New methodologies for studying listening strategies in phoneme categorization tasks

Léo Varnet

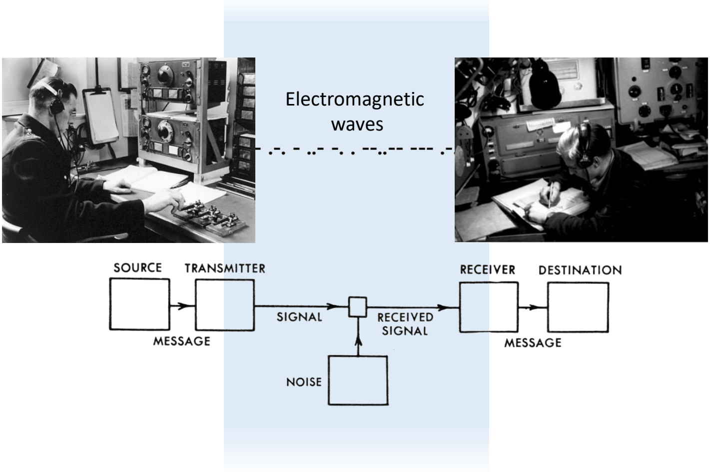
École Normale Supérieure Paris

Institut de l'Audition, 27/05/2021

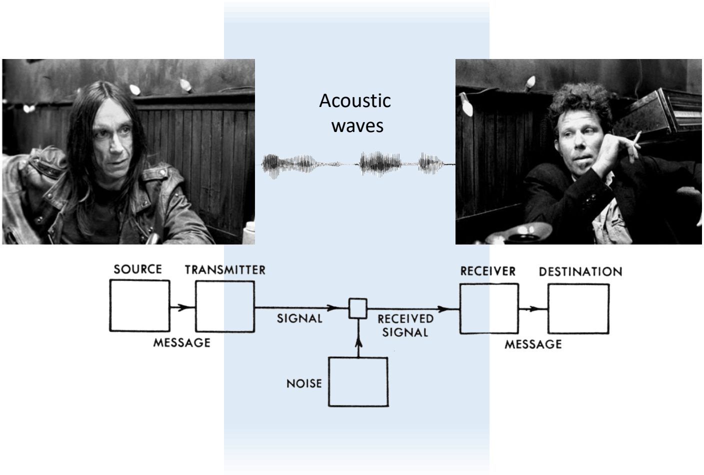
Plan

Decoding speech

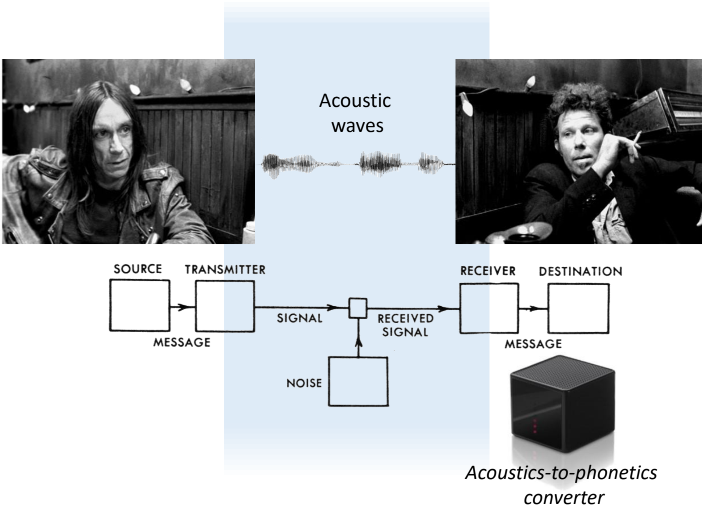
Decoding speech



Decoding speech



Decoding speech



Decoding speech

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

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

Decoding speech

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

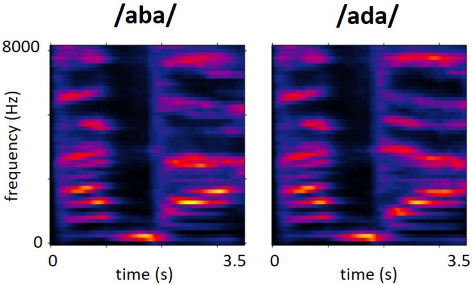
Which **acoustic cues** allow the listener
to differentiate one phoneme from another?

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

Decoding speech

How do we distinguish /aba/ from /ada/?

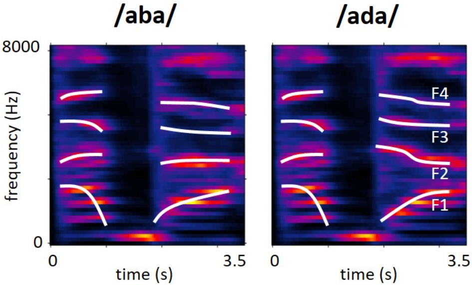
- Many acoustical differences



Decoding speech

How do we distinguish /aba/ from /ada/?

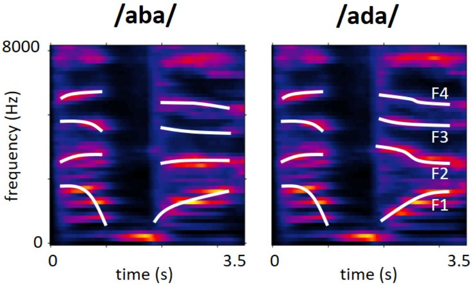
- Many acoustical differences (e.g. formant trajectories)



Decoding speech

How do we distinguish /aba/ from /ada/?

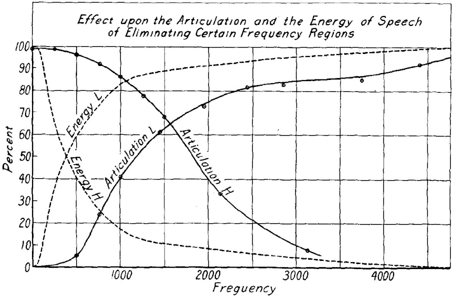
- Many acoustical differences (e.g. formant trajectories)
- Which ones are actually used by the auditory system ?



Decoding speech

One solution: using **reduced speech**

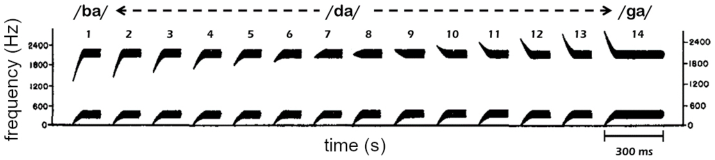
- Low-/high-pass filtered speech *[Fletcher, 1922]*



Decoding speech

One solution: using **reduced speech**

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

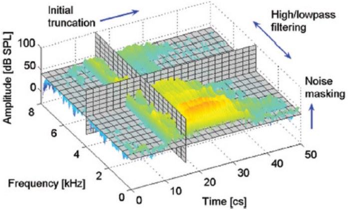


→ Proof that the **F2 onset** is a cue for categorizing /b/-/d/-/g/?

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

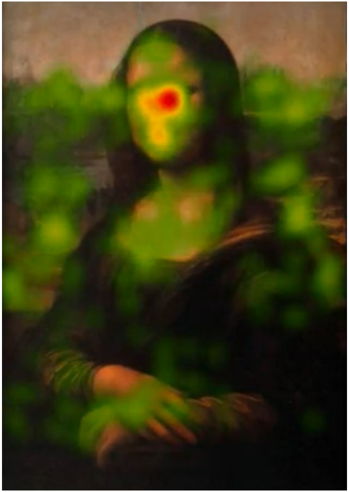
One solution: using **reduced speech**

- 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**.

The need for an **‘ear-tracker’**

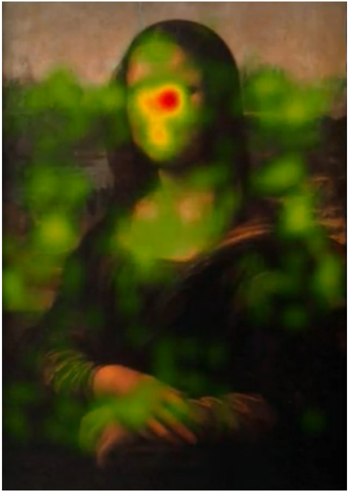
Developing a new method to visualize ‘where’ humans listen inside **natural** speech signals.



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

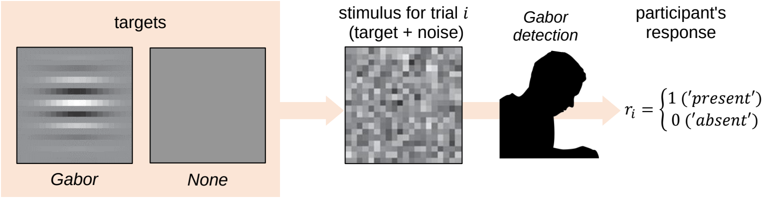
Developing a new method to visualize 'where' humans listen inside **natural** speech signals.

A solution could be provided by the **Classification Images** (CI) approach developed in visual psychophysics.



Visual Classification Images (CI)

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

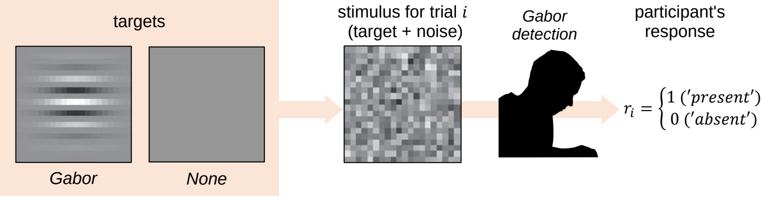


[Solomon, 2002]

Visual Classification Images (CI)

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

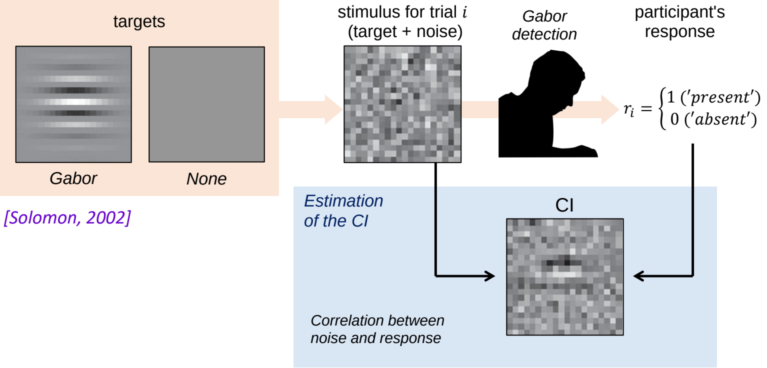
Which information is used to detect whether the target was present or not?



[Solomon, 2002]

Visual Classification Images (CI)

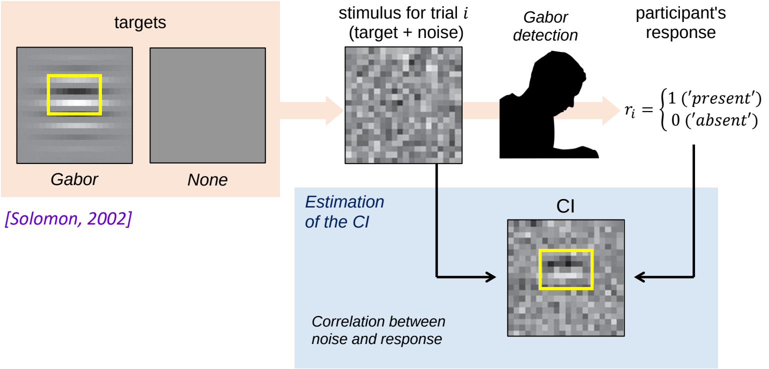
Correlation between the specific noise field in each trial and the response of the observer. The resulting correlation matrix shows how the presence of noise at each point interferes with the decision.



[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)

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
[4] Laboratoire sur le Langage le Cerveau et la Cognition, CNRS UMR5304, Lyon, France

Edited by:
Srikantan S. Nagarajan, University of California, San Francisco, USA

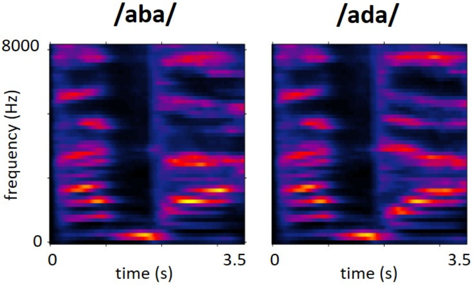
Reviewed by:
Shanqing Cai, Boston University, USA
Nima Mesgarani, Columbia University, USA

Correspondence:
Léo Varnet, Institute for Cognitive Science, 67 boulevard Pinel, 69675 Bron, France
e-mail: leo.varnet@isc.cnrs.fr

An essential step in understanding the processes underlying the general mechanism of perceptual categorization is to identify which portions of a physical stimulation modulate the behavior of our perceptual system. More specifically, in the context of speech comprehension, it is still a major open challenge to understand which information is used to categorize a speech stimulus as one phoneme or another, the auditory primitives relevant for the categorical perception of speech being still unknown. Here we propose to adapt a method relying on a Generalized Linear Model with smoothness priors, already used in the visual domain for the estimation of so-called classification images, to auditory experiments. This statistical model offers a rigorous framework for dealing with non-Gaussian noise, as it is often the case in the auditory modality, and limits the amount of noise in the estimated template by enforcing smoother solutions. By applying this technique to a specific two-alternative forced choice experiment between stimuli "aba" and "ada" in noise with an adaptive SNR, we confirm that the second formantic transition is key for classifying phonemes into /b/ or /d/ in noise, and that its estimation by the auditory system is a relative measurement across spectral bands and in relation

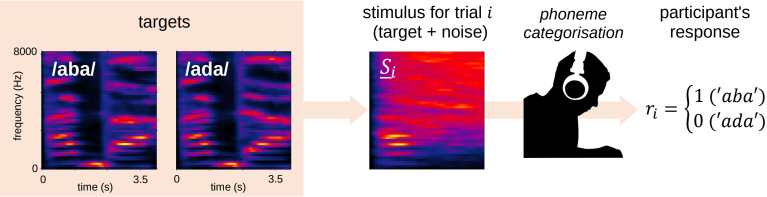
Materials

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%.

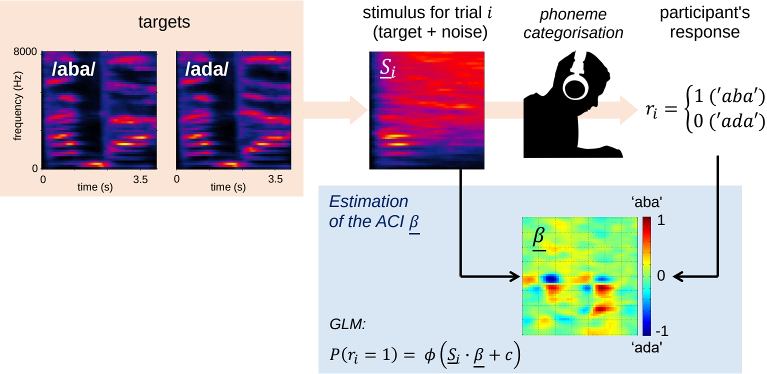
Methods



Two major differences:

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

Methods

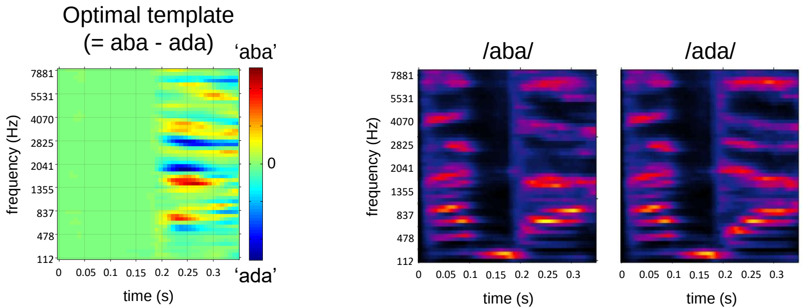


Generalized Linear Model (GLM):

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

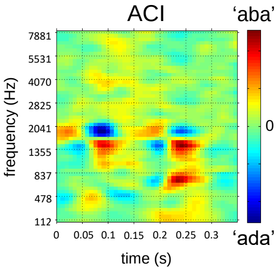
Optimal template

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



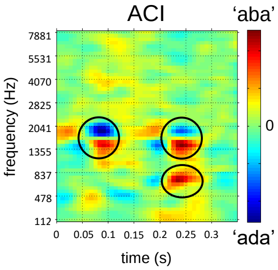
Real participant

- ACI does not look like the optimal template.



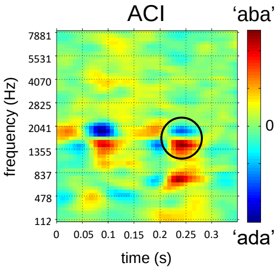
Real participant

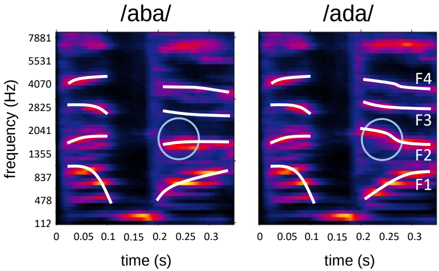
- ACI does not look like the optimal template.
- Clusters of positive and negative weights corresponding to the acoustic cues (preceded with the opposite pattern of weights).



Real participant

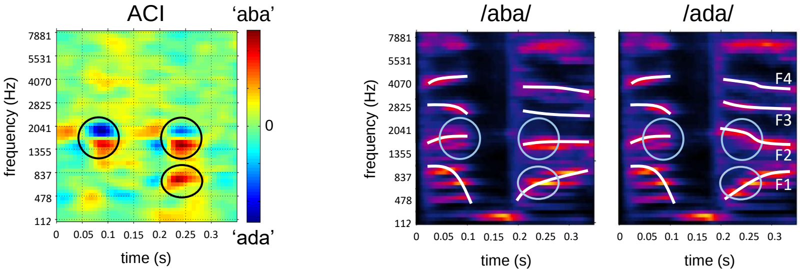
- Confirms that **the F2 onset is a cue** for classifying phonemes into /b/ or /d/ *(Liberman et al., 1954)*.





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

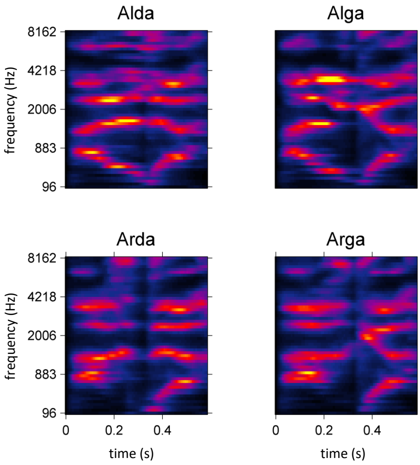
- The method works fine!
- Visualize what cues people listen to
  in natural speech signals (in noise)
- Can even reveal cues that are not present in the targets!

Aba/Ada experiment

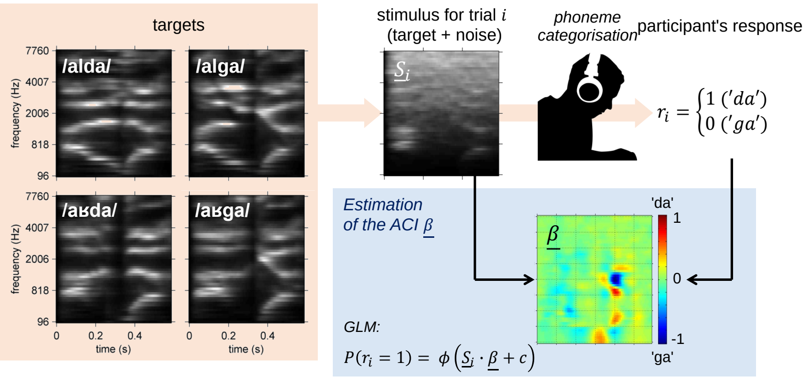
- The method works fine!
- Visualize what cues people listen to in natural speech signals (in noise)
- Can even reveal cues that are not present in the targets!
- … group-level ACIs?

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.

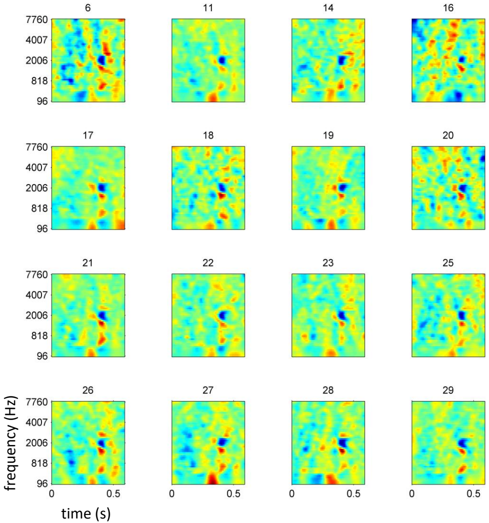


Methods


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)
$\underline{S}_i$
phoneme
categorisation
participant's response
$r_i = \begin{cases} 1 \ ('da') \\ 0 \ ('ga') \end{cases}$
Estimation
of the ACI $\underline{\beta}$
$\underline{\beta}$
'da'
1
0
-1
'ga'
GLM:
$P(r_i = 1) = \phi\left(\underline{S}_i \cdot \underline{\beta} + c\right)$

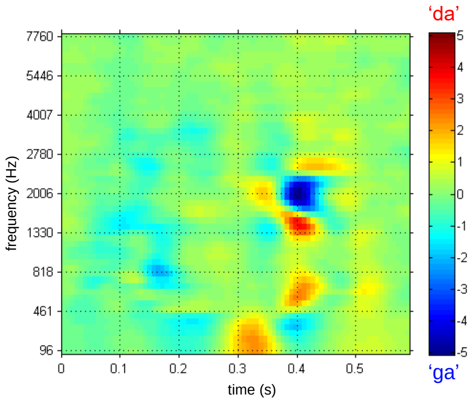
Group ACI

- Similar pattern of weights for all 16 participants.



Group ACI

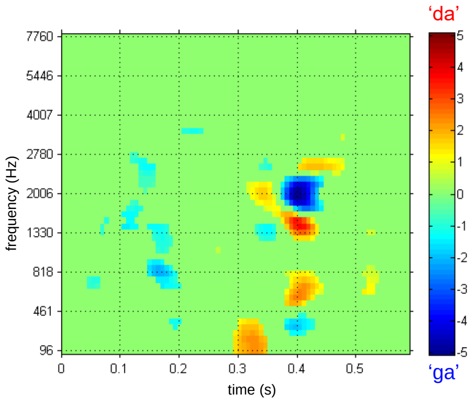
- Similar pattern of weights for all 16 participants.
- Primary cue: negative cluster surrounded by positive cluster.
- Other cues at lower frequencies.



Mean ACI over 16 participants

Group ACI

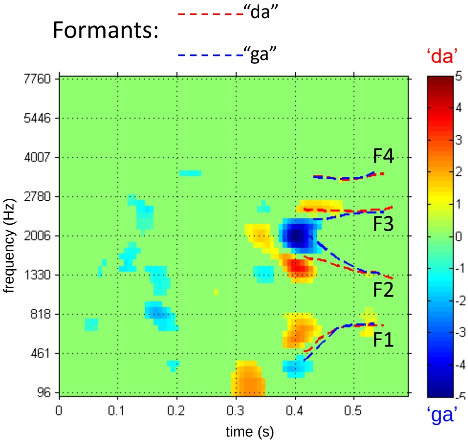
- Similar pattern of weights for all 16 participants.
- Primary cue: negative cluster surrounded by positive cluster.
- Other cues at lower frequencies.



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

Group ACI

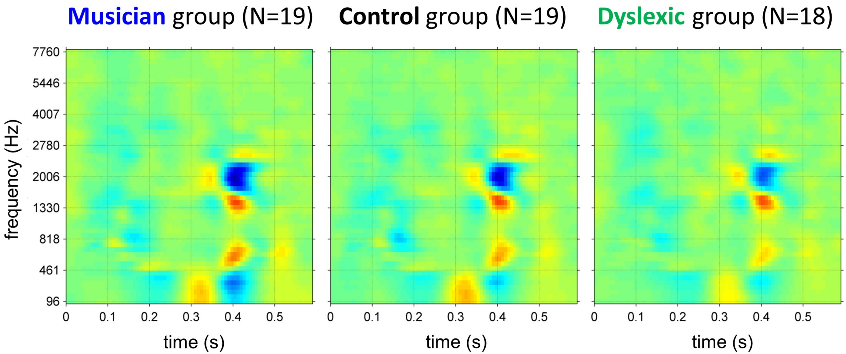
- The F2 and F3 onsets are critical cues for this task.
- The onset of F1 is also a cue for categorization.



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

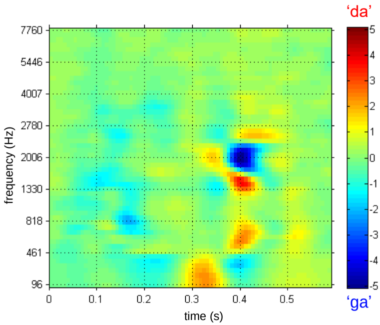
Comparing groups of listeners

Same /da/-/ga/ categorization experiment with musician experts *[Varnet et al., 2015]* or dyslexic participants *[Varnet et al., 2016]*.



Bump noise

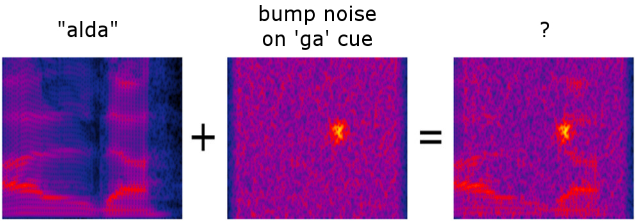
The clusters of weights on the ACI are regions where the presence of noise biases categorization toward /d/ or /g/.



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

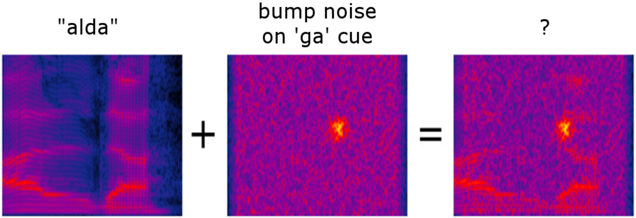
Bump noise

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



Bump noise

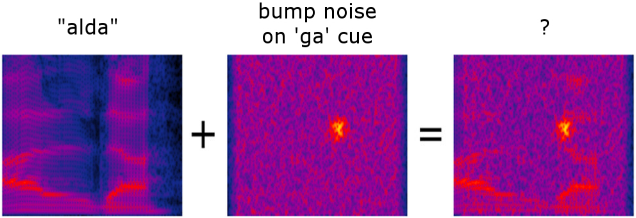
Bump noise: white noise with an additional bump of noise on the location of a cue previously identified.



$\rightarrow$ a noise that shifts perception from da to ga (or from ga to da)!

Bump noise

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



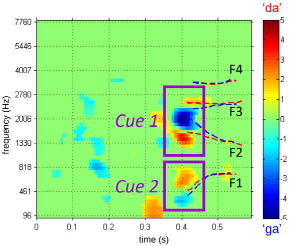
$\rightarrow$ a noise that shifts perception from da to ga (or from ga to da)!

First application: 'parametric bump noise experiment' (see part 2).

Interim summary

The acoustic-to-phonetic conversion is a complex process

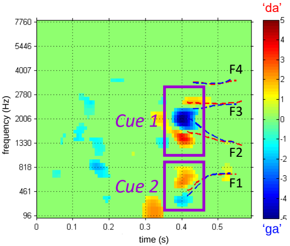
- involving **multiple cues**
- some of which may be **anticipatory**
- cues are associated with **different weights** in the decision



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



How consistent are those strategies across listeners / groups of listeners? In particular for hearing-impaired listeners?

Varnet et al., 2019

Original Article

Trends in Hearing
Volume 23: 1–18

DOI: 10.1177/2331216519886707
journals.sagepub.com/home/tia
SAGE

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

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

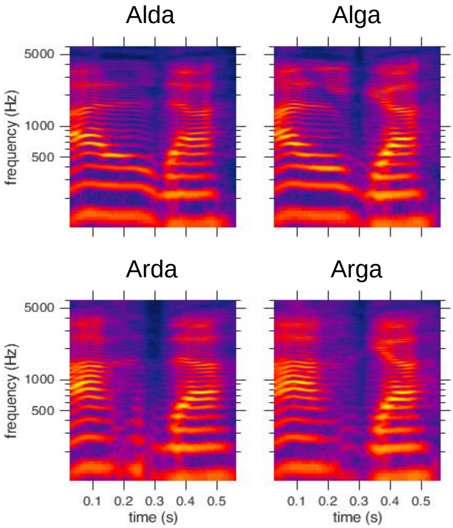
Abstract
There is increasing evidence that hearing-impaired (HI) individuals do not use the same listening strategies as normal-hearing (NH) individuals, even when wearing optimally fitted hearing aids. In this perspective, better characterization of individual perceptual strategies is an important step toward designing more effective speech-processing algorithms. Here, we describe two complementary approaches for (a) revealing the acoustic cues used by a participant in a /d/-/g/ categorization task in noise and (b) measuring the relative contributions of these cues to decision. These two approaches involve natural speech recordings altered by the addition of a "bump noise." The bumps were narrowband bursts of noise localized on the spectrotemporal locations of the acoustic cues, allowing the experimenter to manipulate the consonant percept. The cue-weighting strategies were estimated for three groups of participants: 17 NH listeners, 18 HI listeners with high-frequency loss, and 15 HI listeners with flat loss. HI participants were provided with individual frequency-dependent amplification to compensate for their hearing loss. Although all listeners relied more heavily on the high-frequency cue than on the low-frequency cue, an important variability was observed in the individual weights, mostly explained by differences in internal noise. Individuals with high-frequency loss relied slightly less heavily on the high-frequency cue relative to the low-frequency cue, compared with NH individuals, suggesting a possible influence of supra-threshold deficits on cue-weighting strategies. Altogether, these results suggest a need for individually tailored speech-in-noise processing in hearing aids, if more effective speech discriminability in noise is to be achieved.

Keywords

Acoustic cues and hearing loss

Information in speech is **redundant**.

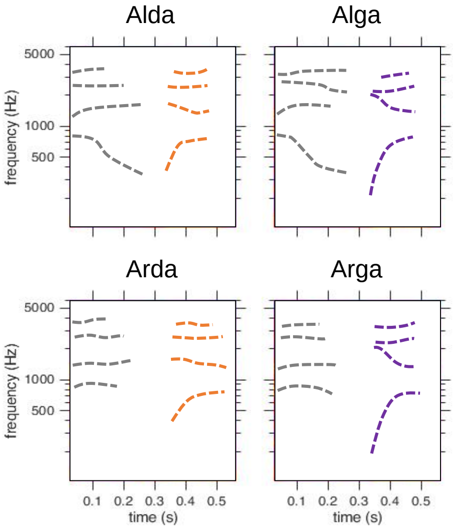
Cues for categorizing da/ga in noise *[Varnet et al., 2015]*:



Acoustic cues and hearing loss

Information in speech is **redundant**.

Cues for categorizing da/ga in noise *[Varnet et al., 2015]*:

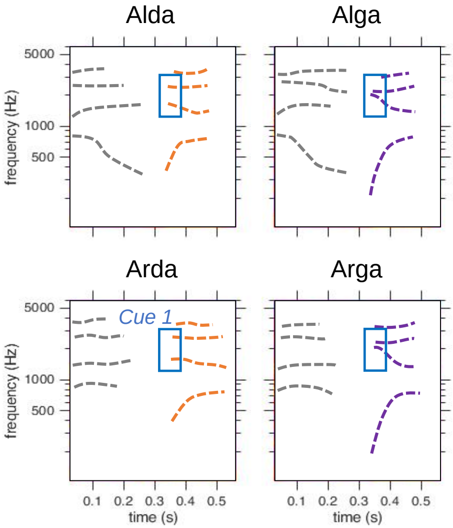


Acoustic cues and hearing loss

Information in speech is **redundant**.

Cues for categorizing da/ga in noise *[Varnet et al., 2015]*:

- Primary HF cue: F2/F3 onsets

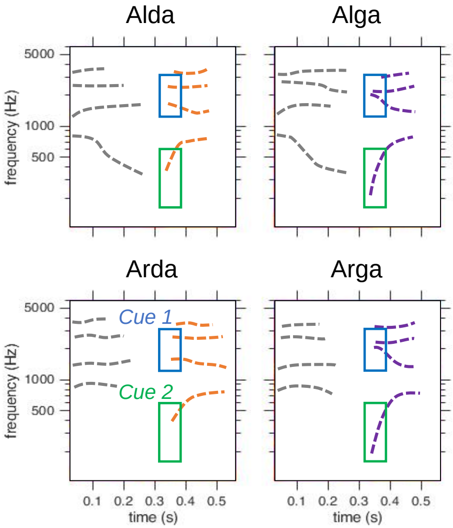


Acoustic cues and hearing loss

Information in speech is **redundant**.

Cues for categorizing da/ga in noise *[Varnet et al., 2015]*:

- Primary HF cue: F2/F3 onsets
- Secondary LF cue: F1 onset



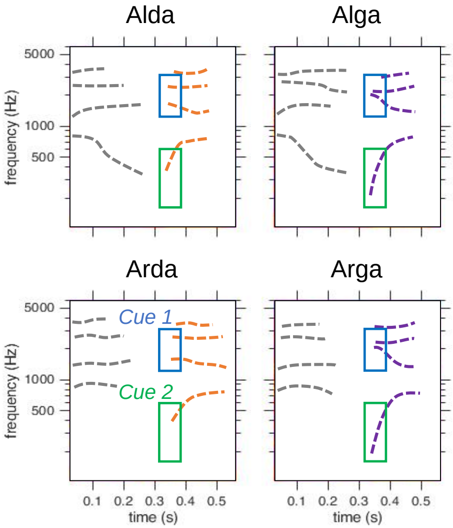
Acoustic cues and hearing loss

Information in speech is **redundant**.

Cues for categorizing da/ga in noise *[Varnet et al., 2015]*:

- Primary HF cue: F2/F3 onsets
- Secondary LF cue: F1 onset

How is the relative importance of the high- and low-frequency cues impacted by hearing loss?



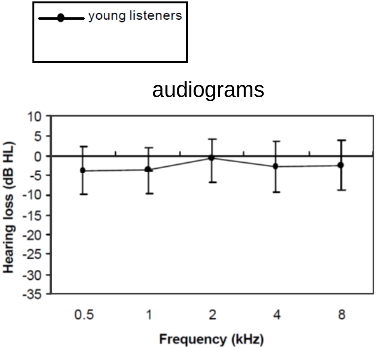
Frequency weighting functions

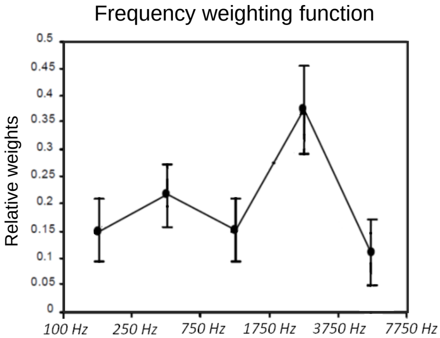
Relative importance of different frequency bands for speech intelligibility.

Frequency weighting functions

Relative importance of different frequency bands for speech intelligibility.

- Listeners with HF hearing loss rely relatively less on the HF information in speech, compared to normal-hearing listeners. *[Gilbert et al., 2002]*

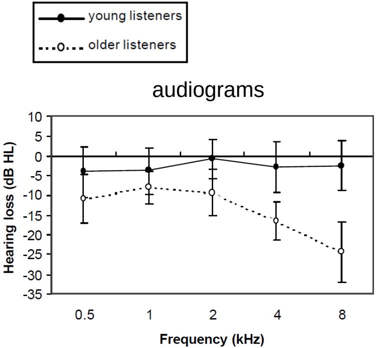


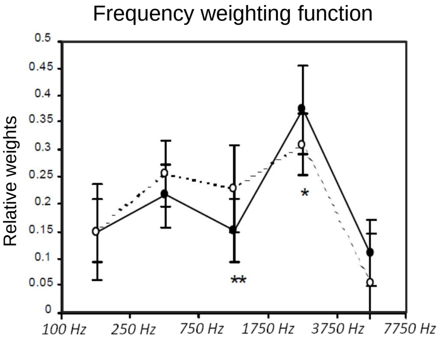


Frequency weighting functions

Relative importance of different frequency bands for speech intelligibility.

- Listeners with HF hearing loss rely relatively less on the HF information in speech, compared to normal-hearing listeners. *[Gilbert et al., 2002]*

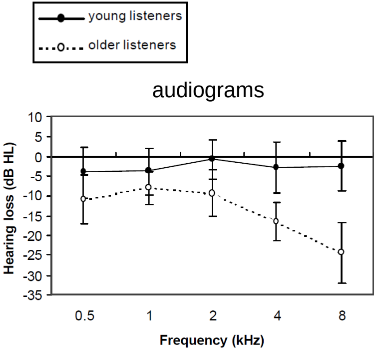


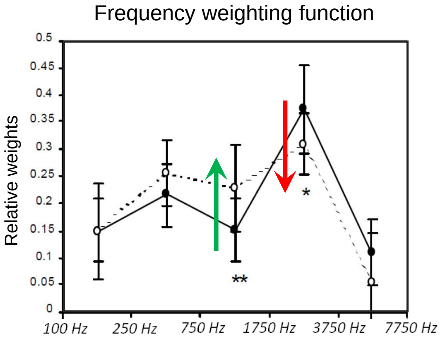


Frequency weighting functions

Relative importance of different frequency bands for speech intelligibility.

- Listeners with HF hearing loss rely relatively less on the HF information in speech, compared to normal-hearing listeners. *[Gilbert et al., 2002]*

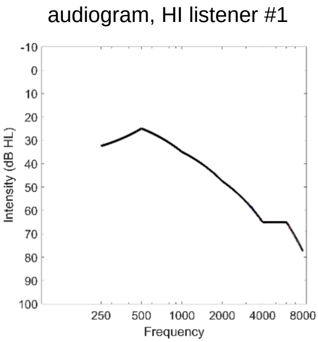


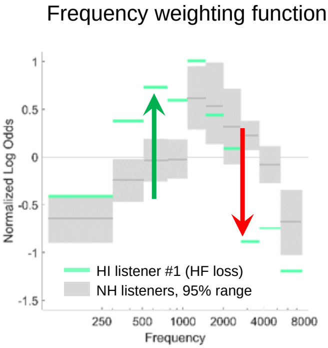


Frequency weighting functions

Relative importance of different frequency bands for speech intelligibility.

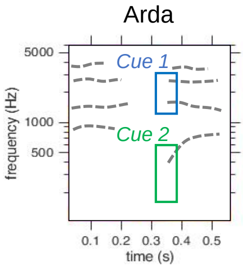
- Listeners with HF hearing loss rely relatively less on the HF information in speech, compared to normal-hearing listeners. *[Gilbert et al., 2002]*
- The same is true **when they are provided with frequency-dependent amplification** (NAL-R). *[Yoho, Long & Bosen, 2018]*





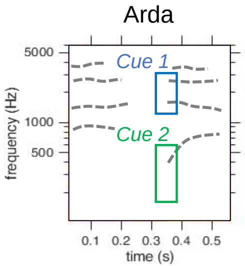
Questions

Frequency weighting functions measure the global importance of a frequency band – how about the weights associated to each cue?



Questions

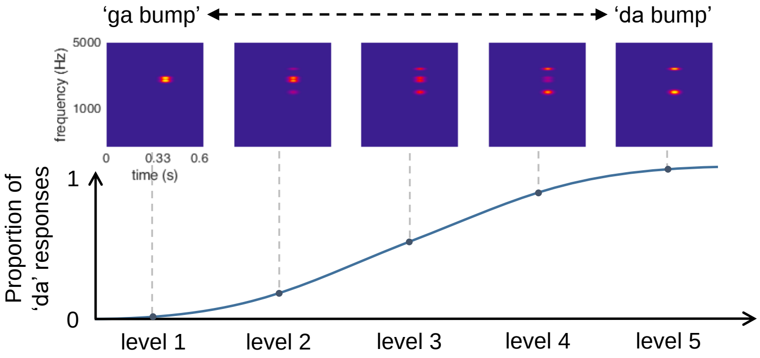
Frequency weighting functions measure the global importance of a frequency band – how about the weights associated to each cue?



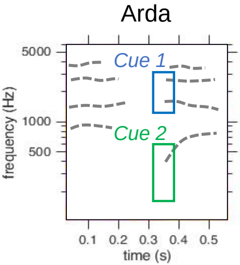
- Do individuals with HF loss rely less on the primary HF cue than individuals with...
  - normal hearing (NH)?
  - 'flat' hearing loss?
- Do they use the LF cue as a primary cue in this case?

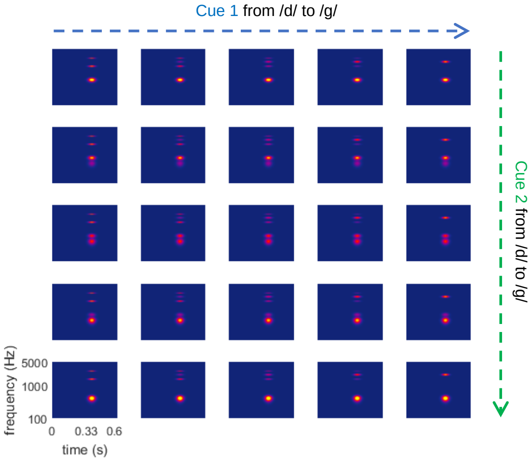
Manipulating acoustic cues

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



Parametric bump noise paradigm

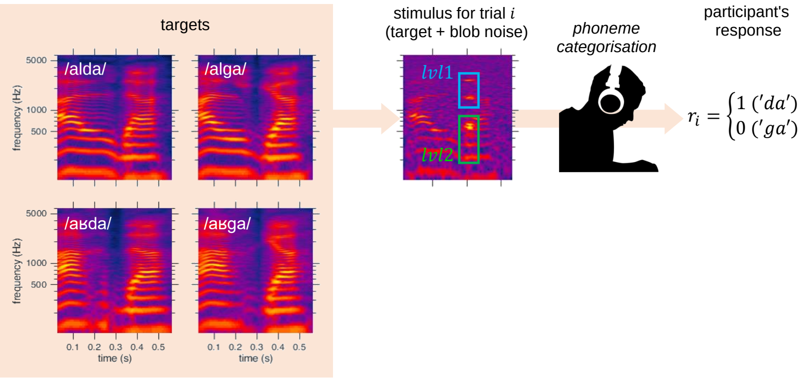




Joint measurement of the weights on **two separate cues**.

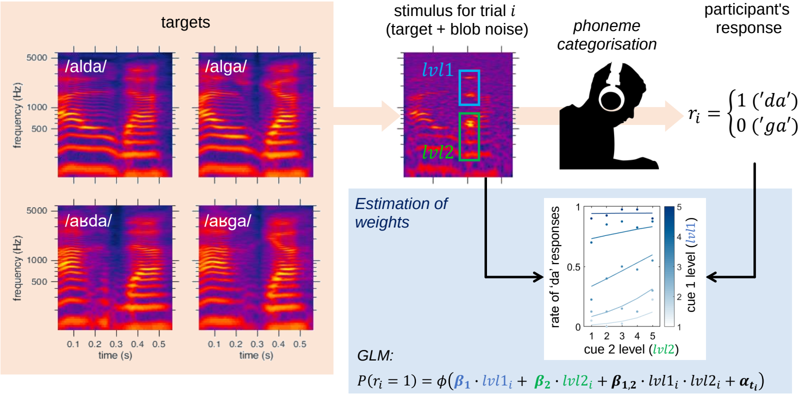
$\Rightarrow$ **2-dimensional** /da/-/ga/ continuum.

Parametric bump noise paradigm



1000 trials per participant (4 targets X 25 bump configs X 10 repetitions).

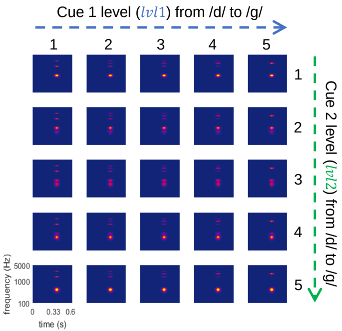
Parametric bump noise paradigm



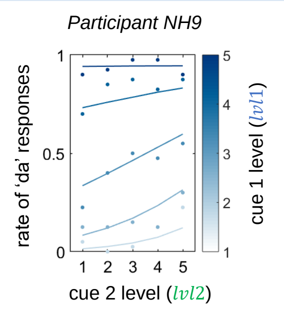
1000 trials per participant (4 targets X 25 bump configs X 10 repetitions).
β_1 and β_2: participant's weights on cue 1 and cue 2.

Parametric bump noise paradigm

Bump noise continuum



Estimation of weights



GLM:

$P(r_i = 1) = \phi(\boldsymbol{\beta_1} \cdot lvl1_i + \boldsymbol{\beta_2} \cdot lvl2_i + \boldsymbol{\beta_{1,2}} \cdot lvl1_i \cdot lvl2_i + \boldsymbol{\alpha_{t_i}})$

1000 trials per participant (4 targets X 25 bump configs X 10 repetitions).

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

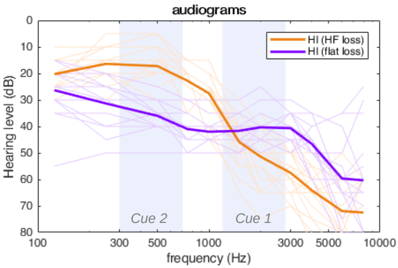
$\rightarrow$ Strong effect of cue 1; weak effect of cue 2.

Participants

- young **Normal-hearing** participants (NH)
  (N=17; 23-36 y.; M=27 y.)

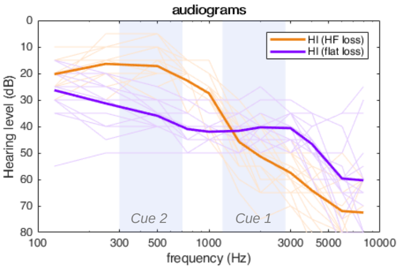
Participants

- young **Normal-hearing** participants (NH)
  (N=17; 23-36 y.; M=27 y.)
- **Hearing-impaired** participants with **HF loss** (HI - HF)
  (N=18; 55-73 y.; M=64 y.)
- **Hearing-impaired** participants with **flat loss** (HI - flat)
  (N=15; 51-72 y.; M=64 y.)



Participants

- young **Normal-hearing** participants (NH)
  (N=17; 23-36 y.; M=27 y.)
- **Hearing-impaired** participants with **HF loss** (HI - HF)
  (N=18; 55-73 y.; M=64 y.)
- **Hearing-impaired** participants with **flat loss** (HI - flat)
  (N=15; 51-72 y.; M=64 y.)



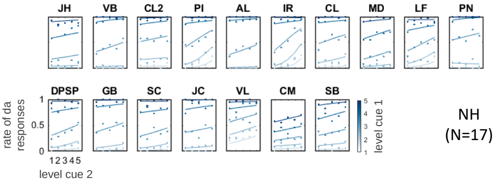
All HI participants had been wearing HAs for $>$ 1 mo. except 2 in each group. Audibility restored using frequency-dependent amplification (NAL-R).

Cue-weighting strategies

3 groups:

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

Audibility restored with NALR.



Large variability **within-groups**...

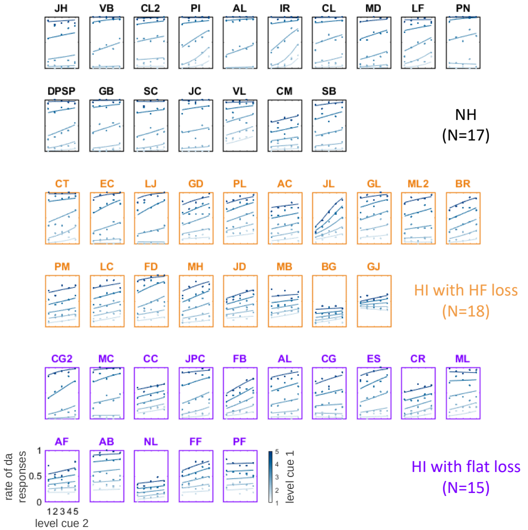
Cue-weighting strategies

3 groups:

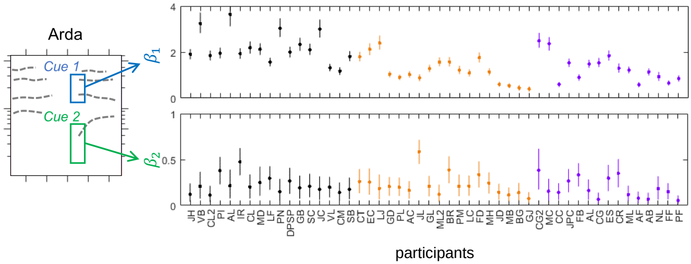
- **NH**
- **HI (HF loss)**
- **HI (flat loss)**

Audibility restored with NALR.

Large variability **within-groups** and **between-groups**.



Cue-weighting strategies

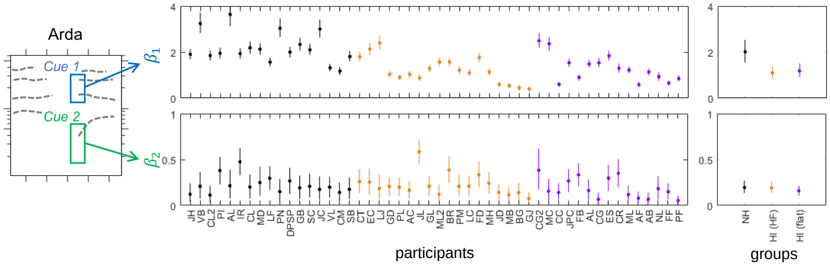


3 groups:

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

Audibility restored with NALR.

Cue-weighting strategies



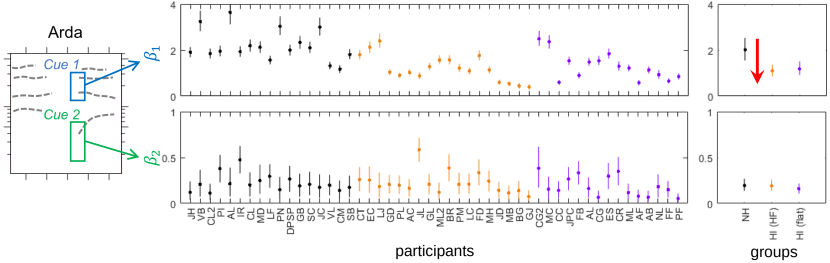
3 groups:

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

Audibility restored with NALR.

After frequency-dependent amplification, **all HI participants still relied more heavily on the primary cue** than on the secondary cue.

Cue-weighting strategies



3 groups:

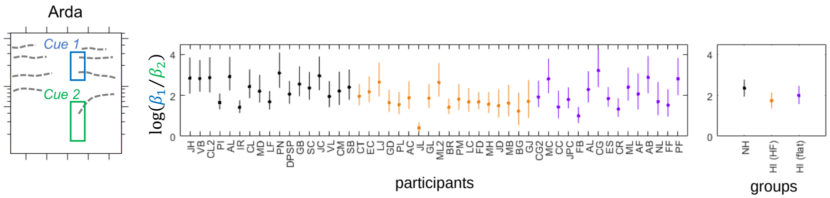
- **NH**
- **HI (HF loss)**
- **HI (flat loss)**

Audibility restored with NALR.

After frequency-dependent amplification, **all HI participants still relied more heavily on the primary cue** than on the secondary cue.

Differences in the exact weights associated to each cue.

Cue-weighting strategies



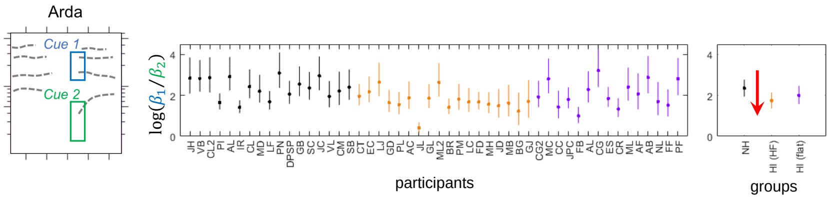
3 groups:

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

Audibility restored with NALR.

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)**

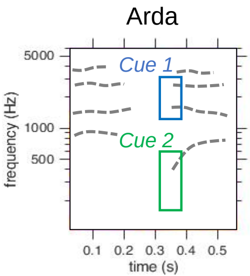
Audibility restored with NALR.

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.

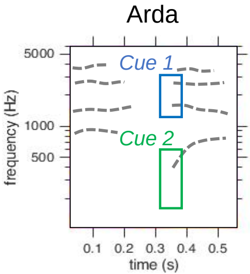
Conclusions

- HF hearing loss affects the use of the high-frequency cues...



Conclusions

- HF hearing loss affects the use of the high-frequency cues...
- ... even after compensating for the audibility deficit.



This may explain why individuals with HF loss **receive only limited benefit from their hearing aid** in phoneme discrimination tasks.
[Scheidiger & Allen, 2013; Abavisani & Allen, 2017]

Conclusions

ACI: a set of tools to explore auditory categorization tasks. The ACI gives an insight into the black box.

The **ANR fastACI** (2021-2024) aims at further improving the method, then applying it to the whole french phonological system.

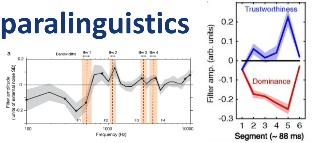
Available on GitHub as an open-source MATLAB toolbox: `https://github.com/aosses-tue/fastACI`

Auditory revcorr studies

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

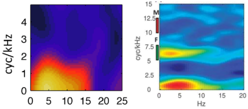
high level

paralinguistics

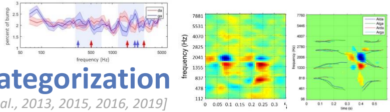


[Ponsot et al., 2018a, 2018b]

sentence recognition



[Venezia et al., 2016, 2019]

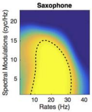


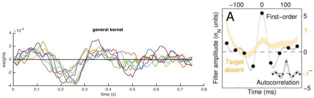
phoneme categorization

[Varnet et al., 2013, 2015, 2016, 2019]

timbre judgment

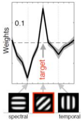
[Thoret et al., 2017]

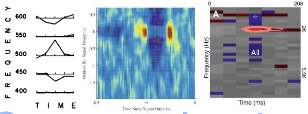




modulation perception

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





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

pure-tone detection

low level

Thanks for your attention!

And thanks to:

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

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