Impact of noise on speech comprehension and identification of the acoustic cues that are encoded and used for speech processing

Fanny Meunier

Laboratoire Base, Corpus et Langage
Université de Nice Sophia-Antipolis &CNRS (UMR 7320)
Nice, France

/s/..

…
soap
spinach
psychologist
spin
spit
sun
spank
…

• Acoustic cues used on-line to modulate the activation of targets and competitors

• Can they be of any help towards correct continuous speech segmentation?

• Due to elision, some spoken utterances in French are phonemically identical (e.g., l'amie 'the friend' vs. la mie 'the crumb', both [lami]).

• There are no clear word boundaries in spoken language.

• Correct segmentation into discrete word units is necessary for comprehension.

• There are, some fine acoustic differences between members of ambiguous sequences that could be exploit by listeners.

Phonemically identical pairs

30 phonemically identical pairs

	C'est la fiche ('it's the sheet')	**C'est l'affiche** ('it's the poster')	→ [selafiʃ]
Acoustic analyses	↓	↓	
Duration	[la] = 146 ms [l] = 59 ms [a] = 89 ms	[la] = 151 ms [l] = 69 ms [a] = 82 ms	

⇨ Longer durations for the beginning of content words

F2	[a] = 1834 Hz	[a] = 1808 Hz	

⇨ More front articulation for the determiner *la*

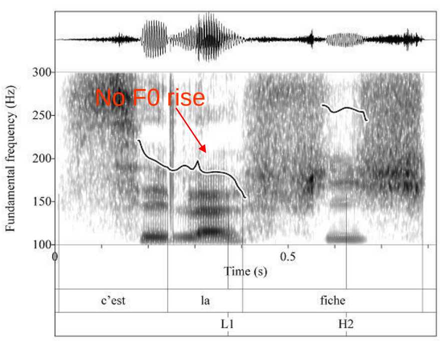
F0	[a] = 179 Hz	[a] = 195 Hz	

⇨ Rise in fundamental frequency beginning at the left edge of the first content word syllable

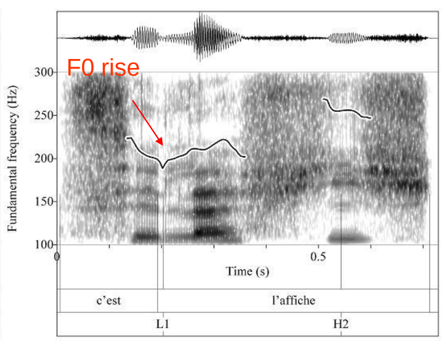
Intonational cue

[selafiʃ]

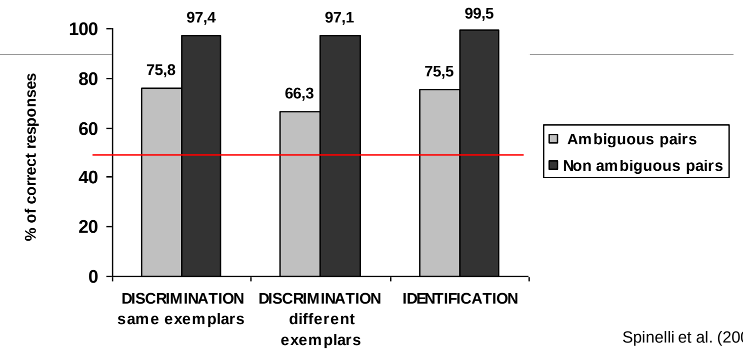
C'est la fiche ('it's the sheet')



C'est l'affiche ('it's the poster')



Do people make the difference ?

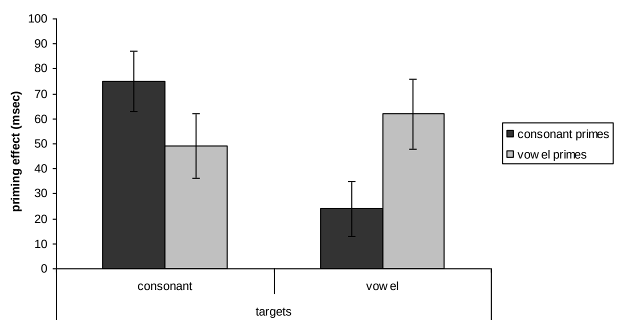


Spinelli et al. (2007).

Listeners can retrieve the correct segmentation in such utterances: they can discriminate between and identify them

• How much are activation of targets and competitors modulated by the fine cues?

Prime		**Target**	
Reliées (Voyelle initiale, *l'affiche*)	[lafiʃ]	**FICHE**	[fiʃ]
Reliées (Consonne initiale, *la fiche*)	[lafiʃ]		
Non Reliées (*la mer / l'amer*)	[lamER]	**AFFICHE**	[afiʃ]



Priming for all conditions, irrespective of the intended segmentation

Priming is greater in the intended segmentation

Acoustic cues are not powerful enough to rule out alternative hypotheses.

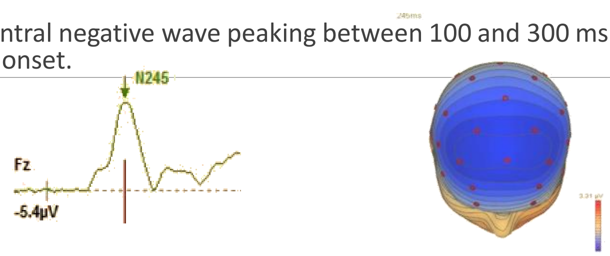
Behavioral experiments:
- focused on word form
- used only one production

Solution:
Use Event related potentials with no task = the MMN.

Pota, Spinelli, Varnet, Hoen, Meunier (2017, sub.)

Mismatch negativity (MMN; Näätänen et al. since 1980)

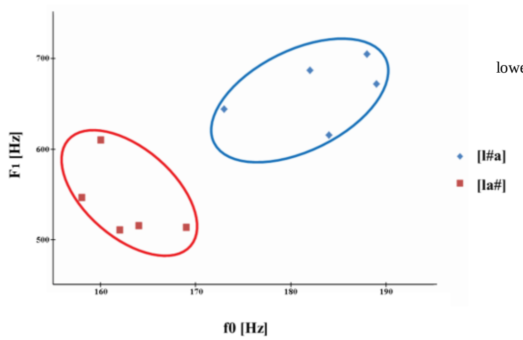
- No task
- fronto-central negative wave peaking between 100 and 300 ms after the deviance onset.



- detection of unexpected changes in some aspects of a regular auditory stream
- oddball paradigm: one rare sound (deviant) occurs in a series of frequent stimuli (standards).
- reflection of the formation of sensory memory traces from statistical regularities in the input signal
- forced the mapping of the signal onto more abstract representations
- also reflect fail of predictive models (allow adjustement)

(Wacongne, 2012)

Variabilities in the tokens



lower tongue position

Différences in duration [l#a]>[la#], fundamental frequency [l#a]>[la#] and formant value
=> [l#a] more canonical [a]

Figure 3. Mean f_0 value in the function of a Mean F_1 value for the 1st vowel of the 1st syllable (i.e., [a]) in all the productions.

MÉTHODE

- Stimuli :
  - Words:
    - *Phonologically similar*:
      - la locution ('the phrase') / l'allocution ('the speech')
    - *Phonologically different*: l'illocution
  - CV :
    - la, l'a, l'i

Used of a modified version of the Oddball paradigm (N. Kraus, 2000)

---------- Std 1 | Std 2 | Std 3 | Std 4 | **Dev** | Std … ----------

Std = • ***4 productions of /la/***

Dev =
- ***phonologically similar deviant LA (*5)***
- ***phonologically different deviant LI (*5)***
- ***Other production of the Std***

1125-1800 stim

Results and Discussion

No MMN for the identical segmentation conditions (in which another token of the standard was presented at the test position), despite the acoustic variability between the different tokens.

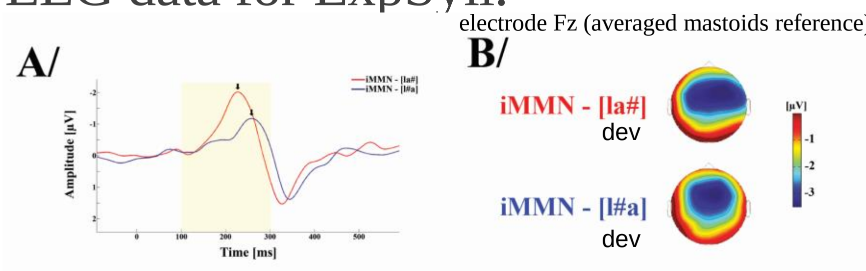
⇒ intra-speaker variability was not considered deviant by the neural system = the MMN is a good tool for tracking the meaningful changes of acoustical stimulation.

Homophone conditions: clear MMNs for both CV and nominal sequences, despite the variability of standard tokens.

⇒ the MMNs observed for the homophone conditions were not merely a change detection answer based solely on acoustic feature divergences.

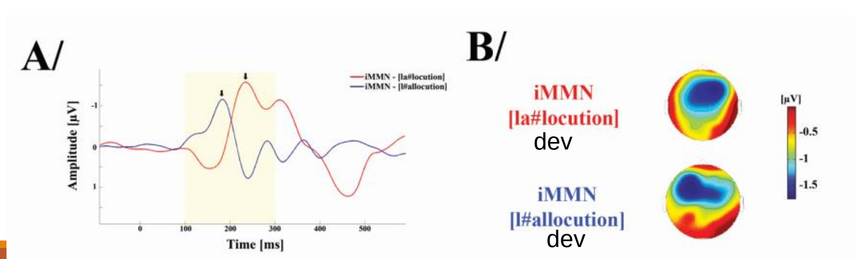
EEG data for ExpSyll.

electrode Fz (averaged mastoids reference).



EEG data for ExpWord

40 ms time-window centered on the amplitude peak



Contrast LA – L'A

l'a-stand = more negative MMNs than *la*-stand :

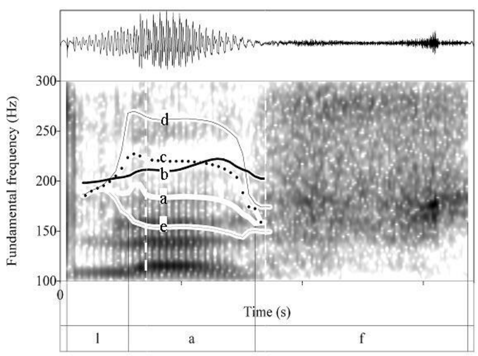
- *la* : greater variability in the different tokens in the initial syllable duration and the mean first formant value on the vowel [a].
- *l'a*, which is less variable et have a rising F0 may produce a more refined sensory memory trace, leading to more precise predictions, thereby increasing sensitivity to deviants and amplifying MMNs.

La : the determiner

The deviant *l'a* elicited a later MMN peak latency than its homophone *la* in ExpSyllable (+39 ms), while it appears earlier (-59 ms) in ExpWord.

- The syllable [la] in Expword stimuli being a determiner and corresponding to one of the most common definite articles in French (thefeminine) could create a memory trace that included syntactic information .
- with its well-differentiated topography on a left-frontal area and its precocity, appears to have the characteristics of a syntactic-MMN (Pulvermüller & Shtyrov, 2003) or an early left anterior negativity (ELAN, Friederici et al., 1993).

F0 manipulation



Spinelli et al. 2010

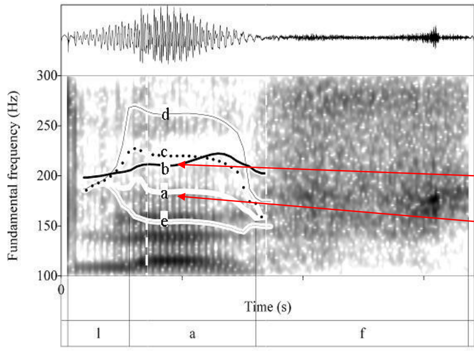
b). natural vowel-initial *l'affiche*

a). natural consonant-initial *la fiche*

+ 3 resynthesized versions created by modifying the F0 of *la fiche* across the first vowel ([a]).

Considering the intonational phonology of the language, the manipulation of F0 values should modify listeners' segmentation.
Decreasing the F0 value in the [a] of of *la fiche* should give rise to more consonant-initial segmentation (*la fiche*) whereas increasing the F0 in the [a] of *la fiche* should give rise to more vowel-initial segmentation (*l'affiche*).

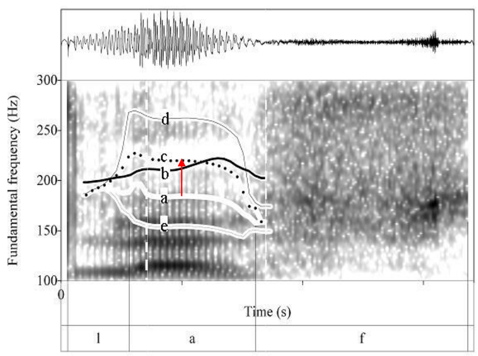
F0 manipulation



b). natural vowel-initial *l'affiche*

a). natural consonant-initial *la fiche*

F0 manipulation

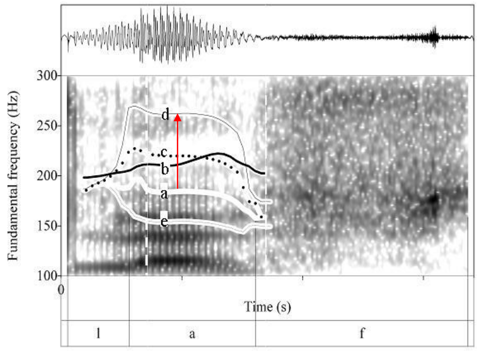


b). natural vowel-initial *l'affiche*

a). natural consonant-initial *la fiche*

c) resynth-consonant-equal condition : F0 of the first vowel [a] resynthesized to *equal* that of the vowel-initial item *l'affiche*).

F0 manipulation



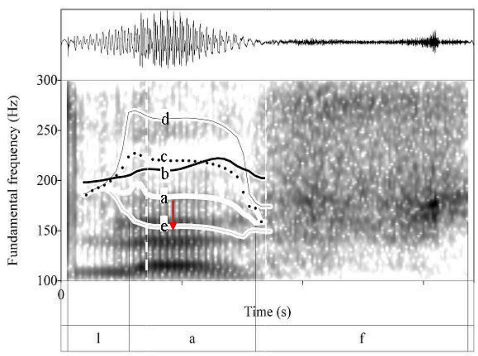
b). natural vowel-initial *l'affiche*

a). natural consonant-initial *la fiche*

c) resynth-consonant-equal condition : F0 of the first vowel [a] resynthesized to *equal* that of the vowel-initial item *l'affiche*).

d) resynth-consonant-higher condition: F0 of the first vowel [a] resynthesized to be *higher* that of the vowel-initial item *l'affiche*).

F0 manipulation



b). natural vowel-initial *l'affiche*

a). natural consonant-initial *la fiche*

c) resynth-consonant-equal condition : F0 of the first vowel [a] resynthesized to *equal* that of the vowel-initial item *l'affiche*).

d) resynth-consonant-higher condition: F0 of the first vowel [a] resynthesized to be *higher* that of the vowel-initial item *l'affiche*).

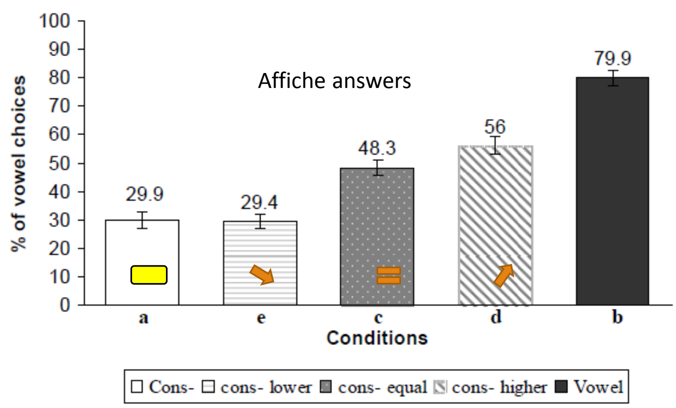
e) resynth-consonant-lower condition: F0 of the first vowel [a] resynthesized to be *lower* that of the consonant-initial item *l'affiche*).

Impact of F0 manipulation on perceived segmentation

50 participants

2AFC

fiche / affiche ?



Summary of on-line experiments

- Acoustic cues used on-line to modulate the activation of targets and competitors
- They guide listeners towards the correct segmentation
- They are not powerful enought to rule out alernative hypotheses
- Some cues seem to be used more than other.
  - F0
- Set up a « new method » to see the cues used on line
  - Auditory Classification Image

Varnet et al., 2013, 2015, 2015, 2016

The need for an “ear-tracker”

Our aim is to develop a new method to directly see where humans listen inside natural speech signals…

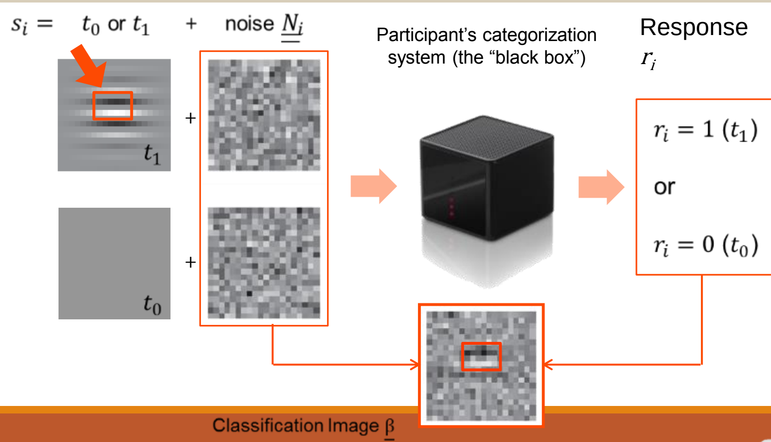
…like an **eye-tracker**:

A solution could be provided by the technique of **Classification Image** (CIm).

Classification Image: brief description

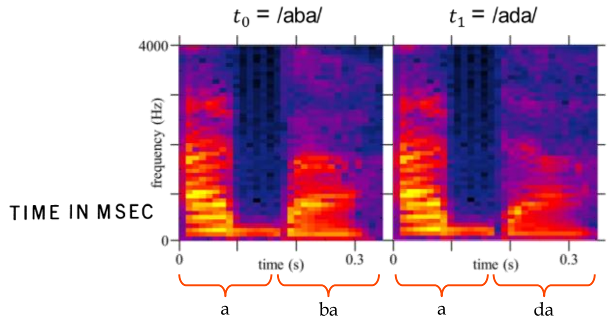
Correlation between the specific **noise field** in each trial and the **response** of the observer. The resulting correlation matrix shows how the presence of noise at each point interferes with the decision.

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



Materials…

Target : 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/ (power-normalized).



(à -12 dB)

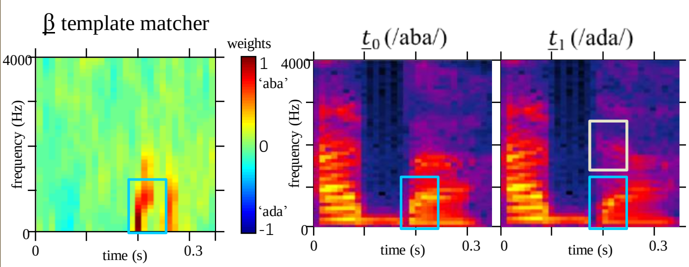
Stimuli: Target sounds in an additive Gaussian noise.

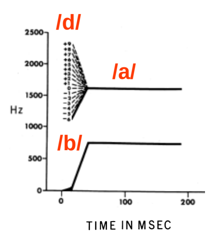
Task: Indicate whether the target was /aba/ or /ada/.

The SNR was adapted continuously to ensure a correct response rate of 75%.

Modeling an ideal template-matcher

What would be the results if the participants performed the task linearly by comparing the input stimulus with templates of the two targets and choosing the most similar?



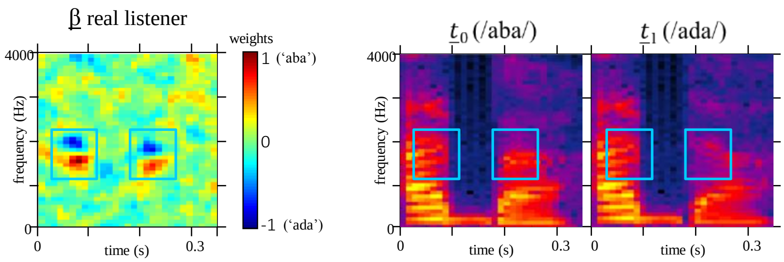


The ideal template-matcher follows the **optimal strategy**: taking into account only **the region where the targets differ most** in terms of energy.

Liberman (1954): The second formantic transition is a key for classifying phonemes into /b/ or /d/.

Classification Images: real listener

Two similar patterns (a cluster of positive weights below a cluster of negative weights) at two precise time-frequency locations.



The critical time-frequency locations exactly match the **F2 transitions** !
The participant also used information from the first syllable, even though this region actually contains **no useful acoustic cue** for performing the task.

This measurement of the F2 onset frequency by the auditory system is a **relative estimation**. This could allow us to compensate for variability in speech production…

Dyslexia

Defined as a durable impairment in the development of normal reading skills despite appropriate education and in the absence of other sensory or cognitive deficits.

Dyslexia

Defined as a durable impairment in the development of normal reading skills despite appropriate education and in the absence of other sensory or cognitive deficits.

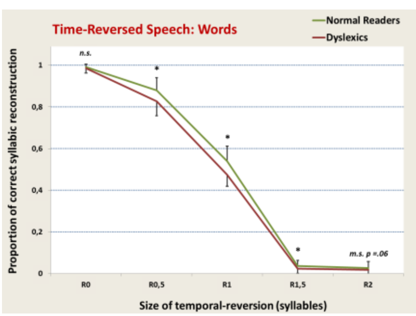
Dyslexia is associated with speech in noise comprehension difficulties

Often ignored : hardly observable in good listening conditions.

- Speech-in-noise comprehension deficit in children.
  Ziegler JC, Pech-Georgel C, George F, Lorenzi C. Dev Sci. 2009.
- Associated to difficulties understanding speech in adverse cognitive or listening situations.
- SpiN deficit in children predictive of dyslexia.
  Boets et al., 2013.

Degraded signal

N= 40 adult Dyslexics vs. 40 Normal Readers

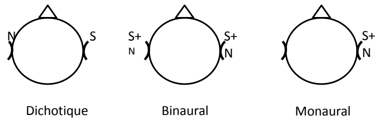


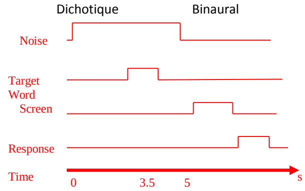
Listening Experiment

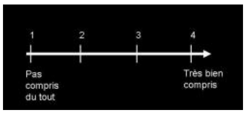
fMRI Study

Methods

Configuration	OG	OD
Dichotique (C1)	Babble	Cible
Binaural (C2)	Cible / Babble (10 dB)	Cible/Bable (0 dB)
Monaural (C3)	Silence	Cible/Babble



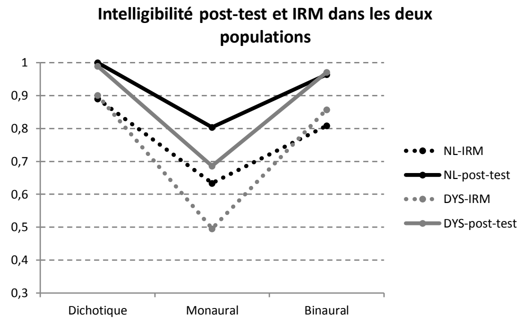




Subjective Intelligibility judgement task

Post-hoc objective intelligibility test

Results: Behavior (N = 16 & 16)



Dole, Hoen & Meunier, *Neuropsychologia*, 2012 Dole, Meunier & Hoen, *2014*

fMRI Study

Binaural – Monaural: spatial release of masking

FWE corrected

In the binaural condition where dyslexics did not get behavioural impairment, they show an increased activation in the right sup. temporal regions.

- **Morphometry (MRI) :** normal readers > dyslexics reveals a significant WM and GM volumes difference in middle right STG (BA22).

… same areas where dys showed functional compensation.

Dole, Hoen & Meunier, *Neuropsychologia*, 2012 Dole et al., 2013 Dole, Meunier & Hoen, *2014*

Listening Experiment

Functional data suggest that also listening abilities of dyslexics appear behaviorally 'normal' in silence and in moderate noise for short durations, they are associated to higher metabolic consumption as in normal readers.

Listening effort is greater for dyslexics in a speech-in-speech and speech in noise listening situation.

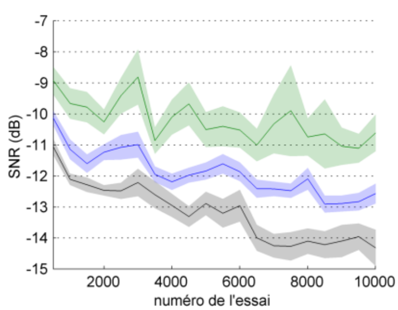
31

Dyslexics and musicians

ada / aga

The SNR was adapted continuously to ensure a correct response rate of 75%.

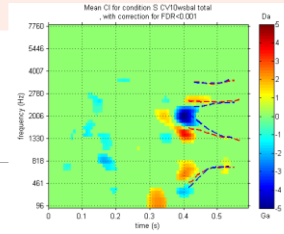
$SNR_{musicien} < SNR_{contrôle} < SNR_{dyslexique}$



Evolution of mean SNRs (± s.e.m.) for Controle (blue, N=19), Dyslexics (green, N=18) and Musician (grey, N=19)

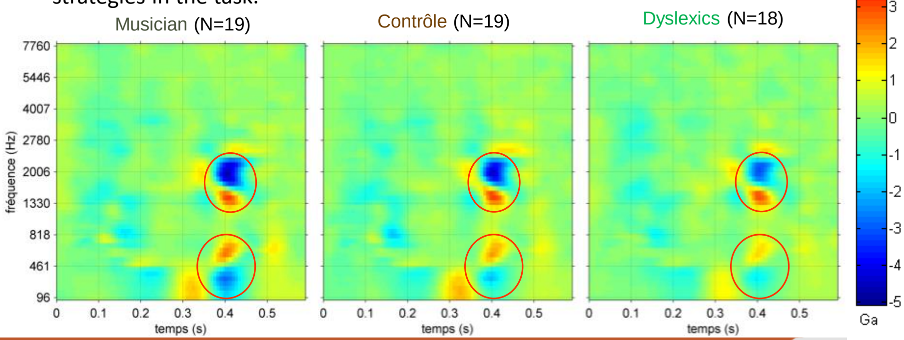
32

ACIs for musicians and dyslexics on [da]-[ga]

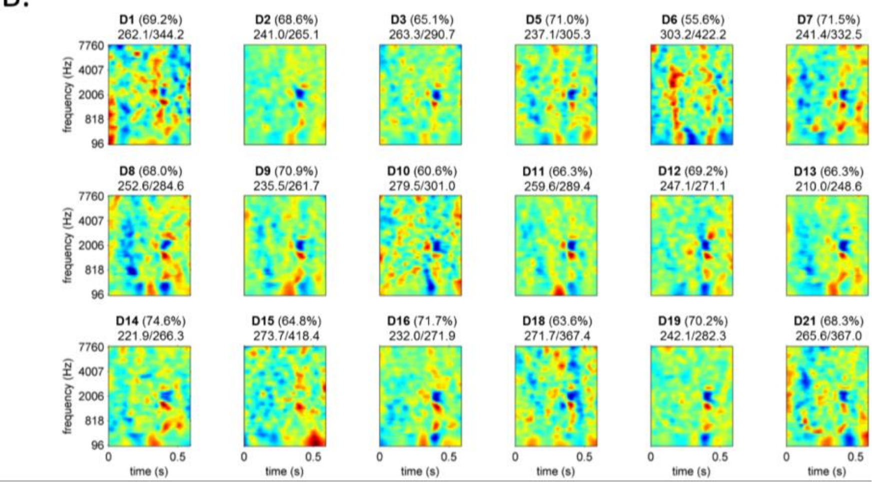


=> overall similar listening strategies: the first formant onset and the onsets of the second and third formants onsets.

- but the musicians relied more heavily on the main acoustic cues and responded more consistently to stimuli.
- participants with dyslexia show more individual, less efficient, weighting strategies in the task.



B.

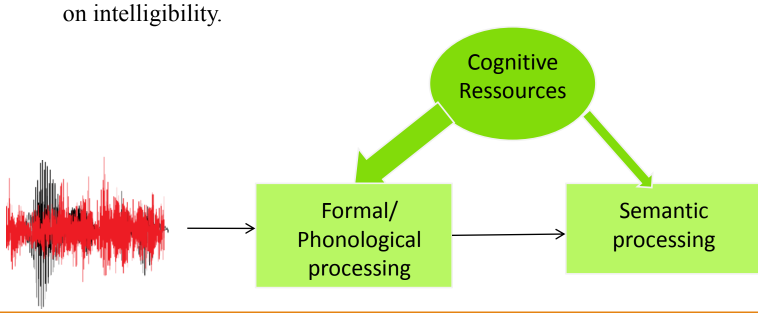


Perspectives **Speech in noise and Executive functions**

- Lack of intelligibility generates a reallocation of cognitive resources on low-level processes. This reallocation is deleterious to semantic processing.

⇒ Effortfulness Hypothesis (Wingfield, 2005).

- Semantic processing depends on cognitive resources themselves depending on intelligibility.



35

Collaborations

PhD students

M. Dole
DDL, Lyon

A. Gautreau
CNRL, Lyon

L. Varnet
LPP, Paris

G. Estivalet
L2C2, Lyon

C. Grattaloup
DDL, Lyon

S. Pota
PC, Grenoble

M. Dekerle
L2C2, Lyon

M. Tassin
L2C2, Lyon

Post-doc

V. Boulenger
DDL, Lyon

J. Meyer
Gipsa-Lab, Grenoble

S. Riès
San Diego U, USA

L. Dentel
Univ. Fed. de Para Belem, Br

Researchers

E. Spinelli
LPNC, Grenoble

N. Grimault
CRNL, Lyon

M. Hoen
oticon

K. Knoblauch
SBRI, Lyon

W. Serniclaes
LPP, Paris

S. Donnadieu
LPNC, Grenoble

M. Perrone-Bertolotti
LPNC, Grenoble

Hospital practitioners

E. Veuillet
CRNL, Lyon

E. Truy
O.R.L.
CRNL, Lyon

Thank you for your attention !