Identification of functional acoustic cues involved in speech perception: recent advances using Auditory Classification Images.

Léo Varnet[1,5], Willy Serniclaes[2], Kenneth Knoblauch[3,5], Fanny Meunier[4,5], Michel Hoen[1,5]

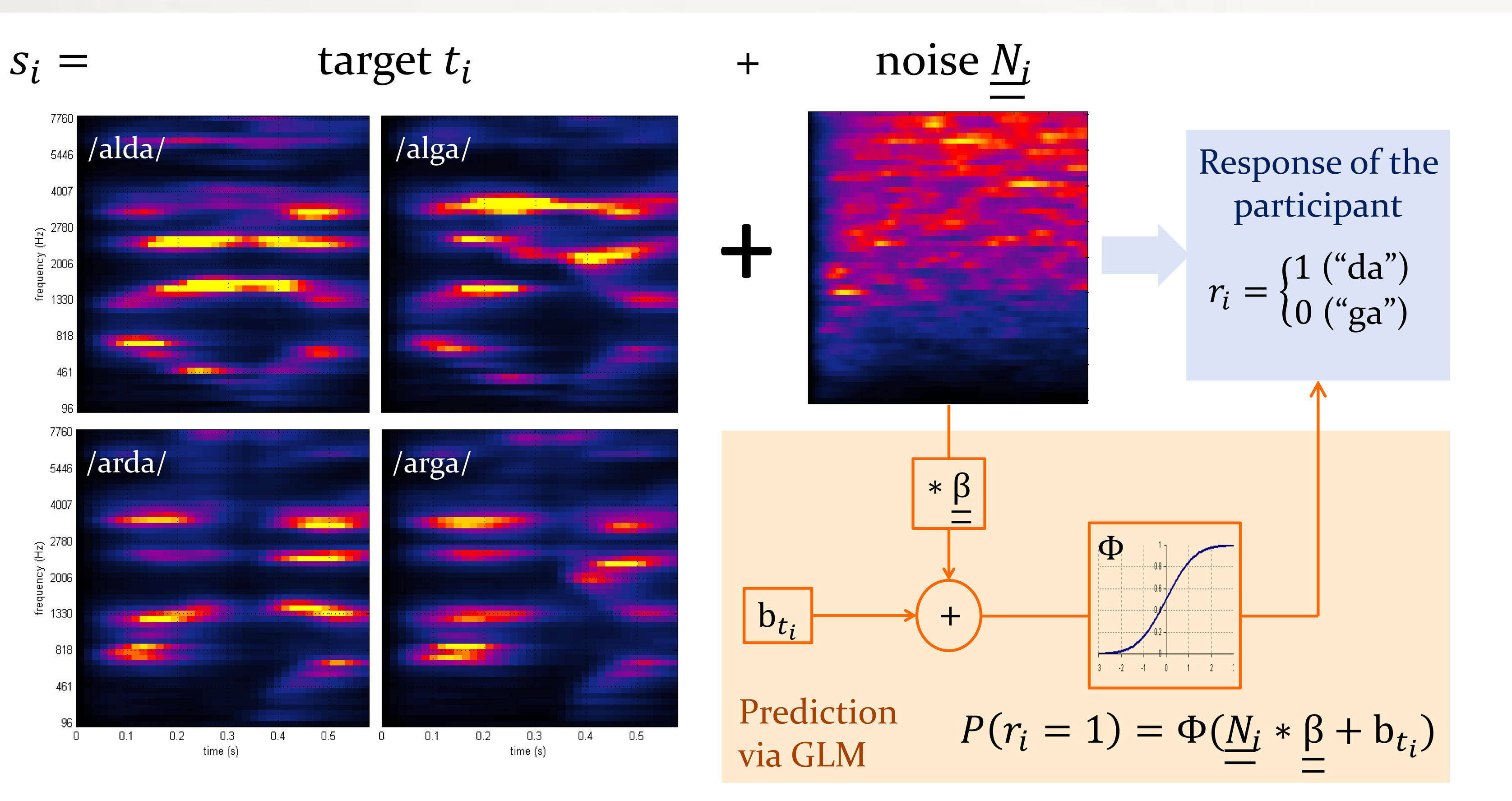
[1]Lyon Neuroscience Research Center, Lyon, France., [2]Université Libre de Bruxelles, Belgique., [3]Stem Cell and Brain Research Institute, Bron, France., [4]Laboratoire sur le Langage le Cerveau et la Cognition, Lyon, France., [5]Université de Lyon, Lyon, France.

1. Background

- Speech comprehension requires **mapping a complex acoustic signal** onto **categorical phonetic representations**. How the auditory system decomposes speech sounds into **elementary features** and combines them to form invariant representations of auditory objects still remains largely unknown.
- **Speech synthesis** and **filtered speech** have been used for the identification of acoustic cues (Delattre et al., 1955 ; Xu et al., 2005). However they require prior knowledge of the cues being sought, and the stimuli are highly unnatural.
- We developed a **psychoacoustic imaging method** for isolating acoustic cues from natural stimuli in a speech-in-noise situation (Varnet et al., 2013). Here we apply this **Auditory Classification Image (ACI)** technique to identify the listening strategies of a group of normal-hearing participants and compare their results to those of a group of participants with dyslexia.

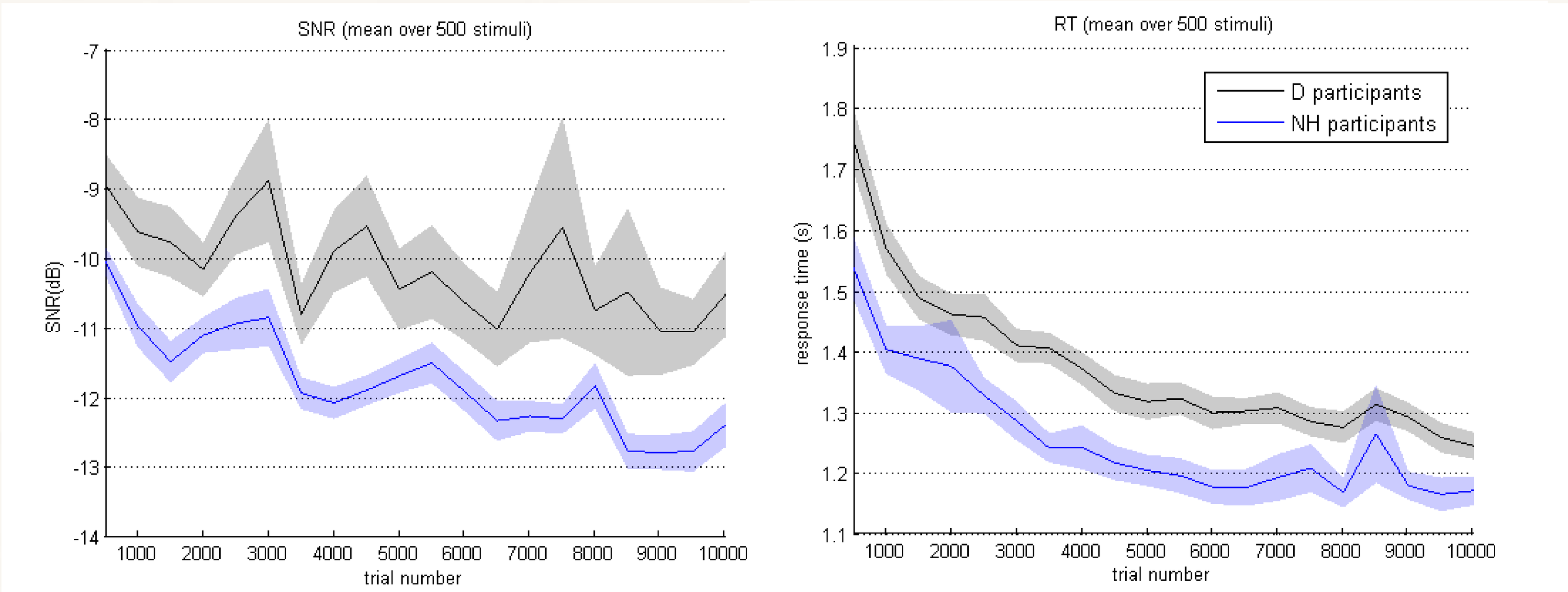
2. Materials and Methods

- **Stimuli**: 4 natural speech productions of **Alda**, **Alga**, **Arda**, and **Arga** equated in duration and root mean square. Presented in random Gaussian noise.
- **Task**: Each participant performed **10.000 phoneme categorizations** (20 sessions of 500 trials over 4 days), indicating whether the last syllable was /da/ or /ga/.
- SNR was adapted continuously to ensure a correct response rate of 79%.
- **Data Analysis:** The probability of "da" answer is linked via a **Generalized Linear Model** (GLM), to the time-frequency distribution of noise in each trial and the target actually presented. The ACI (β) shows how the presence of noise at each point **interferes with the decision** (= which parts of the stimulus serve as cues for categorization).
- The GLM is fitted by **Penalized Likelihood maximization** with **smoothness prior**, a tradeoff between fitting the data well and obtaining a smooth ACI.



3. Behavioral results

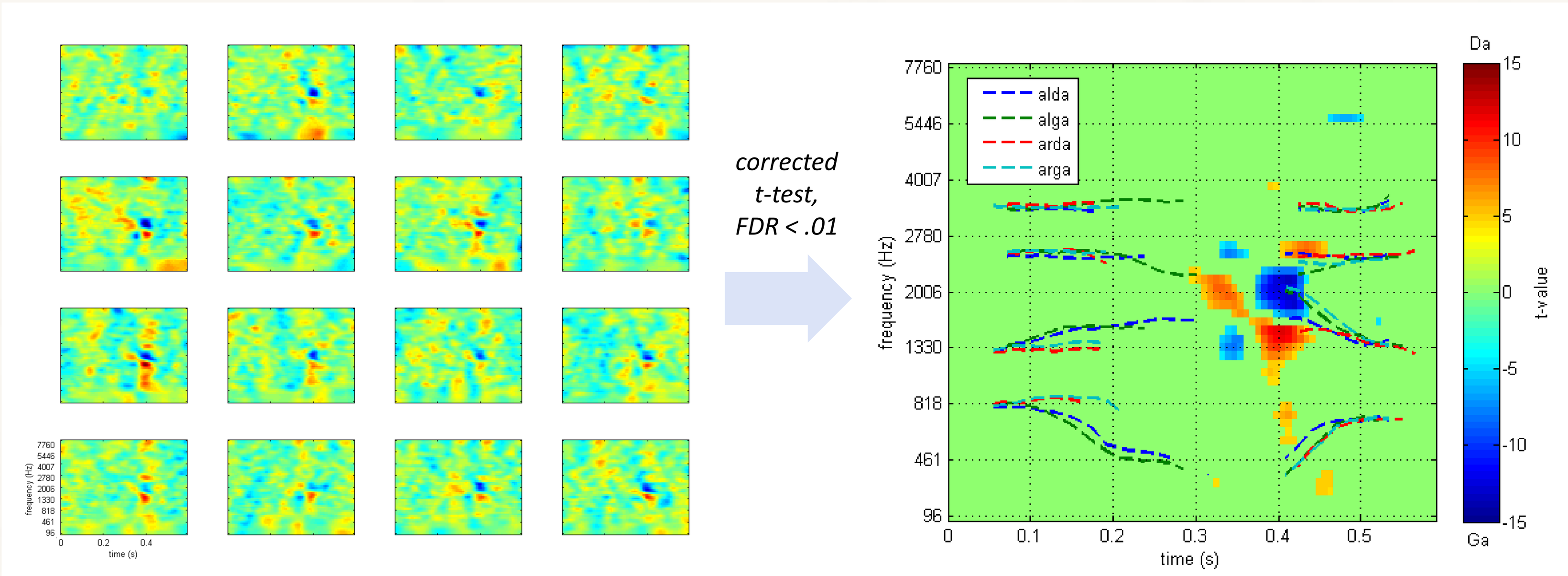
- **32 participants** (native French speakers). **Normal-hearing (NH) group**: **N=16** (aged 19-35, 12 women). **Dyslexic (D) group**: **N=16** (aged 18-44, 11 women).
- Mean correct response rate ≈**79%** throughout the experiment, for each participant.
- Significant effects of group and session number on SNR and Response Time, with no significant interaction effect (2-way ANOVAs, all $p<0.005$).



Evolution of mean SNR and mean RT by session (± s.e.m) over the course of the experiment

4. ACIs for normal-hearing participants

- One individual ACI is derived for each participant. High positive weights (in red) and high negative weights (in blue) are time-frequency locations where noise biases the response of the participant towards "da" or "ga", respectively.

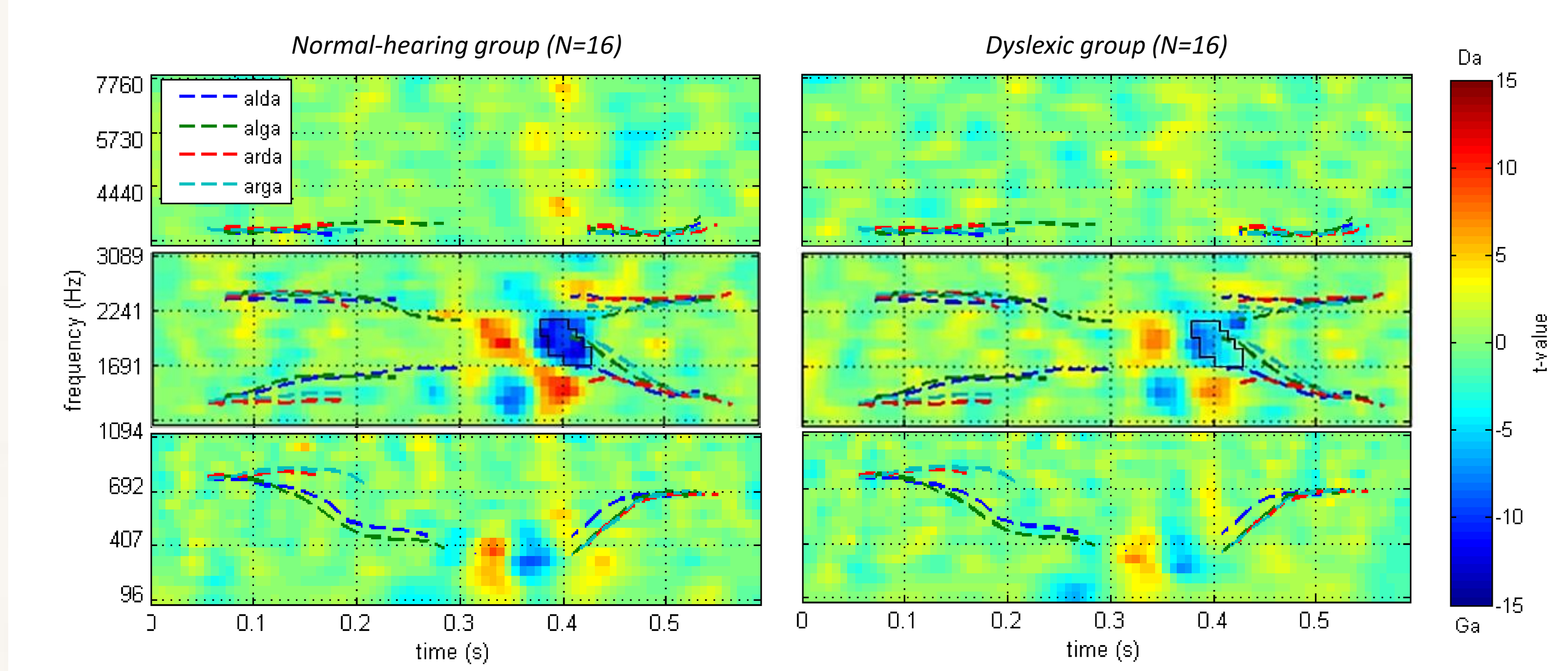


Individual ACIs for the 16 participants *Resulting "group ACI"* *(nonsignificant weights set to 0)*

- The group ACI shows significant clusters of weights on the **onsets of the F2 and F3 formantic transitions** (corrected t-test, FDR < .01) ⇒ these cues are critical for correct categorization of "da" and "ga", as already demonstrated in Stephens & Holt (2003) and Viswanathan et al. (2010).
- To reveal finer acoustic cues, we optimized separately 3 ACIs in the low, medium and high frequency bands (LF, MF and HF). Unexpectedly, a **secondary cue is found in the LF band**. Contrary to what previous studies using synthetic speech have assumed, this simple task involves the parallel processing of several spectral and temporal cues.

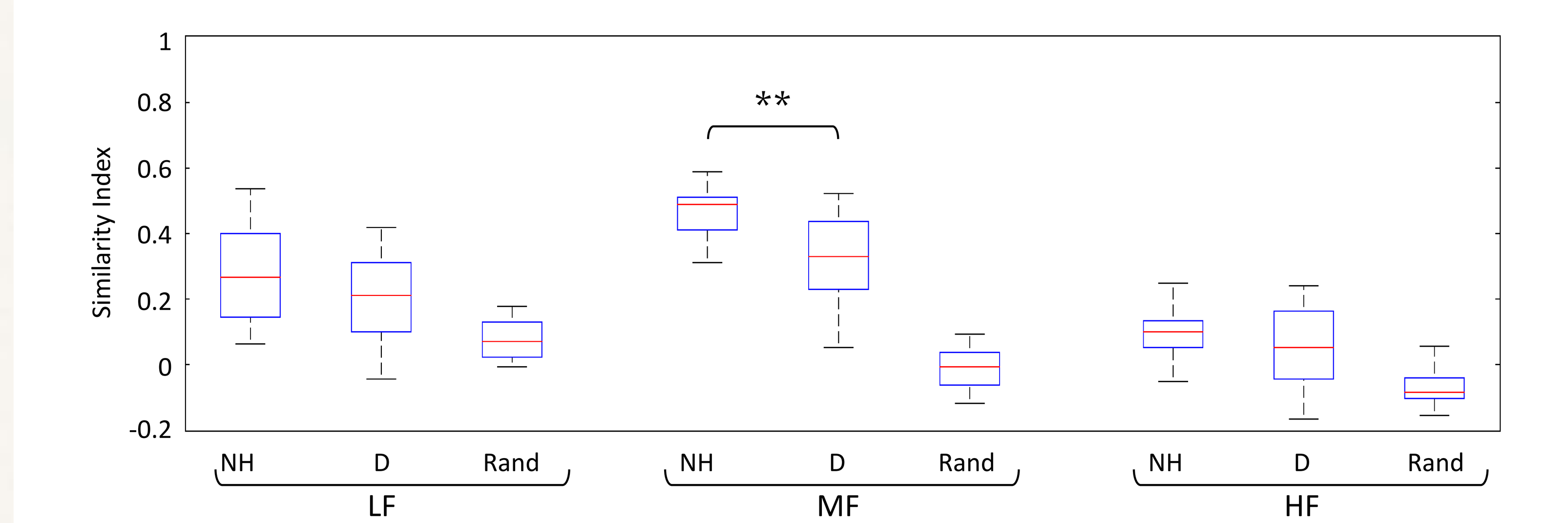
5. Comparison with dyslexic participants

- A very similar pattern of weights was obtained when deriving ACIs for participants with dyslexia. However, a cluster-based nonparametric test indicated a different weighting of the central negative cue.



Band-limited ACIs of the 2 groups. Weights differing significantly are outlined in black, p(cluster)<.05

- Lower t-values for the D group suggest that the individual ACIs are **less consistent** across dyslexic participants. An analysis of similarity between individual ACIs in both groups showed that the dyslexics' ACIs in MF are less similar to each other than those of the normal-hearing participants ($p<0.005$).



Similarity Index between ACIs in the low, medium or high frequency bands, within each group.

6. Conclusions

- The two groups of participants exhibit similar acoustic cues on F2 and F3 onsets for this categorization task. An unexpected temporal cue was found in the LF band.
- The patterns of weights inside dyslexics' ACIs are less coherent than those of normal-hearing participants. They seem to be more sensitive to the presence of noise masking non task-relevant portions of the stimulation, a marker of allophonic perception.
- Altogether these results demonstrate that ACI is a very promising method to study natural speech processing in NH participants and in the context of various speech or auditory perception deficits.

Contact: leo.varnet@isc.cnrs.fr. This research was supported by a European Research Council grant to the SpiN project (n°209234) and an Ecole Docotorale Neuroscience et Cognition (ED NSCo) PhD grant awarded to L.V.