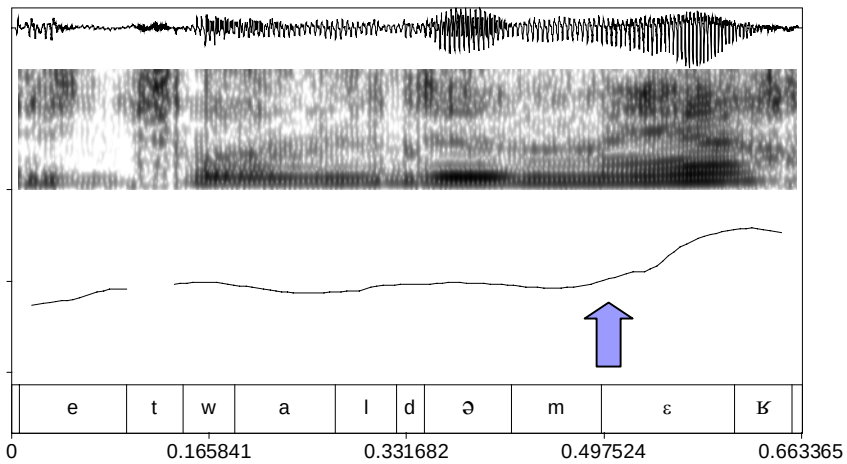
- eight native speakers of French
- 256 completely randomized stimuli of the target word *étoile de mer*
- prosodic parameters of « early » and « standard rising » patterns mixed
- 'synthetic' matched-guised technique:
  - used voice of the « standard » speaker
  - eliminating unwanted effect of non-standard segmental information
  - chose speaker from the community in case voice quality would be a factor

The two contours

“Early rise”



“Standard rise”

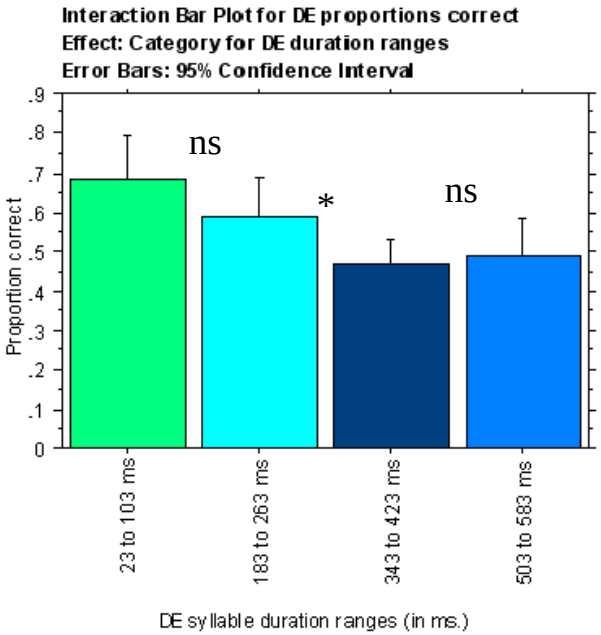


Predictions

1. Length of penultimate is used for socio-indexical information → continuous durational differences categorized into short vs. long penultimate durations => Expect a threshold
2. Length of the final syllable not used to encode socio-indexical information → listeners won't perceive any category between continuous durational differences => No threshold
3. Listeners' will reject phonologically inappropriate durations in both locations

Rate of discrimination to drop down to chance in very long and very short stimuli

Results (1): scores for penultimate syllables

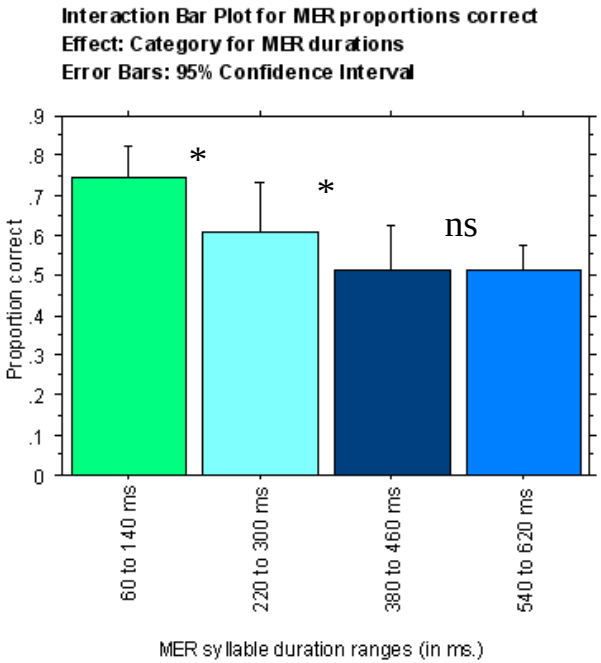


Means Table for DE proportions correct
Effect: Category for DE duration ranges

	Count	Mean	Std. Dev.	Std. Err.
23 to 103 ms	7	.684	.117	.044
183 to 263 ms	7	.589	.108	.041
343 to 423 ms	7	.470	.064	.024
503 to 583 ms	7	.494	.100	.038

- Correct identifications with high error rate → task was difficult
- One listener not on task → taken out
- Repeated-measures ANOVA on proportion of correct answers (all conditions): F (6, 3) = 7.6, p = .001
- Numerical but no statistically significant split between very short vs. short (first two columns)
- "Categorical" difference between short and long(< 263 ms and > 343 ms)
- Last two blocks: listeners ability to identify correctly X as either A or B drops down to chance

Results (2): scores for final syllables



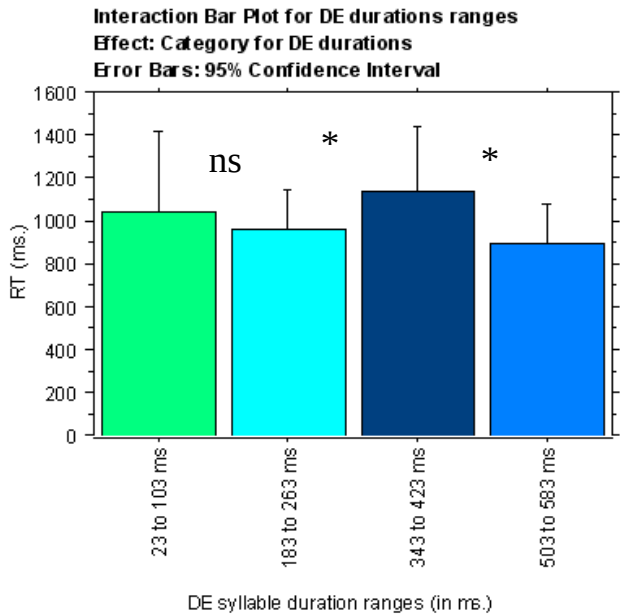
Means Table for MER proportions correct
Effect: Category for MER duration ranges

	Count	Mean	Std. Dev.	Std. Err.
60 to 140 ms	7	.744	.089	.034
220 to 300 ms	7	.611	.129	.049
380 to 460 ms	7	.511	.120	.045
540 to 620 ms	7	.511	.069	.026

- Correct identifications with high error rate → task was difficult
- Repeated-measures ANOVA on proportion of correct answers (all conditions): F (6, 3) = 14.04, p < .0001
- Statistically significant split between short vs. long (first two columns)
- "Continuous" rather than categorical degradation of performance
- Last two blocks: listeners ability to identify correctly X as either A or B drops down to chance

Results (3): RTs for penultimate syllables

- still no sensitivity to shortening
  - shortest duration patterns elicit same reaction times
- Repeated-measures ANOVA: F (6, 3) = 3.09, p = .05
- Barely significant main effect
- Statistically significant split between all other columns
- RTs quite large and very similar across conditions (around 900-1000 ms) overall: long sets of stimuli to judge



Means Table for DE duration ranges

	Count	Mean	Std. Dev.	Std. Err.
23 to 103 ms	7	1042.143	408.003	154.210
183 to 263 ms	7	960.000	196.774	74.374
343 to 423 ms	7	1133.429	334.258	126.338
503 to 583 ms	7	889.857	202.791	76.648

Results (4): RTs for final syllables

- Repeated-measures ANOVA: $F(6, 3) = 1.6$, $p = .21$ (ns.)
- No significant main effect
- Mean RTs very similar
- **Continuous** increase in standard error rate and standard deviation with longer durations
- Nothing overly salient or familiar seems to have speeded up or slowed down listeners' tendencies of discrimination

Means Table for MER duration ranges

	Count	Mean	Std. Dev.	Std. Err.
60 to 140 ms	7	1021.714	200.036	75.606
220 to 300 ms	7	1152.714	378.321	142.992
380 to 460 ms	7	992.000	428.476	161.949
540 to 620 ms	7	1002.286	434.100	164.074

Summary – penultimate: as expected

- Some numerical but no statistically significant split between very short vs. short (first two columns)

It doesn't matter how short the penultimate gets until about 263 ms!

- "Categorical" difference between short and long (< 263 ms and > 343 ms)
- Last two blocks: listeners ability to identify correctly X as either A or B drops down to chance

It doesn't matter how long the penultimate gets beyond 343 ms !

Summary - final: mixed feelings

- NO: Did not expect statistically significant split between very short vs. short (first two columns) and short vs. long syllables
  - Listeners ARE SENSITIVE to very short durations if the final syllable
  - Carries primary stress that is cues by duration
  - Possible contribution of phrase-final lengthening
- YES: “Continuous” rather than categorical degradation of performance
- YES Last two blocks: listeners ability to identify correctly X as either A or B drops down to chance

Discussion

- TBA

Summary - final: mixed feelings

Works Cited

B. Conein, and F. Gadet, “Le « français populaire » des jeunes de la banlieue parisienne entre permanence et innovation”, In : *Jugendsprache / Langue des jeunes / Youth language*, J. Androutsopoulos and A. Scholz (eds.), Peter Lang, Frankfurt, pp. 105-123, 1998.

Fagyal, Z. (2003a). La prosodie du français populaire des jeunes à Paris: traits héréditaires et novateurs. *Le Français aujourd'hui, 143* (n° spécial "Français de l’école et langues des élèves : quel statut, quelles pratiques ?"), 47-55.

Fagyal, Z. (2003b). The Matter with the Penultimate: Prosodic Change in the Vernacular of Lower-Class Immigrant Youth in Paris, *Proceedings of the XVth International Congress of Phonetic Sciences*, Vol. 1 (pp. 671-674), Barcelona.

Fagyal, Z. (in press). Prosodic consequences of being a *Beur*: French in contact with immigrant languages in Paris, *Selected Papers from NWAV 32*, 10(2), Philadelphia: Working Papers in Linguistics, Department of Linguistics, University of Pennsylvania.

Jun, Sun-Ah & Fougeron, Cécile. 2002. Realizations of Accentual Phrase in French Intonation. Probus 14(1), 147-172.

Works Cited (cont.)

Peters, Jörg et al. 2002. Identification of Regional Varieties by Intonational Cues. An Experimental Study on Hamburg and Berlin German. Language and Speech 45(2), 115-139.

van Bezooijen, Renée & Gooskens, Charlotte. 1999. Identification of Language Varieties: The Contribution of Different Linguistic Levels. Journal of Language and Social Psychology 18(1), 31-48.

Walton, Julie & Orlikoff, Robert. 1994. Speaker Race Identification From Acoustic Cues in the Vocal Signal. Journal of Speech and Hearing Research 37, 738-745.