Speech reductions cause a de-weighting of secondary acoustic cues

Léo Varnet [1]*, *Fanny Meunier* [2], *Michel Hoen* [3]

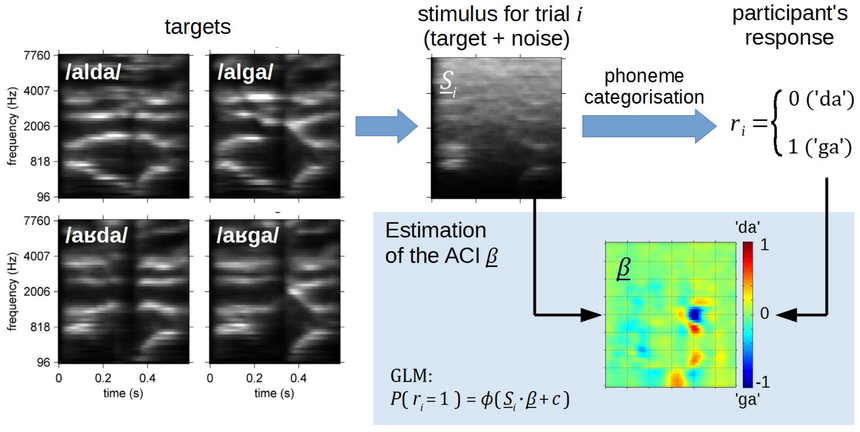
[1] Institut d'Étude de la Cognition, Laboratoire des Systèmes Perceptifs (LSP), CNRS UMR 8248, École Normale Supérieure (ENS), Paris, France.
[2] Laboratoire Base, Corpus, Langage (BCL), CNRS UMR 7320, Nice Sophia Antipolis, France.
[3] Oticon Medical, 2720 Chemin Saint Bernard, Vallauris, France.
*contact: leo.varnet@ens.fr

1. Introduction

- The ability of the auditory system to **change the perceptual weighting of acoustic cues when faced with degraded speech** has been evidenced in several studies. Their results suggest that listeners adapt their weighting strategy to selectively attend to the most reliable cues in a particular context (e.g. Lotto & Holt, 2011). However, the exact changes that occur in the listener's strategy remain mostly unknown.
- The recent development of **Auditory Classification Images (ACI)**, a psychoacoustical tool for studying speech perception, makes it possible to precisely identify the acoustic cues involved during a phoneme categorization task (Varnet et al., 2013, 2015).
- In the present study we applied the ACI technique to **compare the listening strategies of normal-hearing listeners in natural speech comprehension and in reduced** (i.e. noise-vocoded or re-synthesized) **speech comprehension**. This offered us a direct way to visualize the re-weighting of cues by the auditory system in degraded listening situations.

2. Methods (experiments 1a. and 1b.)

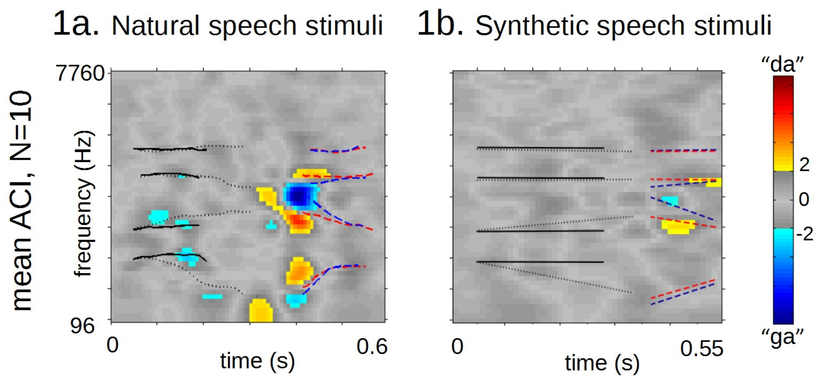
- **Stimuli:** 4 natural male speech productions of **/alda/, /alga/, /aʁda/ and /aʁga/** equated in duration and rms power (exp. 1a.) or re-synthesized versions of these sounds (exp. 1b.). All stimuli were presented in **random Gaussian noise**.
- **Participants: 20** normal-hearing listeners (10 in each experiment).
- **Task:** Each participant performed **10.000 phoneme categorizations**, indicating whether the **last syllable was /da/ or /ga/** (≈ 4 hours). The SNR was continuously adapted to ensure a correct response rate of 79%.



- **Data analysis:** The probability of "ga" answer is linked via a **Generalized Linear Model** (GLM) to the cochleogram of the presented stimulus (target+noise). The GLM is fitted by **Penalized Likelihood maximisation** with **smoothness prior**, a tradeoff between fitting the data well and obtaining a smooth ACI.
- The ACI (β) shows **how the presence of energy at each time-frequency point influences the decision** (i.e. which parts of the stimulus serve as cues for categorization). Positive clusters of weights correspond to regions favoring response "da", whereas negative clusters correspond to "ga" regions.

3. Results (experiments 1a. and 1b.)

- **The average SNR** (at the 79% correct response level) **was different** between the two experiments : -11.8 dB in exp. 1a.; 3.3 dB in exp. 1b.



Mean ACIs in exp. 1a. and 1b., displayed as Z-scores maps. Lines indicate formant positions in the targets (black dotted line: "al", black solid line: "ar", red line: "da", blue line: "ga").

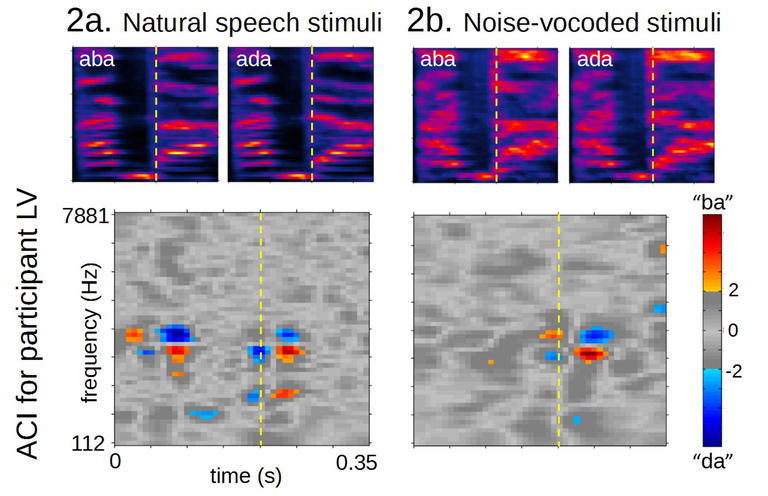
- The mean ACI obtained for experiment 1a. shows **strong clusters of weights on the expected primary cues** (F2 and F3 onsets in the second syllable). This experiment also revealed **secondary cues** on the F1 onset in the second syllable and on the F1 and F2 in the 1st syllable.
- In exp. 1b. **the primary cues for the task are preserved**, although weakened, as revealed by the ACI. However **no secondary cue reaches the significance threshold**. Indeed, while these cues may carry some coarticulatory information in exp. 1a., they are uninformative in exp. 1b.

This research was supported by an ERC grant to the SpiN Project (n°209234) and by an ANR grant to the SpeechCode Project (ANR-15-CE37-0009).

References:
- Lotto & Holt (2011).Psychology of auditory perception, Wiley Interdiscip. Rev. Cogn. Sci., 2(5), 479–489.
- Varnet, Knoblauch, Meunier, and Hoen (2013). Using auditory classification images for the identification of fine acoustic cues used in speech perception. Frontiers in Human Neuroscience, 7:865.
- Varnet, Knoblauch, Serniclaes, Meunier, and Hoen (2015). A psychophysical imaging method evidencing auditory cue extraction during speech perception: a group analysis of auditory classification images. PLoS ONE, 10(3).

4. Experiments 2a. and 2b.

- Same experimental setup, except that only 2 target sounds were used: **/aba/ and /ada/** (**same initial /a/**, followed by a recording of /ba/ or a recording of /da/, equalized in duration and rms normalized).
- In both exp. 2a. and 2b. targets were presented in **random Gaussian noise**. In exp. 2b. the targets were additionally **22-bands noise-vocoded**.
- **One single participant** ran both exp. 2a. and 2b. Its average **SNR was similar in the two sessions** (-13 dB).



ACI for participant LV in exp. 2a. and 2b., displayed as Z-scores maps. The beginning of the 2nd syllable is indicated with a yellow dotted line.

- The ACI for the natural speech condition (exp 2a.) revealed a primary cue on the F2 onset in the 2nd syllable, as well as secondary cues, among which an anticipatory cue in the 1st syllable. However, in the noise-vocoded condition (exp 2b.), only the primary cue is used.
- In both experiments, the **anticipatory cue is uninformative** for the task. Still, this cue is extracted in exp. 2a. but not in exp 2b.

5. Summary

These observations suggest that, **when they have to cope with reduced speech signal, listeners selectively focus on the primary cues by ignoring the secondary cues** (exp 1 and 2). Furthermore, this de-weighting seems not to depend on the actual reliability of the cue, but rather on its **expected change in informativeness** (exp 2).