Using reverse correlation to study speech perception

Léo Varnet

Laboratoire des Systèmes Perceptifs, École Normale Supérieure Paris

Séminaire LPNC, 04/04/2023

Decoding speech

"Cracking the speech code": finding the auditory primitives of speech comprehension

Decoding speech

"Cracking the speech code": finding the auditory primitives of speech comprehension

- Which acoustic cues allow listeners to differentiate one phoneme from another?
- Which acoustic cues underlie the segmentation of the speech signal into words?

Decoding speech

"Cracking the speech code": finding the auditory primitives of speech comprehension

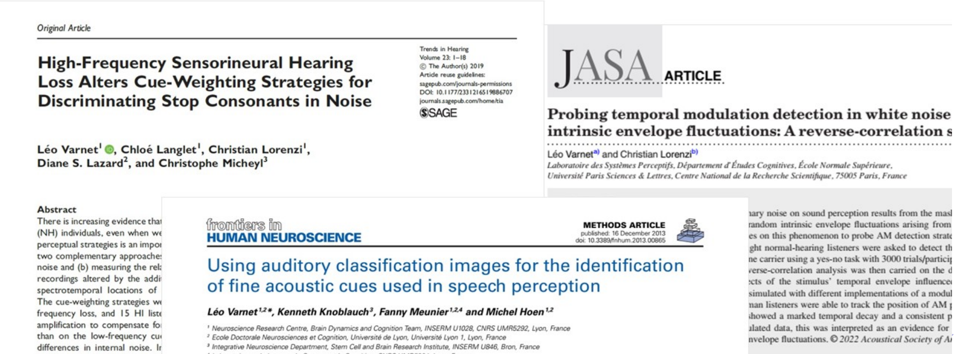
- Which acoustic cues allow listeners to differentiate one phoneme from another?
- Which acoustic cues underlie the segmentation of the speech signal into words?

No easy answer, due to the spectrotemporal complexity of natural speech.

Auditory revcorr

Reverse correlation (aka **revcorr**) is the perfect tool to reveal perceptual cues used in a psychophysical task, based on purely behavioral data... in particular for auditory categorization tasks. *[Varnet et al. 2013, 2015; Osses & Varnet, 2021; Varnet & Lorenzi, 2022]*

Core idea: adding **random fluctuations** to the stimulus and measure how they affect the participant's responses on a trial-by-trial basis.


Original Article

High-Frequency Sensorineural Hearing Loss Alters Cue-Weighting Strategies for Discriminating Stop Consonants in Noise

Trends in Hearing
Volume 23: 1–18
© The Author(s) 2019
Article reuse guidelines: sagepub.com/journals-permissions
DOI: 10.1177/2331216519886707
journals.sagepub.com/home/tia
SAGE

Léo Varnet[1], Chloé Langlet[1], Christian Lorenzi[1], Diane S. Lazard[2], and Christophe Micheyl[3]

Abstract
There is increasing evidence that (NH) individuals, even when we perceptual strategies is an impor two complementary approache noise and (b) measuring the rel recordings altered by the addi spectrotemporal locations of The cue-weighting strategies w frequency loss, and 15 HI liste amplification to compensate fo than on the low-frequency cu differences in internal noise. Ir

JASA ARTICLE

Probing temporal modulation detection in white noise intrinsic envelope fluctuations: A reverse-correlation s

Léo Varnet[a)] and Christian Lorenzi[b)]
Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, Université Paris Sciences & Lettres, Centre National de la Recherche Scientifique, 75005 Paris, France

ıary noise on sound perception results from the mas random intrinsic envelope fluctuations arising from es on this phenomenon to probe AM detection strate ght normal-hearing listeners were asked to detect th ne carrier using a yes-no task with 3000 trials/partici verse-correlation analysis was then carried on the d ɛcts of the stimulus' temporal envelope influence simulated with different implementations of a modul nan listeners were able to track the position of AM showed a marked temporal decay and a consistent p ılated data, this was interpreted as an evidence for nvelope fluctuations. © 2022 Acoustical Society of A

frontiers in HUMAN NEUROSCIENCE
METHODS ARTICLE
published: 16 December 2013
doi: 10.3389/fnhum.2013.00865

Using auditory classification images for the identification of fine acoustic cues used in speech perception

Léo Varnet[1,2]*, Kenneth Knoblauch[3], Fanny Meunier[1,2,4] and Michel Hoen[1,2]
[1] Neuroscience Research Centre, Brain Dynamics and Cognition Team, INSERM U1028, CNRS UMR5292, Lyon, France
[2] Ecole Doctorale Neurosciences et Cognition, Université de Lyon, Université Lyon 1, Lyon, France
[3] Integrative Neuroscience Department, Stem Cell and Brain Research Institute, INSERM U846, Bron, France

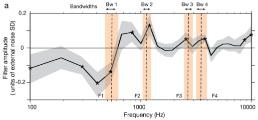
high level

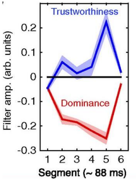
low level

Auditory revcorr studies (full diagram on https://dbao.leo-varnet.fr/)

paralinguistics

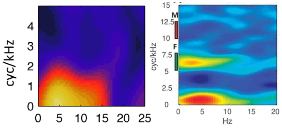
[Ponsot et al., 2018a, 2018b]





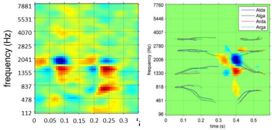
sentence recognition

[Venezia et al., 2016, 2019]

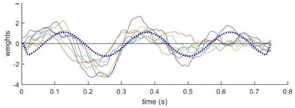


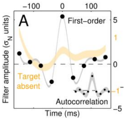
phoneme categorization

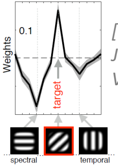
[Varnet et al., 2013, 2015]



modulation perception



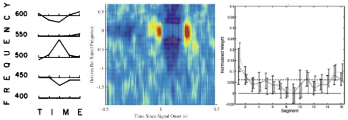




[Ponsot et al., 2020; Joosten & Neri, 2012; Varnet & Lorenzi, 2022]

pure-tone detection & loudness perception

[Ahumada & Lovell, 1971; Shub & Richards, 2009; Ponsot et al. 2013]

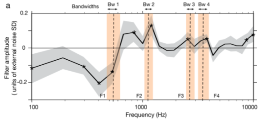


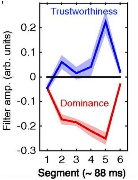
Auditory revcorr studies (full diagram on https://dbao.leo-varnet.fr/)

high level

paralinguistics

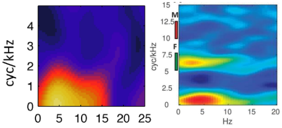
[Ponsot et al., 2018a, 2018b]





sentence recognition

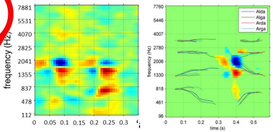
[Venezia et al., 2016, 2019]



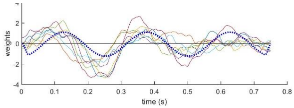
YOU ARE HERE

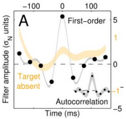
phoneme categorization

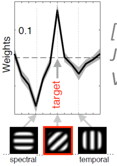
[Varnet et al., 2013, 2015]



modulation perception





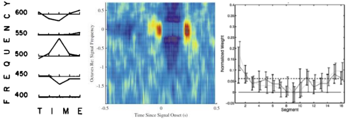


[Ponsot et al., 2020; Joosten & Neri, 2012; Varnet & Lorenzi, 2022]

low level

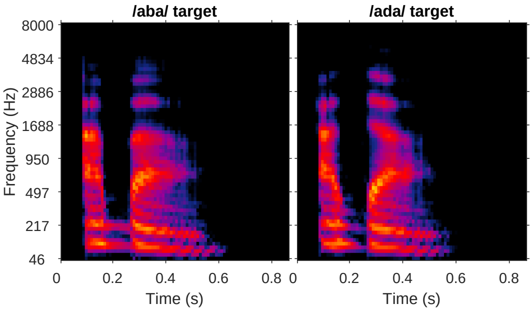
pure-tone detection & loudness perception

[Ahumada & Lovell, 1971; Shub & Richards, 2009; Ponsot et al. 2013]



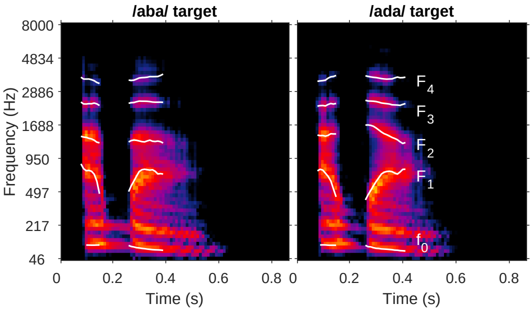
Aba/Ada experiment *[Osses & Varnet, submitted]*

Topic: perception of stop consonants /b/ and /d/.
Targets: 2 VCV sounds (t_0=/aba/ and t_1=/ada/) from the Oldenburg Logatome Corpus *[Wesker et al., 2005]*, equalized in duration and rms.



Aba/Ada experiment *[Osses & Varnet, submitted]*

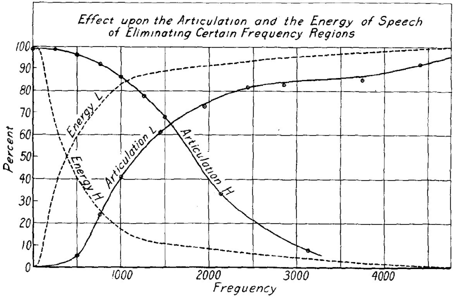
Topic: perception of stop consonants /b/ and /d/.
Targets: 2 VCV sounds (t_0=/aba/ and t_1=/ada/) from the Oldenburg Logatome Corpus *[Wesker et al., 2005]*, equalized in duration and rms.



Decoding speech

One solution: using **reduced speech**

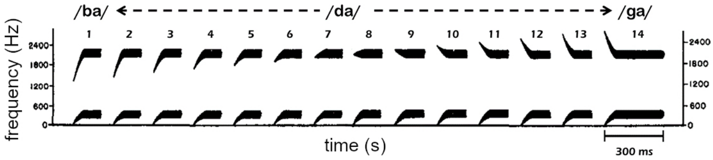
- Low-/high-pass filtered speech *[Fletcher, 1922]*



Decoding speech

One solution: using **reduced speech**

- Low-/high-pass filtered speech *[Fletcher, 1922]*
- Synthetic speech continuum *[Liberman et al., 1954]*

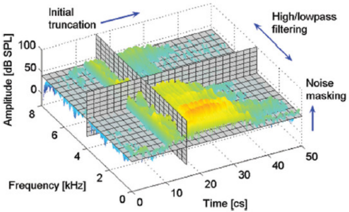


→ Proof that the **F2 onset** is a cue for categorizing /b/-/d/-/g/?

Decoding speech

One solution: using **reduced speech**

- Low-/high-pass filtered speech *[Fletcher, 1922]*
- Synthetic speech continuum *[Liberman et al., 1954]*
- 3-Dimensional Deep Search *[Li & Allen, 2012]*, etc...



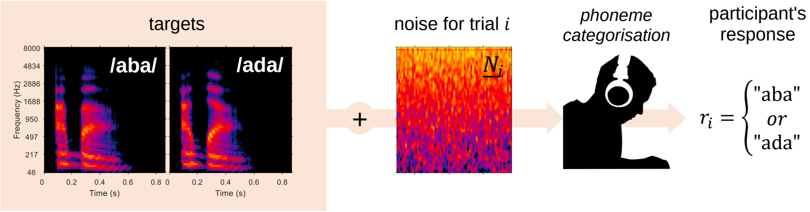
Decoding speech

One solution: using **reduced speech**

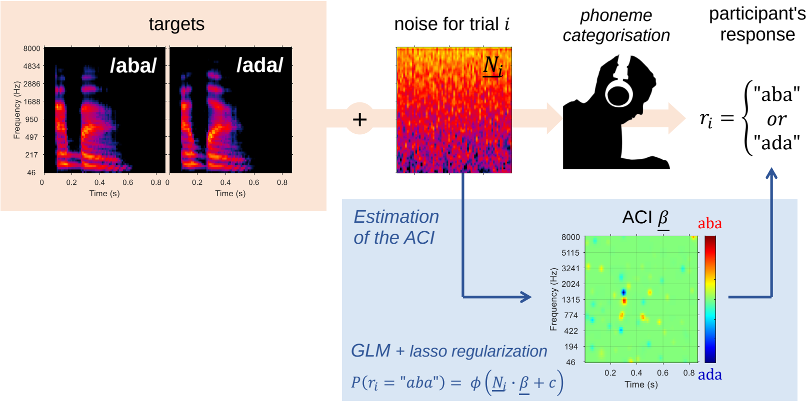
- Low-/high-pass filtered speech *[Fletcher, 1922]*
- Synthetic speech continuum *[Liberman et al., 1954]*
- 3-Dimensional Deep Search *[Li & Allen, 2012]*, etc...

Problem: the speech comprehension system shows very efficient **strategy adaptation**.

Aba/Ada experiment *[Osses & Varnet, submitted]*

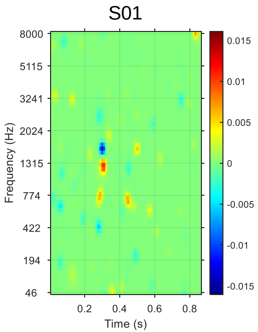


Aba/Ada experiment *[Osses & Varnet, submitted]*

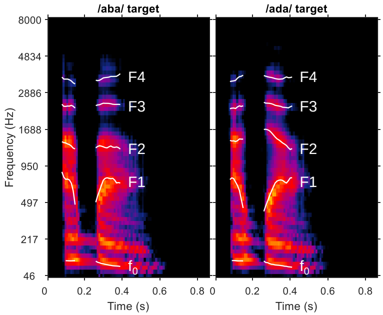


Auditory Classification Image (ACI): time-frequency matrix of decision weights. Shows how a specific noise configuration can mislead the participant.

Results

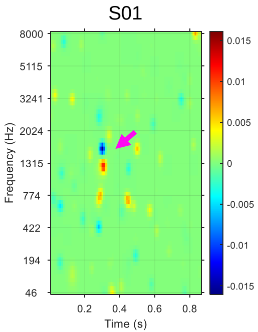


S01
Frequency (Hz)
8000
5115
3241
2024
1315
774
422
194
46
0.2
0.4
0.6
0.8
Time (s)
0.015
0.01
0.005
0
-0.005
-0.01
-0.015

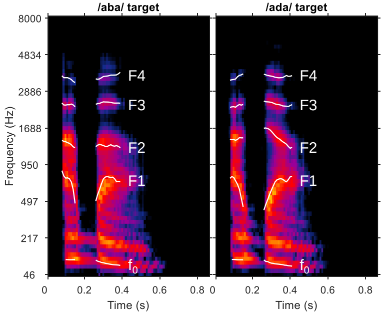


/aba/ target
/ada/ target
Frequency (Hz)
8000
4834
2886
1688
950
497
217
46
F4
F3
F2
F1
f_0
0
0.2
0.4
0.6
0.8
Time (s)

Results

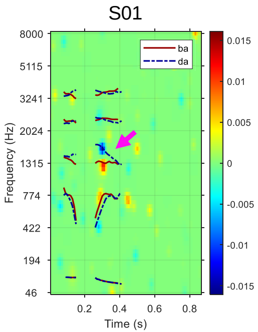


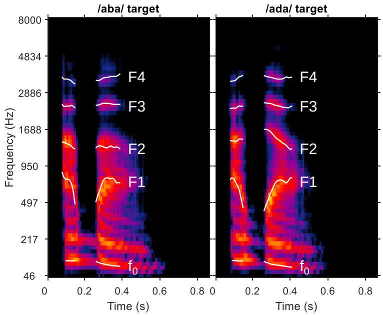
S01
8000
5115
3241
2024
1315
774
422
194
46
Frequency (Hz)
0.2
0.4
0.6
0.8
Time (s)
0.015
0.01
0.005
0
-0.005
-0.01
-0.015



/aba/ target
/ada/ target
8000
4834
2886
1688
950
497
217
46
Frequency (Hz)
F4
F3
F2
F1
f_0
0
0.2
0.4
0.6
0.8
Time (s)

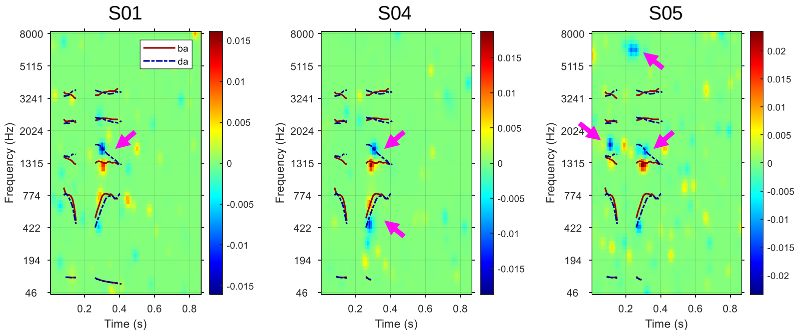
Results





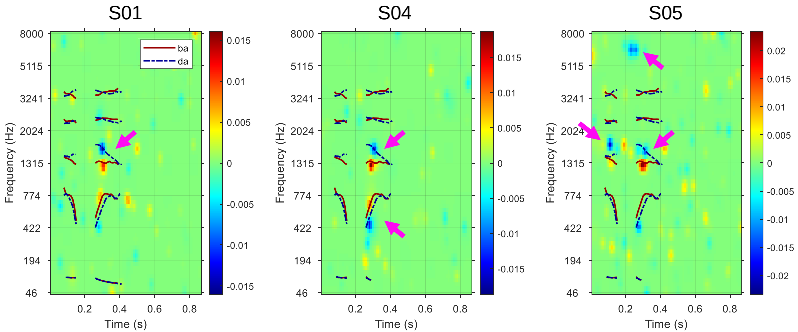
- The analysis successfully identified the **primary cue** for the task (F2 onset), consistent with the phonetics literature *[Liberman et al., 1954]*...

Results



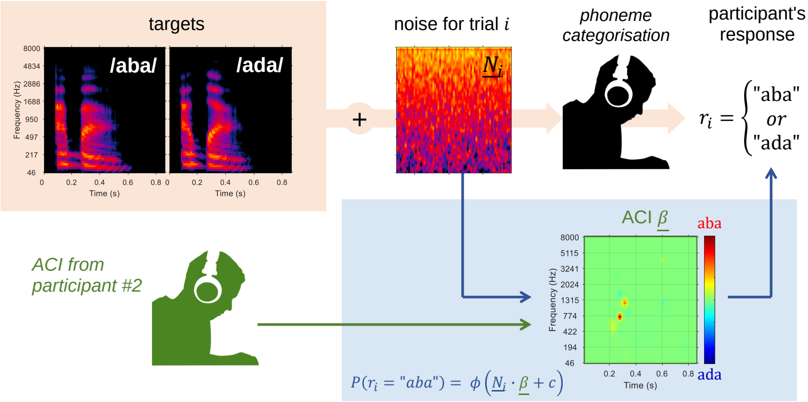
- The analysis successfully identified the **primary cue** for the task (F2 onset), consistent with the phonetics literature *[Liberman et al., 1954]*...
- ...as well as several **secondary cues** (e.g., F1 onset, coarticulation cue, burst).

Results



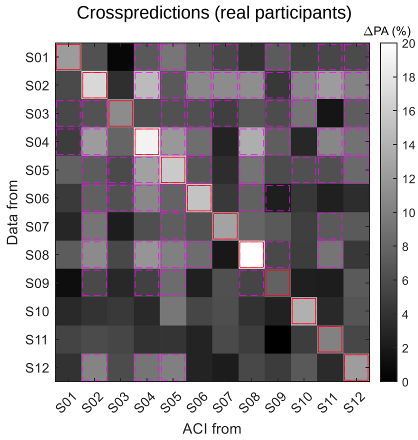
- The analysis successfully identified the **primary cue** for the task (F2 onset), consistent with the phonetics literature *[Liberman et al., 1954]*...
- ...as well as several **secondary cues** (e.g., F1 onset, coarticulation cue, burst).
- Contrary to our preregistered hypothesis, we observed some meaningful **interindividual variability** in the pattern of secondary cues.

Cross-predictions



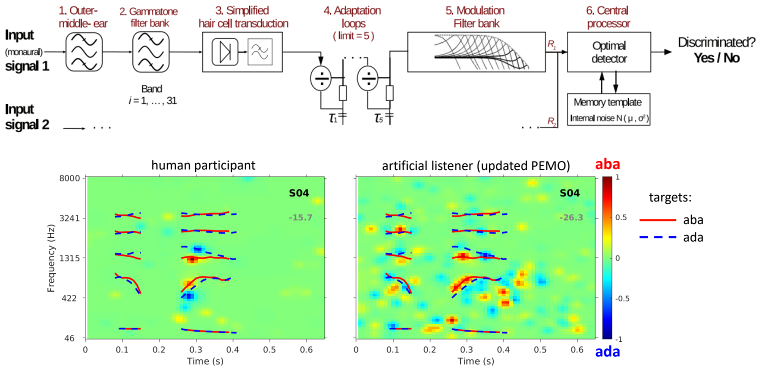
The similarity between listening strategies can be quantified by attempting to predict the responses of one participant using the ACI of another.

Cross-predictions



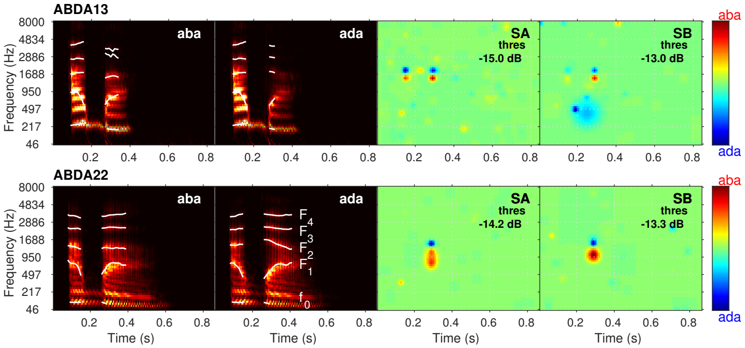
Our group of normal-hearing participants shows significant heterogeneity in their listening strategies: the ACI of one participant is better at predicting new data from this participant, rather than new data from another participant.

Artificial listener



An ideal observer based on a modulation filterbank front-end *[Osses & Kohlrausch, 2021]* obtains a (somewhat) different ACI on the same aba/ada categorization task.

Talker variability *[Varnet et al., 2013; Osses & Varnet, 2022]*



- Across several utterances (and types of noise), the primary cue is always found, but secondary cues are not.
- The revcorr method can reveal cues that are not present in the targets.

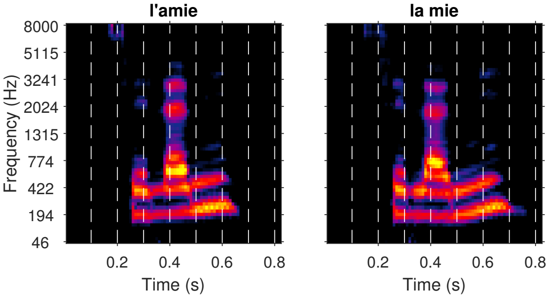
Current project

Identification of cues to place and voicing in French stop consonants.
Several new (preregistered) experiments will be ran in order to extend the dataset collected on /aba/-/ada/:

- /ada/-/aga/
- /apa/-/ata/
- /aba/-/apa/
- /ada/-/ata/
- (also /ata/-/aka/ and /aga/-/aka/ if there is time...)
- /ibi/-/idi/
- /ubu/-/udu/

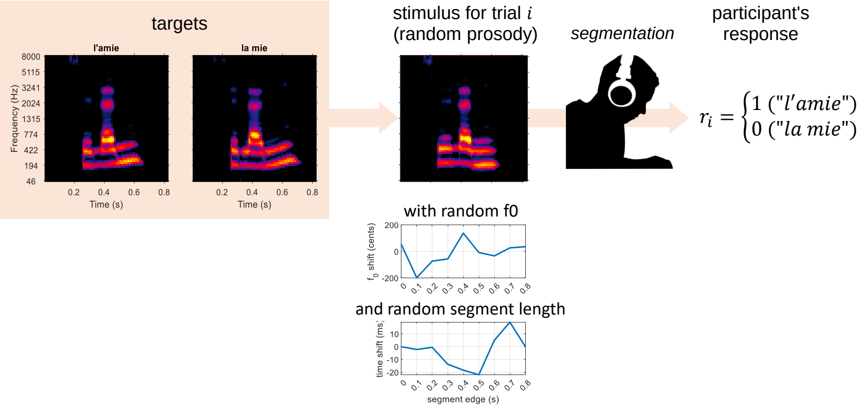
L'amie/La mie experiment *[Osses et al., in prep.]*

Topic: Acoustic cues for the segmentation of a speech sentence into words.
Targets: 2 phonetically identical sentences /selami/ (t_0= "c'est l'amie" and t_1= "c'est la mie"), equalized in duration and rms *[Spinelli al., 2010]*.

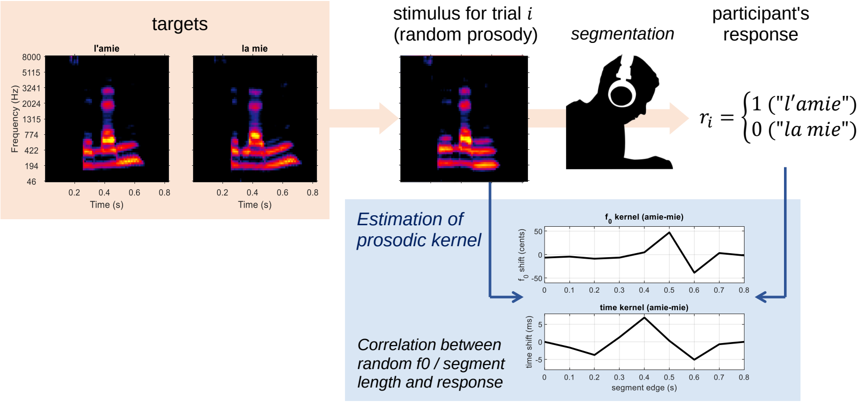


The targets were divided into 100-ms segments. Then, the f_0 trajectory in each segment was replaced by a random f_0 trajectory and each segment was compressed or stretched by a random amount, using WORLD *[Morise et al., 2016]*.

L'amie/La mie experiment *[Osses et al., in prep.]*



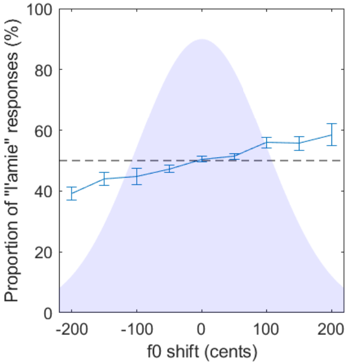
L'amie/La mie experiment *[Osses et al., in prep.]*



We obtain two kernels (f0 kernel and time kernel), indicating which aspects of the prosody are used as segmentation cues.

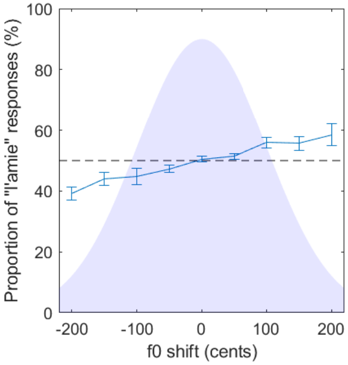
Results: effect of f_0 on /a/ segment

- There is a systematic effect of the random f_0 at 0.5 s (/a/ segment) on the response of participants.



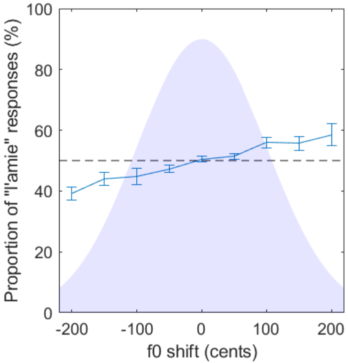
Results: effect of f_0 on /a/ segment

- There is a systematic effect of the random f_0 at 0.5 s (/a/ segment) on the response of participants.
- This is in line with the notion of "early f_0 rise" in French prosody: a high $f0$ on the first vowel of a content word is a cue for segmentation. *[Spinelli et al. 2010; Welby, 2007]*



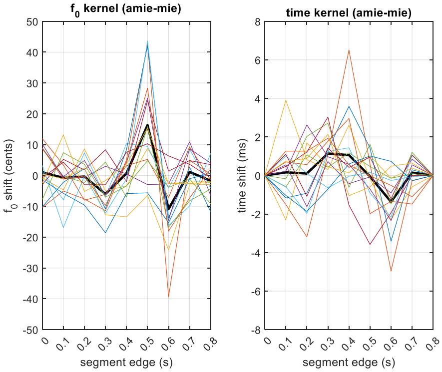
Results: effect of f_0 on /a/ segment

- There is a systematic effect of the random f_0 at 0.5 s (/a/ segment) on the response of participants.
- This is in line with the notion of “early f_0 rise” in French prosody: a high $f0$ on the first vowel of a content word is a cue for segmentation. *[Spinelli et al. 2010; Welby, 2007]*
- The strength of this relationship corresponds to the 0.5-s weight in the f_0 kernel.



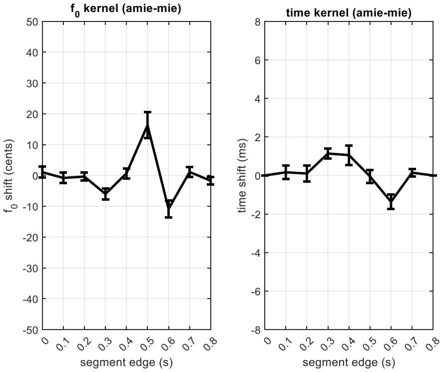
Results: prosody kernels

- Considerable variability at the group level (N=16)



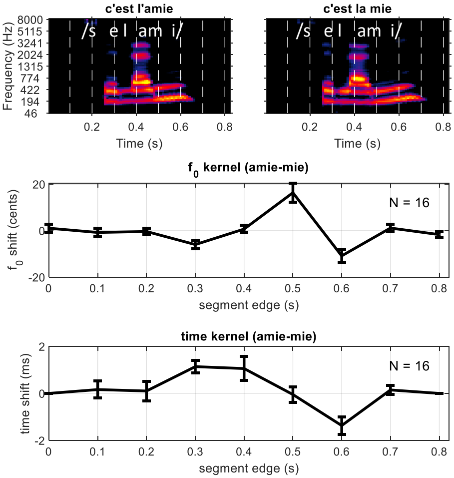
Results: prosody kernels

- Considerable variability at the group level (N=16)
- Nevertheless, a clear prosodic pattern emerges



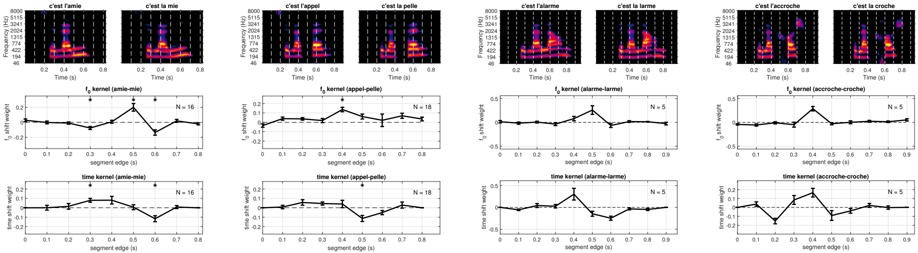
Results: prosody kernels

- Considerable variability at the group level (N=16)
- Nevertheless, a clear prosodic pattern emerges
- Two main segmentation cues can be identified: the relative f_0 of vowels /e/, /a/ and /i/ and relative duration of vowels /e/ and /i/.
- Again, consistent with previous findings obtained using manipulation-based methods *[Spinelli et al. 2010; Welby, 2007; Shoemaker, 2014]*



Results

- Similar results were found for other pairs of ambiguous stimuli.
- Secondary cues seem to depend on the phonetical content of the targets (e.g. duration of the second syllable).



Conclusions

- The **auditory revcorr approach** can reveal listening strategies and acoustic cues in psychoacoustic tasks (e.g., pitch perception, AM detection) and psycholinguistic tasks (e.g., phoneme discrimination, segmentation).

Code available on GitHub as an open-source MATLAB toolbox with documentation and turnkey experiments *[Osses & Varnet, 2021]*: `https://github.com/aosses-tue/fastACI`

Conclusions

- The **auditory revcorr approach** can reveal listening strategies and acoustic cues in psychoacoustic tasks (e.g., pitch perception, AM detection) and psycholinguistic tasks (e.g., phoneme discrimination, segmentation).
- It can also produce reliable results at the individual level, making it possible to explore **individual listening strategies**.

Code available on GitHub as an open-source MATLAB toolbox with documentation and turnkey experiments *[Osses & Varnet, 2021]*:
`https://github.com/aosses-tue/fastACI`

GitHub

Conclusions

- The **auditory revcorr approach** can reveal listening strategies and acoustic cues in psychoacoustic tasks (e.g., pitch perception, AM detection) and psycholinguistic tasks (e.g., phoneme discrimination, segmentation).
- It can also produce reliable results at the individual level, making it possible to explore **individual listening strategies**.
- Listening strategies for speech perception are complex and rely on the **weighted combination of many cues**.

Code available on GitHub as an open-source MATLAB toolbox with documentation and turnkey experiments *[Osses & Varnet, 2021]*: `https://github.com/aosses-tue/fastACI`

GitHub

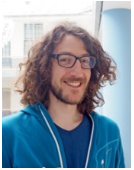
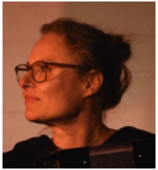
Thanks for your attention!

The "phonetic cues" team



Alejandro **Osses**, Géraldine **Carranante**, Maria **Giavazzi**

The "segmentation cues" team

Fanny **Meunier**, Étienne **Gaudrain**, Elsa **Spinelli**