New methodologies for studying listening strategies in phoneme categorization tasks

Léo Varnet

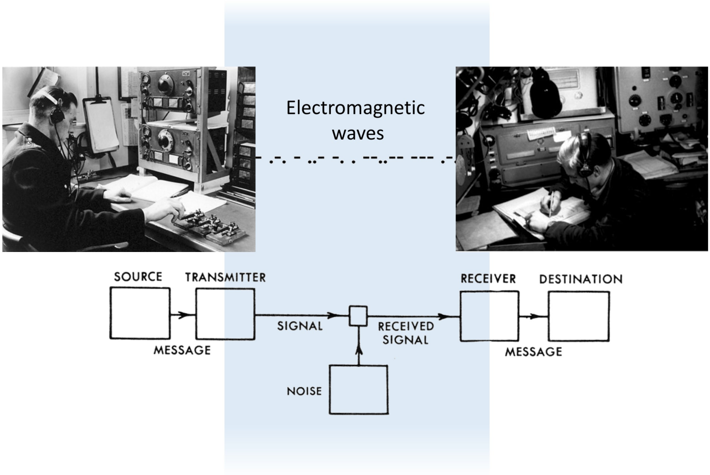
École Normale Supérieure Paris

Speech Science Forum, 18/10/2018

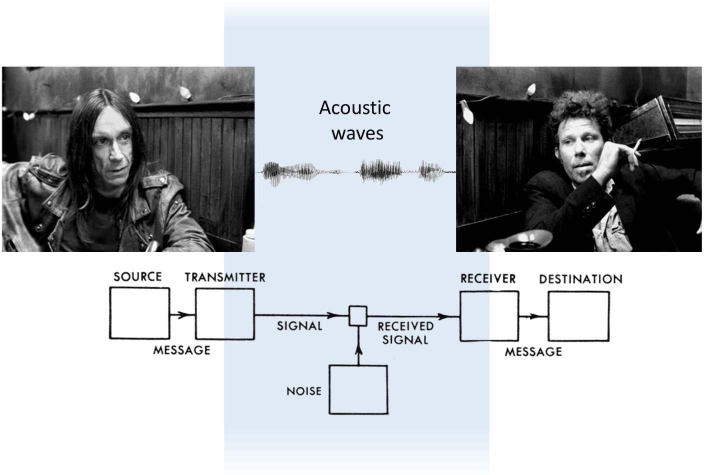
Plan

Decoding speech

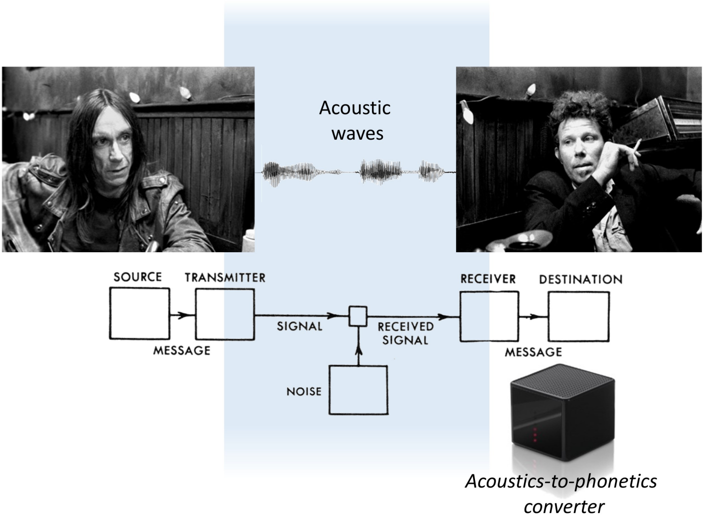
Decoding speech



Decoding speech



Decoding speech



Decoding speech

- Speech is a **complex code** (acoustics → phonetics).

Decoding speech

- Speech is a **complex code** (acoustics $\rightarrow$ phonetics).
- **Cracking the speech code**: finding the auditory primitives of speech comprehension.

Decoding speech

- Speech is a **complex code** (acoustics → phonetics).
- **Cracking the speech code**: finding the auditory primitives of speech comprehension.

Which **acoustic cues** allow the listener
to differentiate one phoneme from another?

Decoding speech

- Speech is a **complex code** (acoustics $\rightarrow$ phonetics).
- **Cracking the speech code**: finding the auditory primitives of speech comprehension.

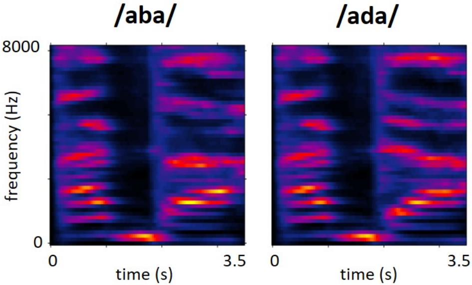
> Which **acoustic cues** allow the listener
> to differentiate one phoneme from another?

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

Decoding speech

How do we distinguish /ba/ from /da/?

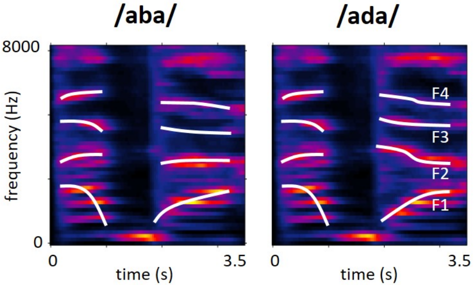
- Many acoustical differences



Decoding speech

How do we distinguish /ba/ from /da/?

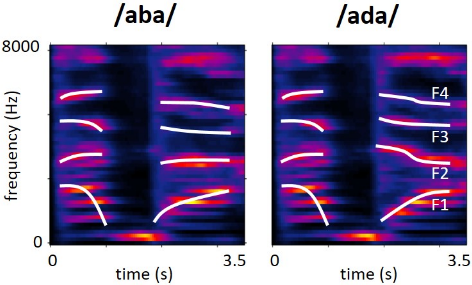
- Many acoustical differences (e.g. formant trajectories)



Decoding speech

How do we distinguish /ba/ from /da/?

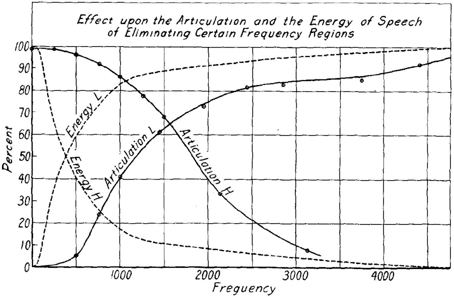
- Many acoustical differences (e.g. formant trajectories)
- Which ones are actually used by the auditory system ?



Decoding speech

One solution: using **reduced speech**

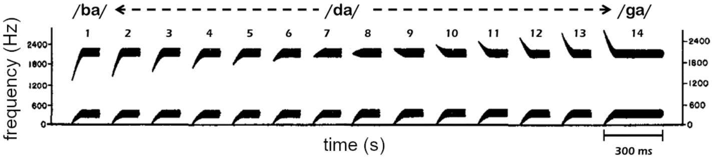
- Low-/high-pass filtered speech *(Fletcher, 1922)*



Decoding speech

One solution: using **reduced speech**

- Low-/high-pass filtered speech *(Fletcher, 1922)*
- Synthetic speech continuum *(Haskins in the 50's)*

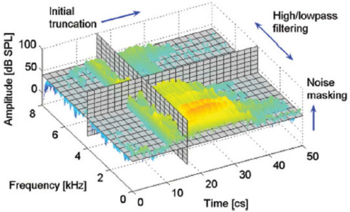


→ Proof that the **F2 onset** is a cue for categorizing /b/-/d/-/g/?

Decoding speech

One solution: using **reduced speech**

- Low-/high-pass filtered speech *(Fletcher, 1922)*
- Synthetic speech continuum *(Haskins in the 50's)*
- 3-Dimensional Deep Search *(Li & Allen, 2012)*, etc...



Decoding speech

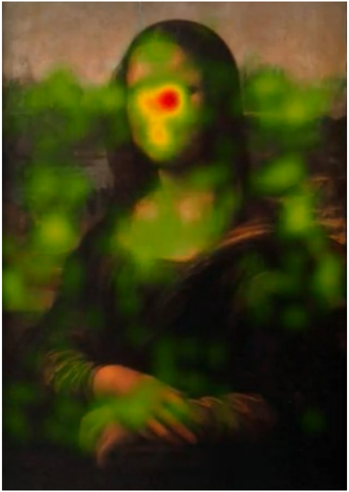
One solution: using **reduced speech**

- Low-/high-pass filtered speech *(Fletcher, 1922)*
- Synthetic speech continuum *(Haskins in the 50's)*
- 3-Dimensional Deep Search *(Li & Allen, 2012)*, etc...

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

The need for an **'ear-tracker'**

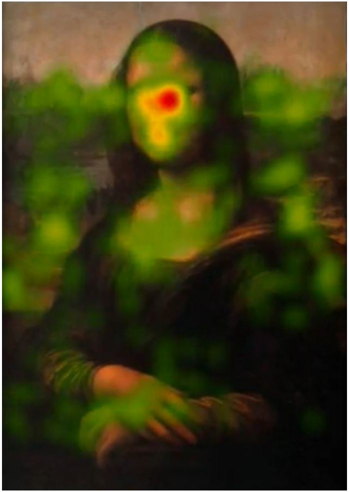
Developing a new method to visualize 'where' humans listen inside **natural** speech signals.



The need for an **‘ear-tracker’**

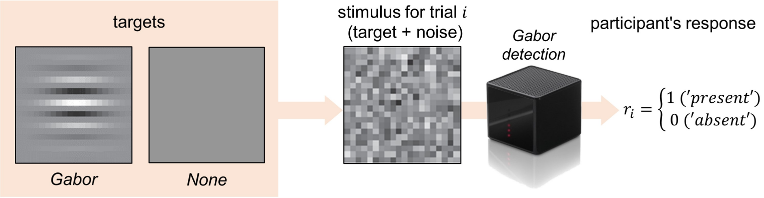
Developing a new method to visualize ‘where’ humans listen inside **natural** speech signals.

A solution could be provided by the technique of **Classification Images** (CI).



Visual Classification Images (CI)

Correlational technique *(Ahumada, 1971)* primarily used for applications in visual psychophysics. Example: visual detection of a Gabor target in noise.

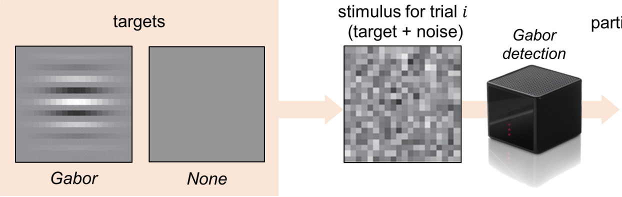


(Solomon, 2002)

Visual Classification Images (CI)

Correlational technique *(Ahumada, 1971)* primarily used for applications in visual psychophysics. Example: visual detection of a Gabor target in noise.

Which information is used to detect whether the target was present or not?

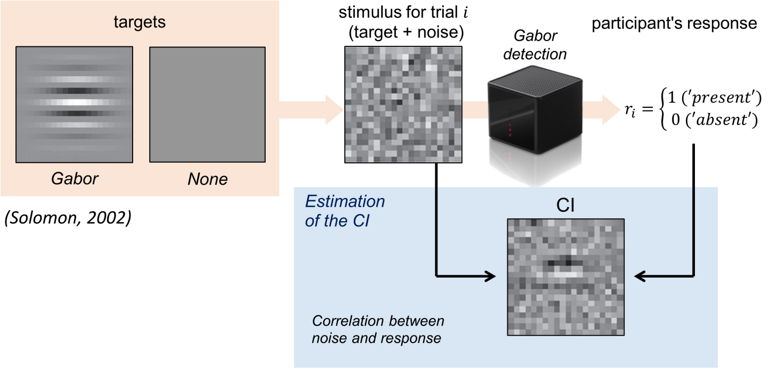


$$r_i = \begin{cases} 1 \ ('present') \\ 0 \ ('absent') \end{cases}$$

(Solomon, 2002)

Visual Classification Images (CI)

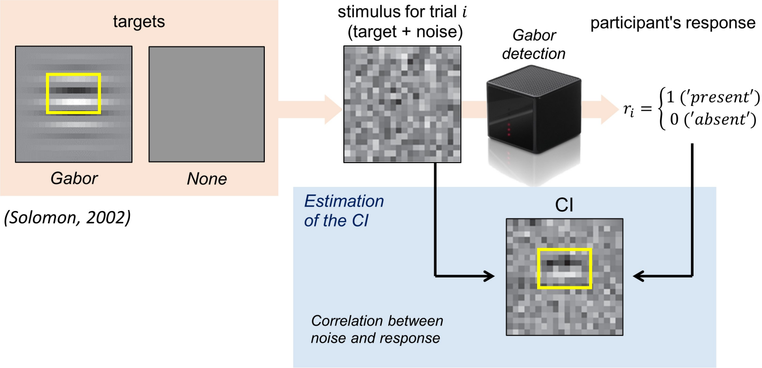
Correlation between the specific noise field in each trial and the response of the observer. The resulting correlation matrix shows how the presence of noise at each point interferes with the decision.



(Solomon, 2002)

Visual Classification Images (CI)

Correlation between the specific noise field in each trial and the response of the observer. The resulting correlation matrix shows how the presence of noise at each point interferes with the decision.



(Solomon, 2002)

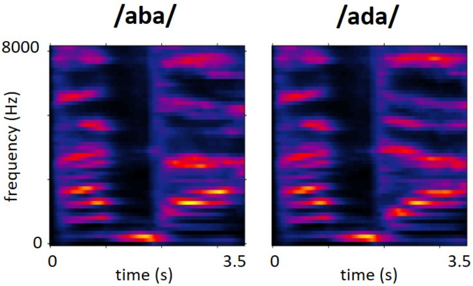
Aba/Ada experiment

Applying CI approach to the auditory modality

→ **Auditory Classification Images** (ACIs)

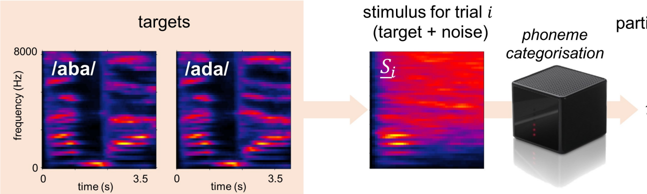
Materials

Targets: 2 speech sounds (t_0=/aba/ and t_1=/ada/) obtained by concatenating the same utterance of /a/ with two single utterances of /ba/ and /da/ (equalized in duration and rms).



Stimuli: Target sounds in an additive Gaussian noise.
Task: Indicate whether the target was /aba/ or /ada/.
SNR adapted continuously to ensure a correct response rate of 75%.

Methods

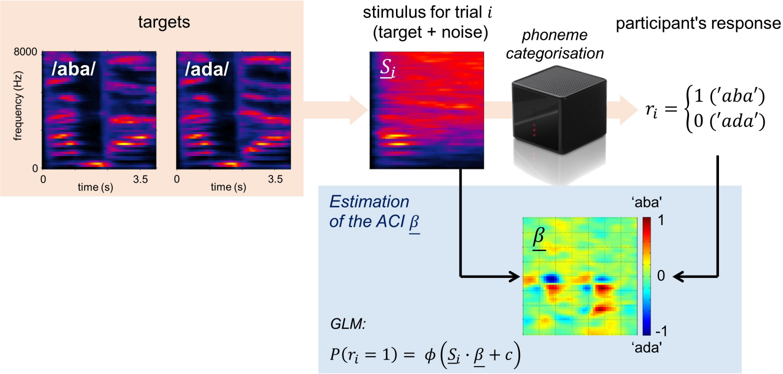


$$r_i = \begin{cases} 1 \ ('aba') \\ 0 \ ('ada') \end{cases}$$

Two major differences:

- Analysis based on time-frequency representations.
- Complexity of the speech targets.

Methods

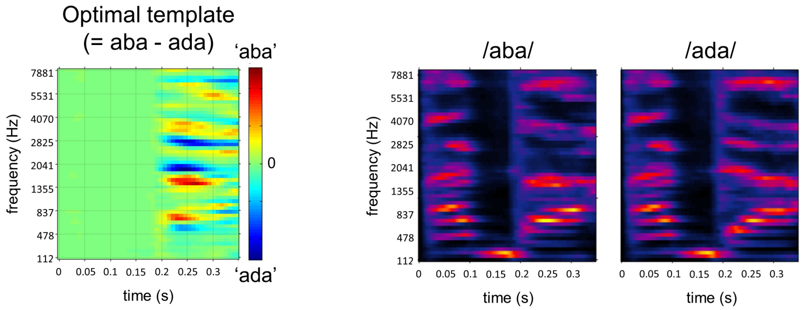


Generalized Linear Model (GLM):

- Works with arbitrary stimuli.
- Can be regularized to alleviate the overfitting problem.

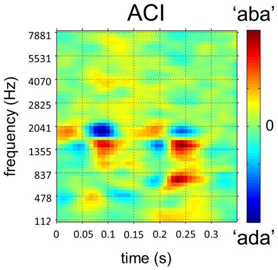
Ideal template

What would be the template used by an ideal observer performing the task linearly by template-matching?



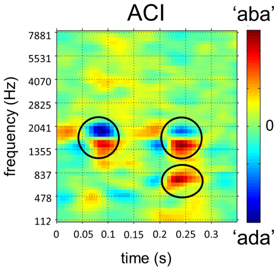
Real participant

- ACI does not look like the optimal template.



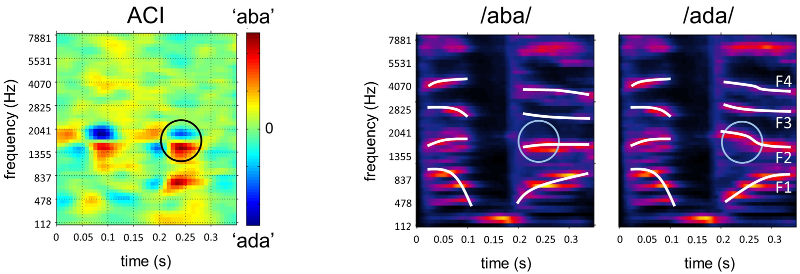
Real participant

- ACI does not look like the optimal template.
- Clusters of positive and negative weights corresponding to the acoustic cues (preceded with the opposite pattern of weights).



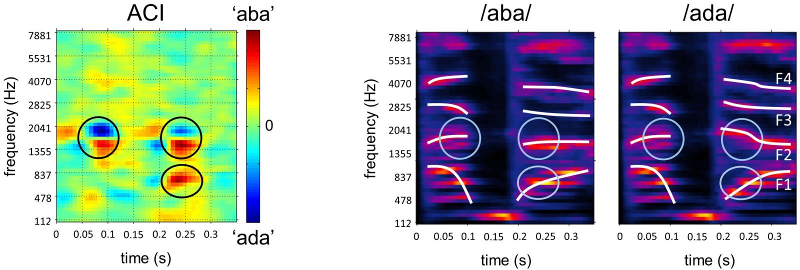
Real participant

- Confirms that **the F2 onset is a cue** for classifying phonemes into /b/ or /d/ *(Liberman et al., 1954)*.



Real participant

- Confirms that **the F2 onset is a cue** for classifying phonemes into /b/ or /d/ *(Liberman et al., 1954)*.
- Two unexpected cues.
- **Coarticulation cue** on the (uninformative) first syllable.



Aba/Ada experiment

- The method works fine!

Aba/Ada experiment

- The method works fine!
- Visualize what cues people listen to in natural speech signals (in noise)

Aba/Ada experiment

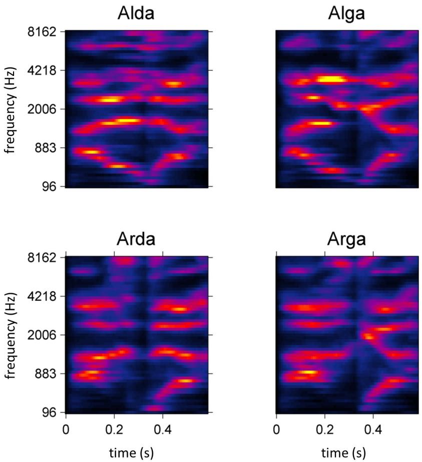
- The method works fine!
- Visualize what cues people listen to
  in natural speech signals (in noise)
- Can even reveal cues that are not present in the targets!

Aba/Ada experiment

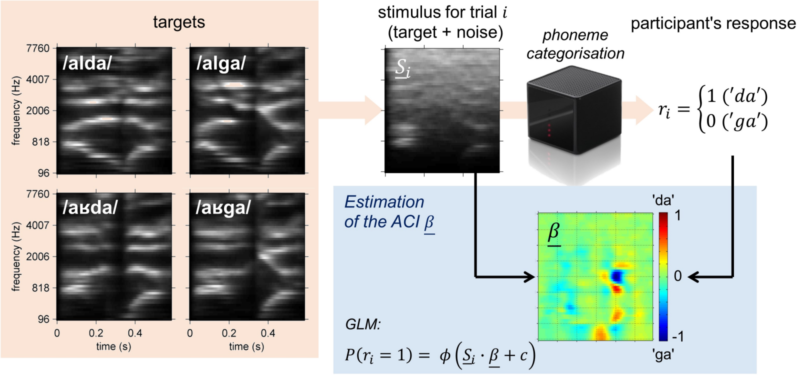
- The method works fine!
- Visualize what cues people listen to
  in natural speech signals (in noise)
- Can even reveal cues that are not present in the targets!
- … group-level ACIs?

Alda/Alga/Arda/Arga experiment

- **Participants**: **16** native French speakers. Each participant completed a set of 10.000 trials (20 sessions of 500 trials over 4 days).
- **Targets**: 4 CVVC sequences (/alda/-/alga/-/aʁda/-/aʁga/). Natural speech productions equated in duration and rms.
- **Task**: Indicate whether the last syllable was **/da/ or /ga/**.
- **Stimuli**: Targets in Gaussian noise. SNR was adapted continuously to ensure 79% correct response rate.

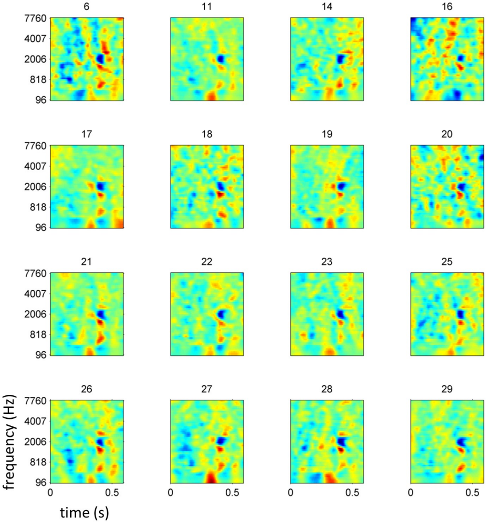


Methods


targets
7760
4007
2006
818
96
frequency (Hz)
/alda/
/alga/
/aʁda/
/aʁga/
0
0.2
0.4
time (s)
stimulus for trial i
(target + noise)
S_i
phoneme categorisation
participant's response
$r_i = \begin{cases} 1\ ('da') \\ 0\ ('ga') \end{cases}$
Estimation of the ACI β
β
'da'
1
0
-1
'ga'
GLM:
$P(r_i = 1) = \phi\left(\underline{S_i} \cdot \underline{\beta} + c\right)$

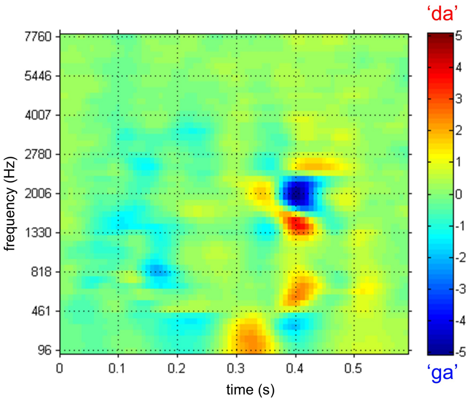
Group ACI

- Similar pattern of weights for all 16 participants.



Group ACI

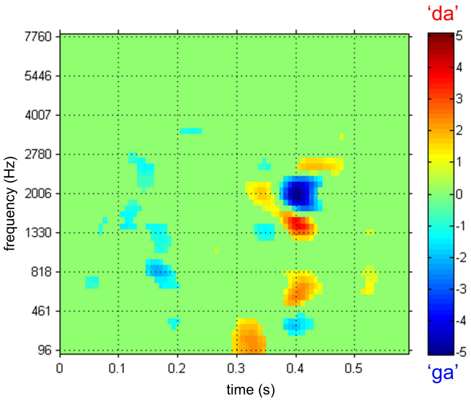
- Similar pattern of weights for all 16 participants.
- Primary cue: negative cluster surrounded by positive cluster.
- Other cues at lower frequencies.



Mean ACI over 16 participants

Group ACI

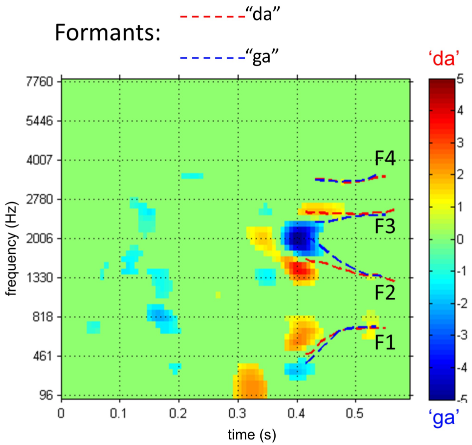
- Similar pattern of weights for all 16 participants.
- Primary cue: negative cluster surrounded by positive cluster.
- Other cues at lower frequencies.



Mean ACI over 16 participants
t-test against 0 with FDR correction (FDR$<$ .001)

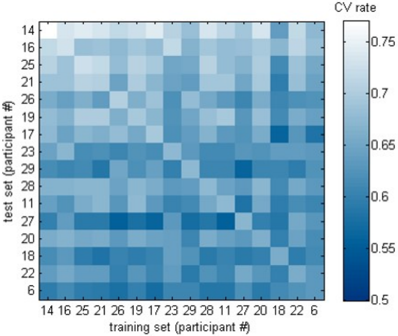
Group ACI

- The F2 and F3 onsets are critical cues for this task.
- The onset of F1 is also a cue for categorization.



Mean ACI over 16 participants
t-test against 0 with FDR correction (FDR$<$ .001)

Cross-validation

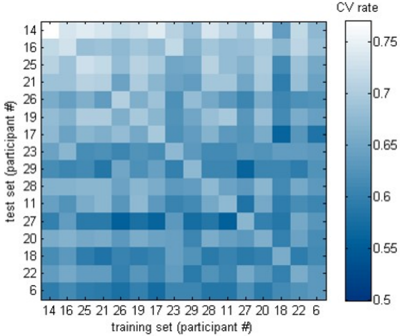


The ACI is a statistical model → **cross-validate**!

Within-participants *k*-fold cross-validation: CV rate ≈ 70%.

Between-participants *k*-fold cross-validation: CV rate ≈ 65%.

Cross-validation



The ACI is a statistical model → **cross-validate**!

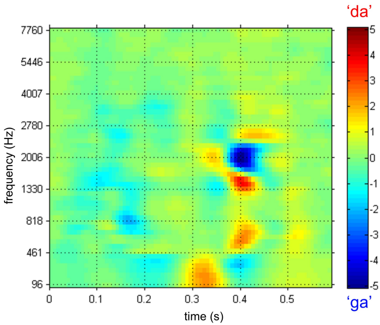
Within-participants *k*-fold cross-validation: CV rate ≈ 70%.

Between-participants *k*-fold cross-validation: CV rate ≈ 65%.

Evidence that normal-hearing individuals are using the same **listening strategy** for categorizing speech.

Blob noise

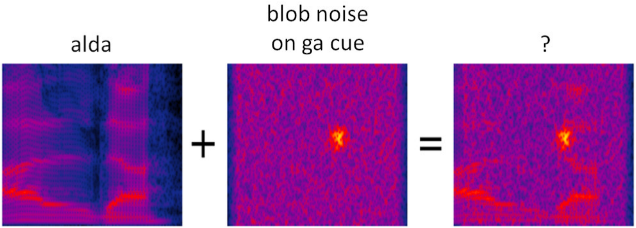
The clusters of weights on the ACI are regions where the presence of noise biases categorization toward /d/ or /g/.



→ What happens if we superimpose an additional bump of noise on the location of a cue previously identified?

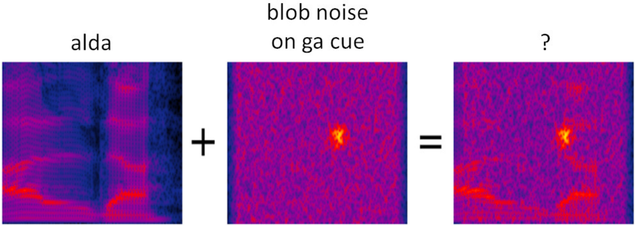
Blob noise

Blob noise: white noise with an additional bump of noise on the location of a cue previously identified.



Blob noise

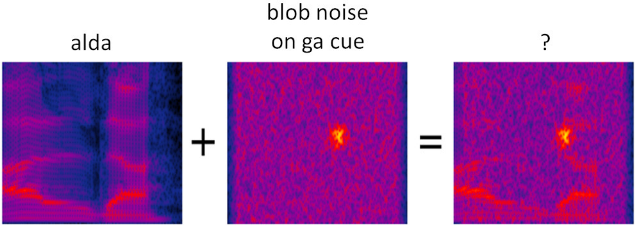
Blob noise: white noise with an additional bump of noise on the location of a cue previously identified.



→ a noise that shifts perception from da to ga (or from ga to da)!

Blob noise

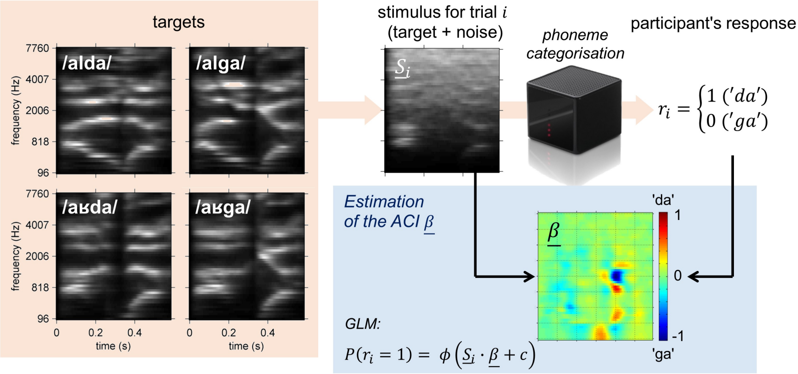
Blob noise: white noise with an additional bump of noise on the location of a cue previously identified.



→ a noise that shifts perception from da to ga (or from ga to da)!

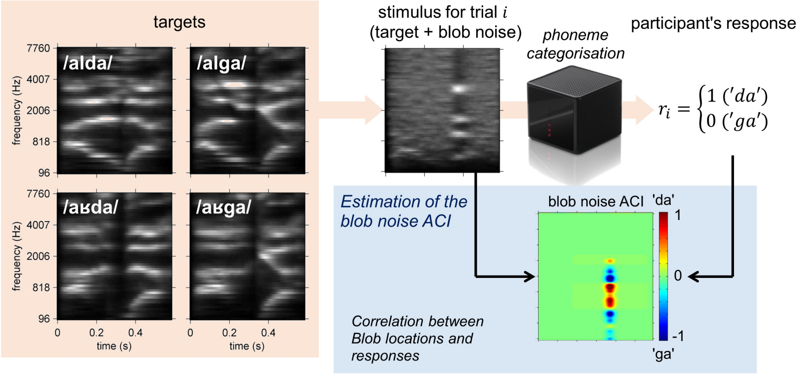
First application: 'blob noise ACI'.

ACI method



Painful experiment (10.000 trials ≈ **4h** of da/ga in noise)

Blob noise ACI method

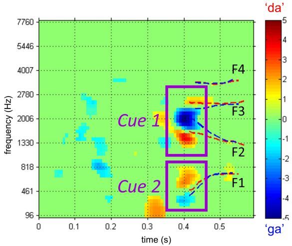


Blob noise ACI: ACI calculated using **random blob noises**.
→ reduces the number of trials required to ≈ 1000!

Interim summary

The acoustic-to-phonetic conversion is a complex process

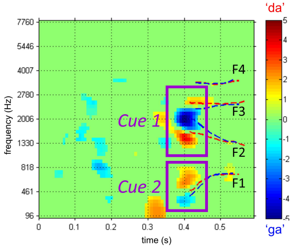
- involving **multiple cues**
- some of which may be **anticipatory**
- cues are associated with **different weights** in the decision (e.g. primary F2/F3 cue vs. secondary F1 cue)



Interim summary

The acoustic-to-phonetic conversion is a complex process

- involving **multiple cues**
- some of which may be **anticipatory**
- cues are associated with **different weights** in the decision (e.g. primary F2/F3 cue vs. secondary F1 cue)



- The ACI gives an insight into the black box.

How consistent are those strategies across listeners / groups / listening conditions?

Comparing groups of listeners

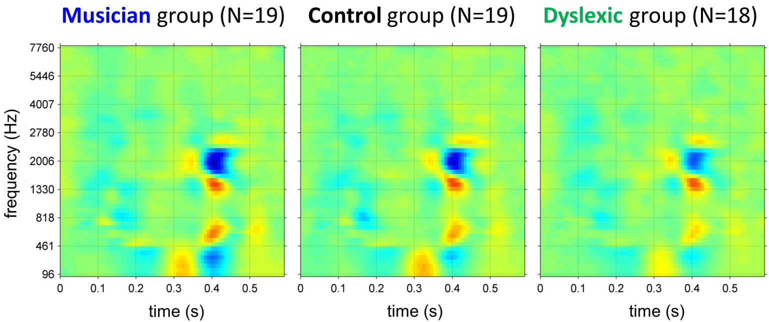
- **Musicians** are better than non-musicians at understanding speech in noise.
- **Dyslexics** have impaired speech in noise comprehension.

→ Do they use different cues?

Comparing groups of listeners

- **Musicians** are better than non-musicians at understanding speech in noise.
- **Dyslexics** have impaired speech in noise comprehension.

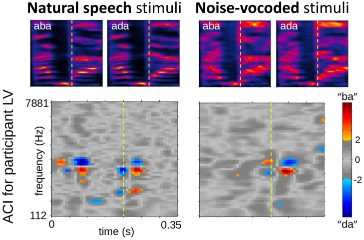
→ Do they use different cues?



Comparing listening conditions

Replicating the /ba/-/da/ and /da/-/ga/ experiments with reduced speech stimuli:

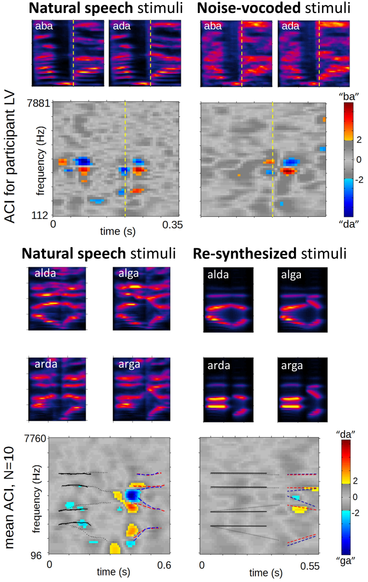
- Noise-vocoded stimuli



Comparing listening conditions

Replicating the /ba/-/da/ and /da/-/ga/ experiments with reduced speech stimuli:

- Noise-vocoded stimuli
- Re-synthesized stimuli

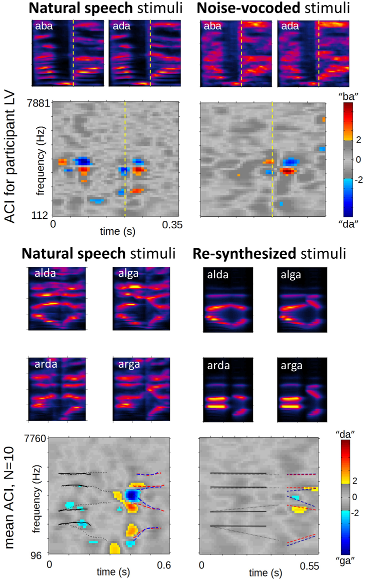


Comparing listening conditions

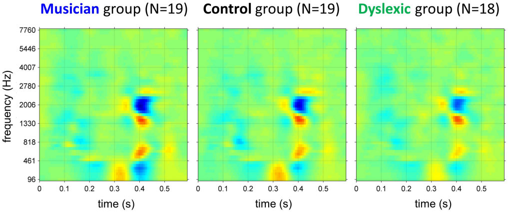
Replicating the /ba/-/da/ and /da/-/ga/ experiments with reduced speech stimuli:

- Noise-vocoded stimuli
- Re-synthesized stimuli

→ How does the auditory system adapt to speech reductions?

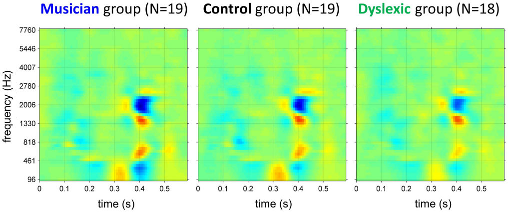


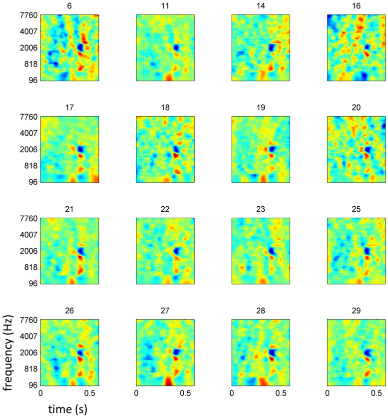
Cue-weighting variability



- **Between groups**: cue-weighting differences, while the overall strategy remains the same.

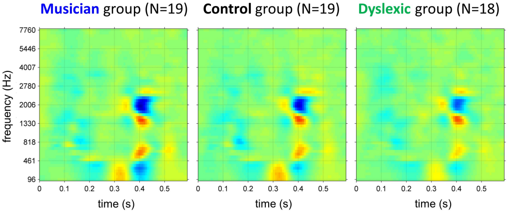
Cue-weighting variability

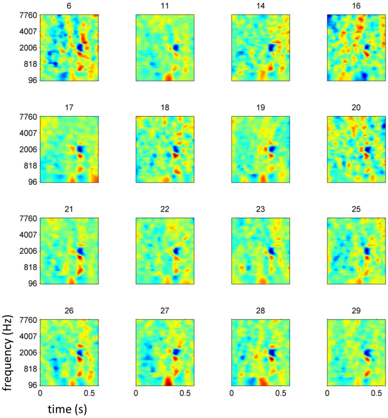




- **Between groups**: cue-weighting differences, while the overall strategy remains the same.
- **Within groups**: all participants appear to use the same cues... But do they have the same weighting strategy?

Cue-weighting variability

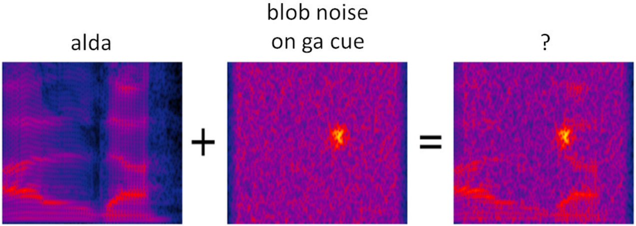




- **Between groups**: cue-weighting differences, while the overall strategy remains the same.
- **Within groups**: all participants appear to use the same cues... But do they have the same weighting strategy?

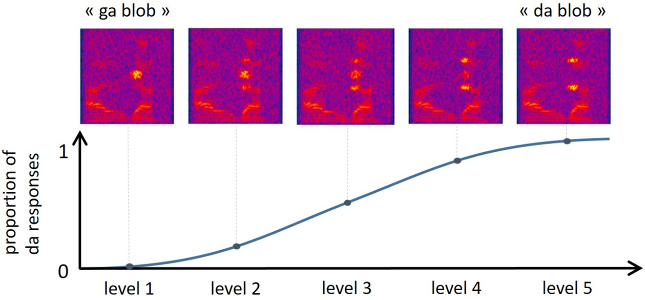
Quantify cues weightings, at the individual and group levels?

Parametric blob noise paradigm

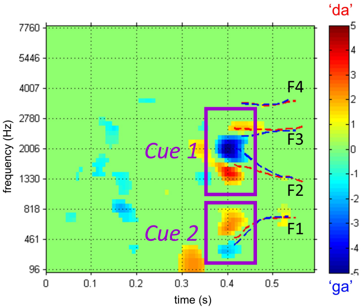


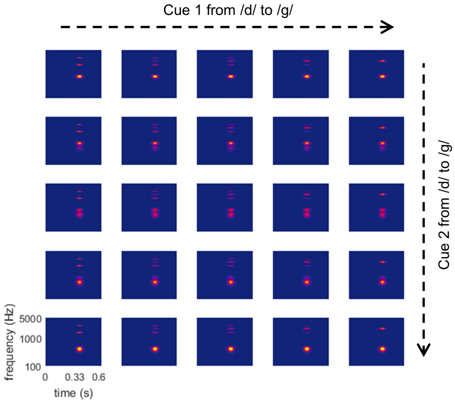
Parametric blob noise paradigm

By varying the energy of the blobs and measuring the proportion of confusions, we should be able to estimate the **sensitivity** of a listener to the corresponding cue.



Parametric blob noise paradigm

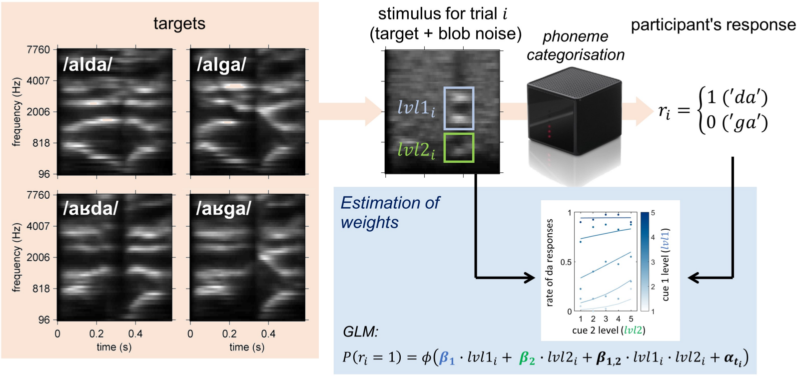




Joint measurement of the weightings of **two separate cues**.

→ **2-dimensional** /da/-/ga/ continuum.

Parametric blob noise paradigm



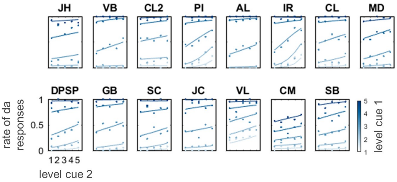
1000 trials per participant (4 target X 25 blob noises X 10 repetitions).

β_1 and β_2: participant's weights on cue 1 and cue 2.

→ Strong effect of primary cue; weak effect of secondary cue.

Cue-weighting strategies

17 **Normal-hearing** (NH) participants



NH
(N=17)

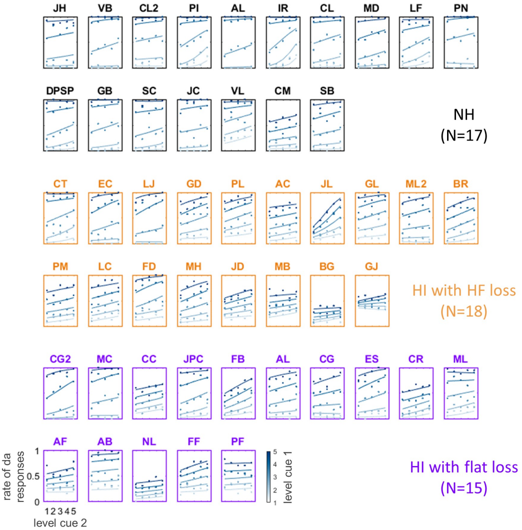
Cue-weighting strategies

17 **Normal-hearing** (NH) participants

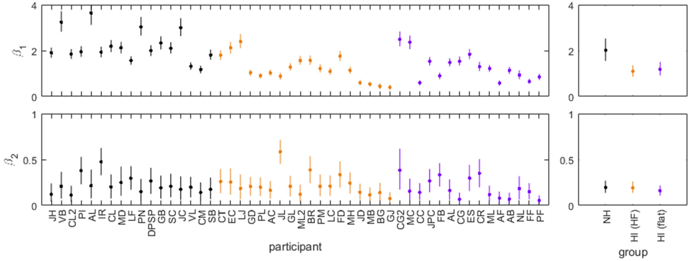
18 **Hearing-impaired** (HI) participants with **high-frequency** loss

15 **Hearing-impaired** (HI) participants with **flat** loss

Audibility restored with simulated hearing aid.



Cue-weighting strategies

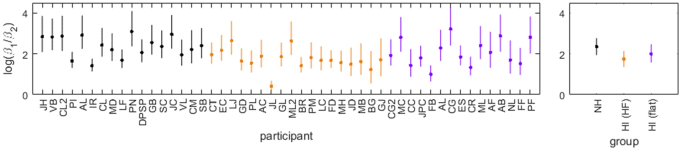


3 groups:

- **NH**
- **HI (HF loss)**
- **HI (flat loss)**

Audibility restored with simulated hearing aid.

Cue-weighting strategies



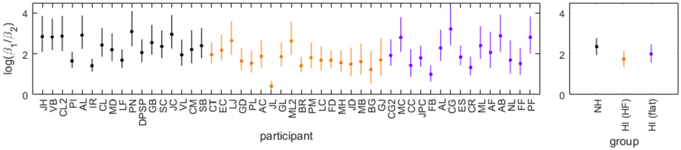
3 groups:

- **NH**
- **HI (HF loss)**
- **HI (flat loss)**

Audibility restored with simulated hearing aid.

log sensitivity ratio $log(\beta_1/\beta_2)$: relative importance of cue 1 and cue 2 in the decision.

Cue-weighting strategies



3 groups:

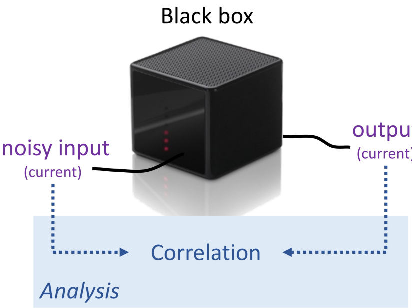
- **NH**
- **HI (HF loss)**
- **HI (flat loss)**

Audibility restored with simulated hearing aid.

log sensitivity ratio $log(\beta_1/\beta_2)$: relative importance of cue 1 and cue 2 in the decision.

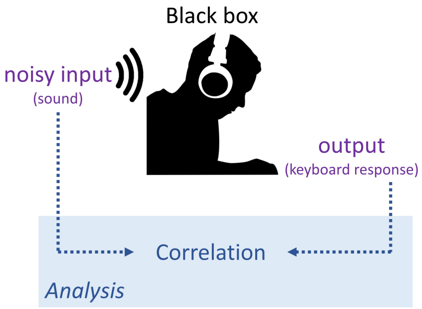
HI (HF loss) participants have a different cue-weighting strategy, even though their hearing loss was corrected through amplification.

Conclusion: using noise to characterize a black box



- Wiener kernel analysis *(Wiener, 1958)*: characterizing an **electrical circuit** by giving it a white noise input and measuring correlations between its input and output.

Conclusion: using noise to characterize a black box



- Wiener kernel analysis *(Wiener, 1958)*
- Auditory Classification Images: characterizing the **auditory system** by giving it a noisy input and measuring correlations between its input and output.

Thanks for your attention!

And thanks to:

Christian **Lorenzi**, Christophe **Micheyl**, Chloé **Langlet**

Michel **Hoen**, Kenneth **Knoblauch**, Gwendoline **Trollé**, Chloé **Peter**, Fanny **Meunier**