Psycholinguistics and neurolinguistics of spoken word identification

Fanny Meunier

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

What is psycholinguistics?

- Psycholinguistics deals with the mental processes a person uses in producing and understanding language.
  - **language comprehension** (how we perceive and understand speech and written language)
  - **language production** (how we construct an utterance from idea to completed sentence)
  - **language acquisition** (how human beings learn language).

- Two main questions:
  - What knowledge of language is needed for us to use language?
  - What processes are involved in the use of language?

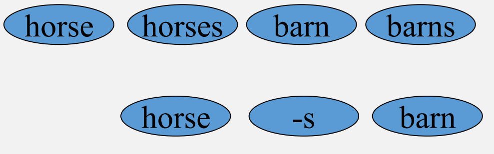
How language units are organised in the brain ?

- Mental lexicon
  - High capacity: 40,000 – 60,000 words
  - Fast: Recognition in as little as 200ms (often before word ends)
    - How do we search that many, that fast

    – > suggests that there is a high amount of organization

What is the format of the units stored in LTM?

Finding the balance between storage weight and processing cost:



- Words primitives
  - Need a lot of representations
  - Fast retrievial
- Morpheme primitives
  - Economical - fewer representations
  - Slow retrieval - some assembly required
    - Decomposition during comprehension
      - *apartment–apart*
      - *Bikini* (1946) named after the Bikini Atoll in the Pacific Ocean where the US did nuclear weapons test. The swimsuit would create an "explosive commercial and cultural reaction".

      => « *–kini* family » : monokini, tankini, trikini, pubikini, burkini ...
    - Composition during production

=> Identifying parameters that modulated word recognition

Lexical Decision task (e.g., Taft, 1981)

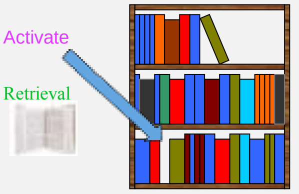
- See a string of letters / listen spoken items
- As fast as you can determine if it is a real English word or not
  - "yes" if it is
  - "no" if it isn't
- Typically speed and accuracy are the dependent measures

table	Yes
vanue	No
daughter	Yes
tasp	No
cofef	No
hunter	Yes

Factors affecting lexical organization

- Morphological structure
- Phonological structure
- Frequency of use
- Imageability, concreteness, abstractness
- Semantics
- Grammatical class

Lexical access



- While structure is important, so are the processes that may be involved in activating and retrieval the information
- How do we retrieve the linguistic information from Long-term memory?
- What are the information needed to reach lexical access?

=> The modality of input comes into play.

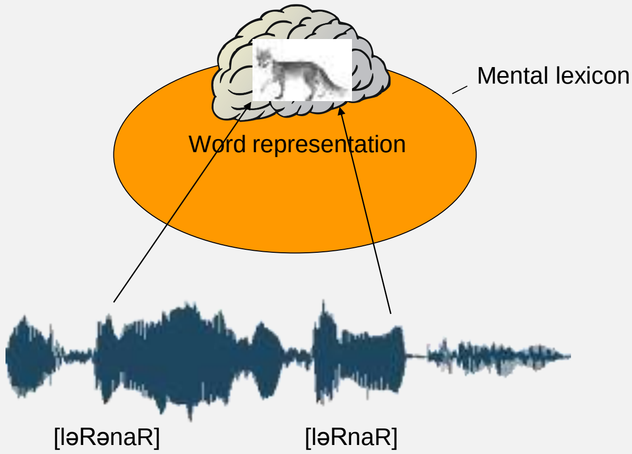
Speech

Major theoretical issue: lack of invariance.

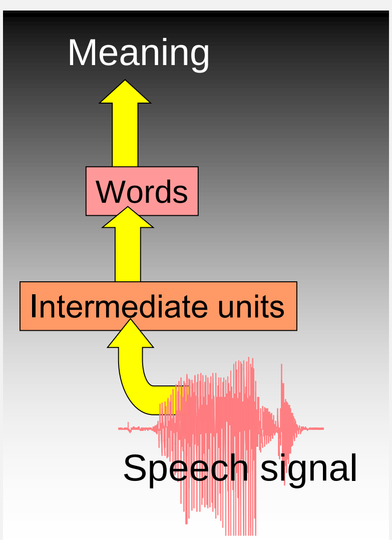
- Talker variation (Allen & Miller, 2003; Jongman, Wayland & Wang, 2000; Peterson & Barney, 1955).
- Influence of neighboring phonemes (coarticulation) (Fowler & Smith, 1986; Delattre, Liberman & Cooper, 1955)
- Speaking rate variation (Miller, Green & Reeves, 1986; Summerfield, 1981)
- Dialect variation (Clopper, Pisoni & De Jong, 2006)

« Many to one problem »

Match sensory input that comes from acoustic signal to the mental representations of words stored in long term memory.
=> Many computational steps to parse and convert acoustic waves into discrete linguistic units from which meaning can be extracted.



The interface between speech perception and spoken word recognition.



By the end of *Speech Perception* processing, a noisy acoustic signal is carved up into discrete units.

Mapping of acoustics to sublexical units and then Sublexical units to lexical ones.

What is the proccessing/functional unit of speech?

- Syllable
  - unit of organization for a sequence of speech sounds.
    - *wa.ter*.
  - typically made up of a syllable nucleus (most often a vowel) with optional initial and final margins (typically, consonants).
- Phoneme
  - hypothesized to be the smallest contrastive units that change a word's meaning
    - e.g., /b/ and /d/ as in bad versus dad)
- Set of Phonetic features
  - distinctive features composing phonemes
  - Allow underspecification

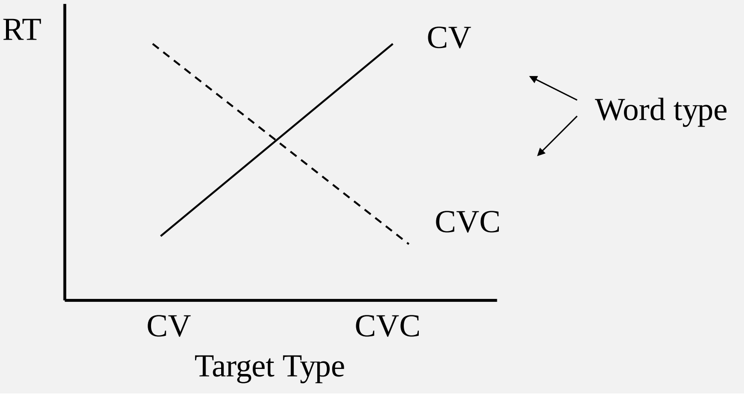
Role of the syllable in the recognition of spoken words.

Syllable Detection and Syllabic Compatibility

(Mehler et al., 1981, JVLVB)

Word type: CV / CVC (ex : BA.LANCE vs. BAL.CON)

Target Type: CV / CVC (ex : BA vs. BAL)



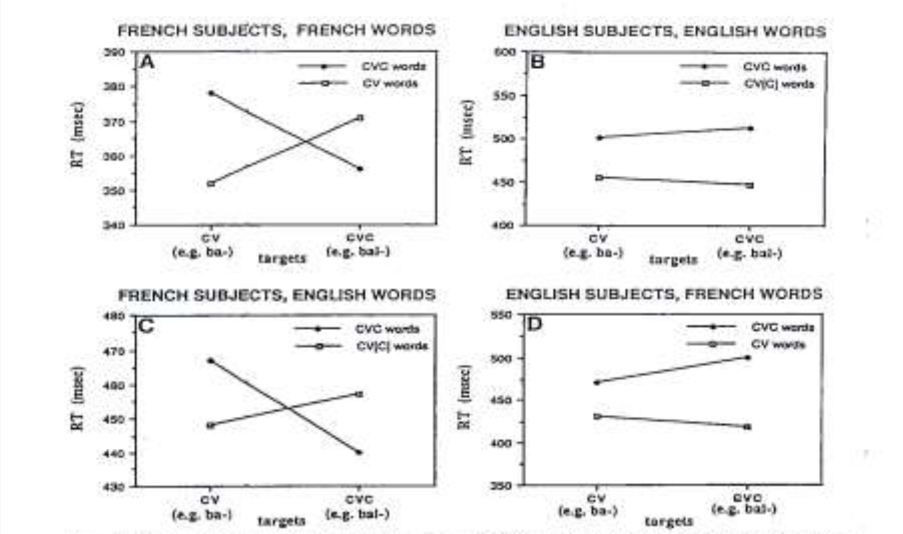


FIG. 1. Mean target detection response time (RT) in msec as a function of size of target sequence (CV, e.g., *ba-*, versus CVC, e.g., *bal-*) and size of initial syllable of stimulus word (CV versus CVC for French; CV[C] versus CVC for English) in preceding studies, for four combinations of subjects' native language and stimulus presentation language: (A) French subjects and words; (B) English subjects and words; (C) French subjects, English words; (D) English subjects, French words.

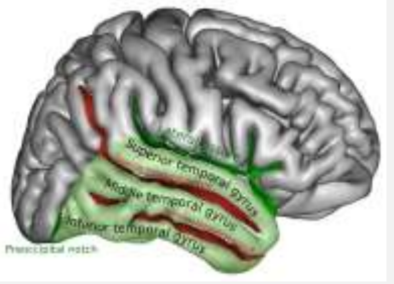
(Cutler et al., 1986, JML)

"Sequencing at the syllabic and supra-syllabic levels during speech perception: an fMRI study." Deschamps & Tremblay, 2014

- The level of processing (e.g., acoustical/phonetic, phonemic, syllabic, supra- syllabic) at which mechanisms implemented within the the supratemporal plane (STP) and superior temporal sulcus (STS) that might be involved in the conversion of acoustical information into "phonological" representations? operate remains unclear.

- Syllabic and supra-syllabic information are processed automatically during passive speech listening. [..]

- "Future studies need to examine whether the processing of sub-lexical information is automatic and necessary during language comprehension using more nat-uralistic stimuli such as words or connected speech."



Variability in the production

The same sentence can vary considerably in its acoustic forms:

« he is telling you he doesn’t know it »

[il.t .di.kil.n .l.s .pa]	" il te dit qu’il ne le sait pas "
[il.t .di.kil.n .ls .pa]	" il te dit qu’il ne l’sait pas "
[i.t .di.kil.n .ls .pas]	" ‘ite dit qu’il ne l’sait pas "
[it.di.ki .s .pa]	" ‘it’dit qu’il’sait pas "

SCHWA ELISION : Influence of frequency of variants

2 types of schwa :	Frequence +	Frequence -
Mandatory : casserole	/ kasr**ol** /	/ kasər**ol** /
Optional : renard	/ RənaR /	/ RnaR /

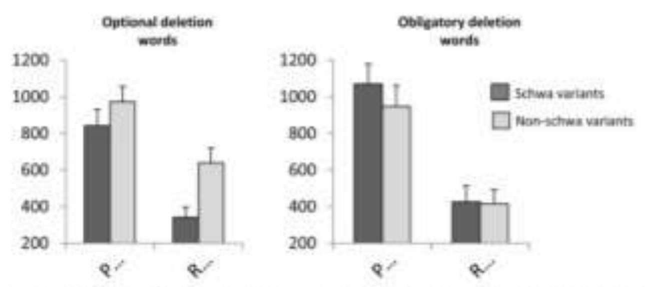


Figure 1. Mean reaction times from word offset as a function of group (pre-readers [P] vs. readers [R]) and variant type (schwa vs. non-schwa variants) for optional deletion words (left panel) and obligatory deletion words (right panel). The bars represent the 95% confidence intervals.

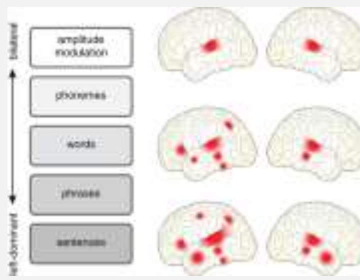
Racine, Bürki & Spinelli (2014)

The success of the phoneme

"Neurocognition of Language/Speech Comprehension and Speech Production" **WikiBook**

"The starting point of our way to comprehend an utterance is the acoustic sequence reaching our auditory system. As a first step, we have to separate the speech signal from other auditory input. […]. Next, we have to identify as segments the individual sounds that form the sequential speech signal, so that we can relate them to meaning.

This early part of speech comprehension is also referred to as **decoding**. The smallest unit of meaning is the **phoneme**, which is a distinguishable single sound that in most cases corresponds to a particular letter of the alphabet. However, letters can represent more than one phoneme like "u" does in "hut" and "put". […]



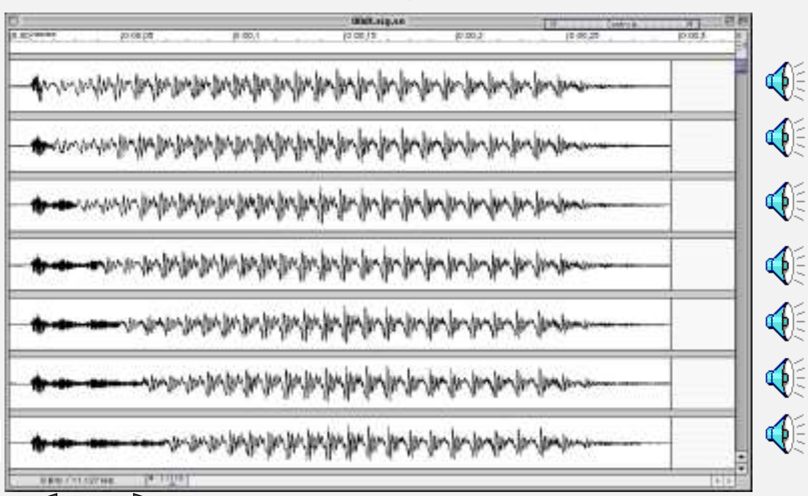
Peelle, 2012

Categorical Perception

Voice Onset Time (VOT): a feature of the production of stop consonant = the length of time that passes between the release of the consonant and the onset of voicing = the vibration of the vocal folds.

[dæ]

[tæ]



60 ms

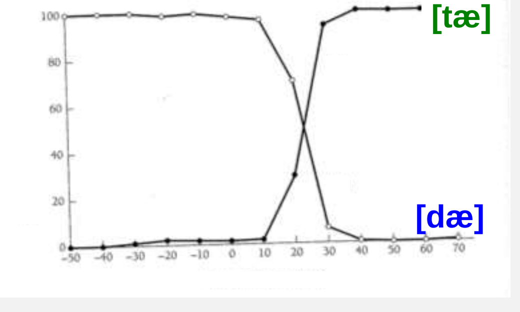
Categorical Perception

Within-category discrimination is hard, across-category discrimination is easy

D	0ms	20ms	D
D	20ms	40ms	T
T	40ms	60ms	T

Categorical Perception

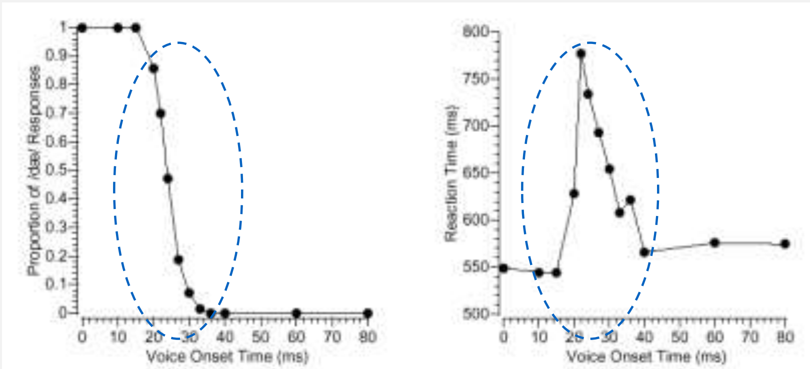
% of responses as either [tæ] or [dæ]



Voice onset time in msec

Categorical Perception

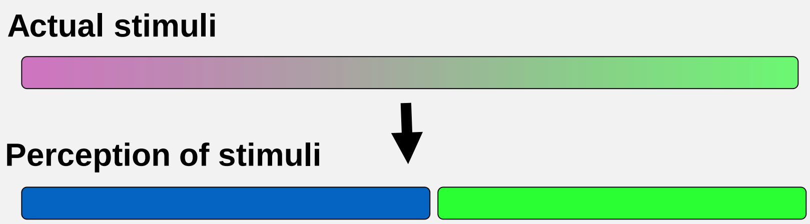
Uncertainty at category boundary



Decision between dæ/tæ　　**Time to make decision**

Categorical Perception meets our contraints

A range of stimuli that differ continuously are perceived as belonging to only two categories with no degrees of difference within those categories.



- Sharp identification of tokens on a continuum.
- Discrimination poor *within* a phonetic category.

Subphonemic variation in VOT is discarded in favor of a discrete symbol (phoneme).

Categorical Perception

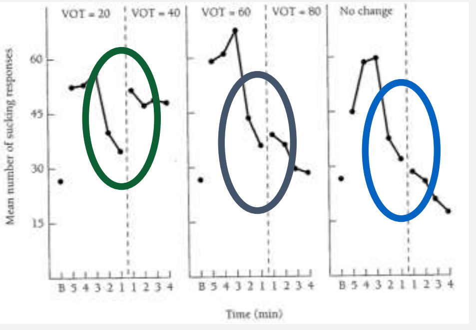
Infant (4 months) categorical perception:
Voice Onset Time (VOT)

Eimas et al. 1971: HAS technique

High Amplitude Sucking (HAS):

across category | within category | control (baseline)

Infants notice, compared to control

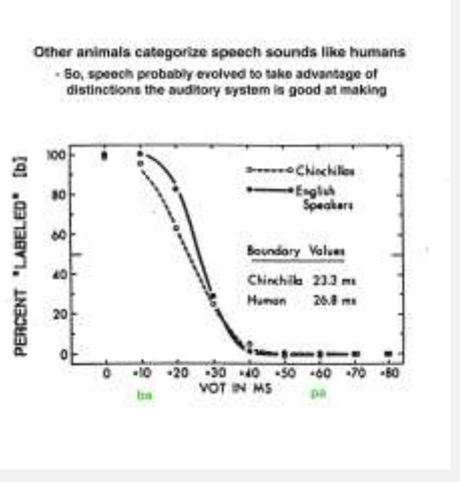


Infants don't notice, compared to control

Categorical perception is not specific to the human speech, it's a feature shared with other mammals like chinchillas.

Japanese quail

Human speech **takes advantage of** properties of auditory system by generally using the differences that are easy to hear to signal important contrasts in the language.



(*chinchillas*) Kuhl & Miller (1978). *JASA*, *63*, 905-917

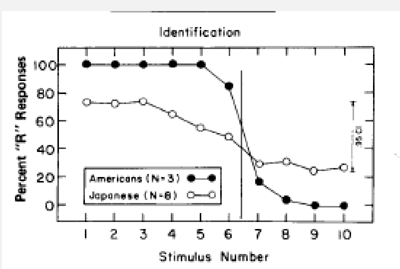
- Also observed for chords (e.g. A minor/Amajor) (McMurray et al. 2008)
- Not accessible to illiterates

When subjects respond *phonologically*, performance looks categorical…:

- But much of the experiemental work is characterized by : explicit, metalinguistic tasks; off-line measures; non-word stimuli (see Schouten (2003). The end of categorical perception as we know it).
- Depends to some extent on the particular task demands (Scott & Evans, 2010). Evidence against categorical perception from:
    - Discrimination task variants (Pisoni and Tash, 1974, Pisoni & Lazarus, 1974, Schouten, Gerrits & Van Hessen, 2003; Carney, Widden & Viemeister, 1977)
    - Rating tasks (Massaro & Cohen, 1983; Miller, 1997)
    - Identification tasks.

Japanese vs. English /r/-/l/ distinction

(Logan et al., 1991)



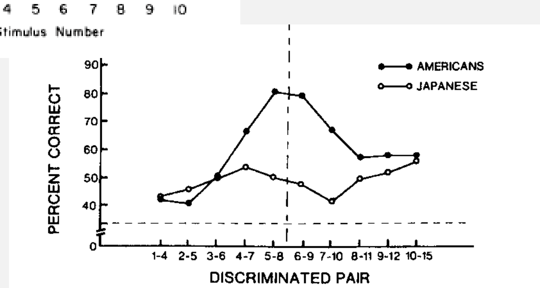


Figure 12.2. Test of the categorical perception of /ra/ and /la/ by American and Japanese adults. American listeners show the characteristic peak in discrimination at the phonetic boundary; Japanese listeners do not. (From Miyawaki et al., 1975.)

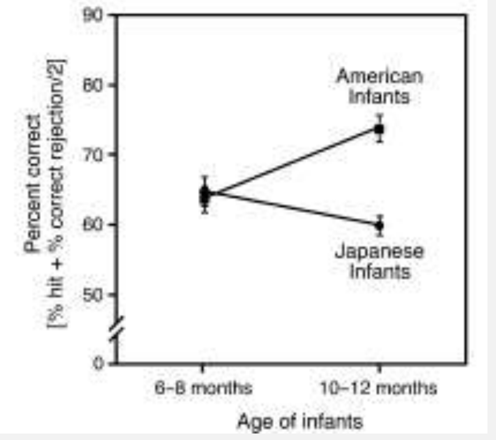
RA

LA

Infants lose the ability to discriminate nonnative contrasts

(Kuhl et al., 2006, *Dev. Science*)

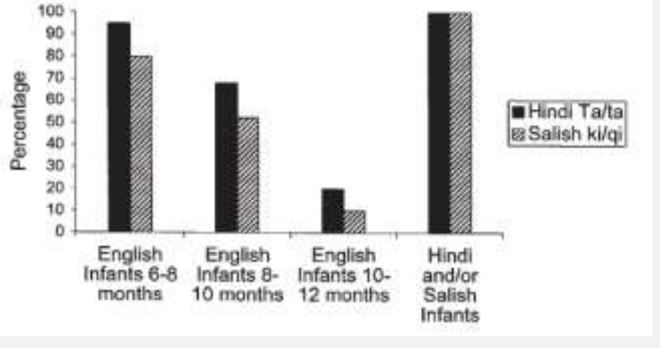
Effects of age on discrimination of **/ra-la/** by American and Japanese infants at 6–8 and 10–12 months (percent correct scores).



Between 6/8months and 10/12months:

• clear improvement for americans
• decline for japanese

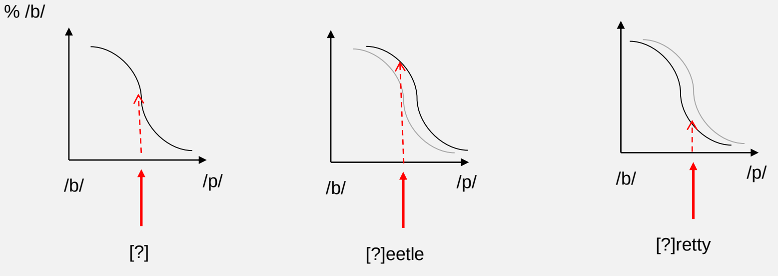
Different patterns for different contrasts



Werker & Tees, 1984

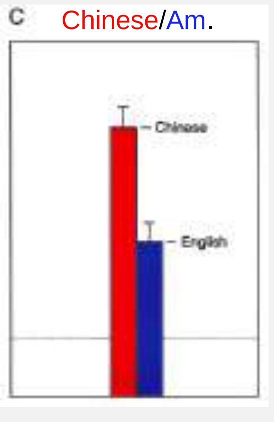
Bottom up Influence

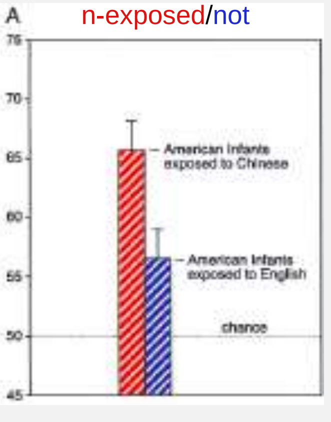
Ganong Effect (1980)
Lexical biais in the interpretation of the ambiguous phoneme



Foreign-language intervention in infancy affects phonetic perception (Kuhl, Tsao, & Liu, 2003, *PNAS*)

Exposure to live Mandarin Chinese speakers in 12 laboratory sessions (totaling ≈5 h) at 9 mo of age resulted in phonetic learning and altered the typical developmental time course of foreign-language speech perception.





Contrast discrimination of Mandarin qi-xi at 10-12 months

WHISTLED LANGUAGES AND SPEECH PERCEPTION

julien.meyer@gipsa-lab.fr (Gipsa-lab, UMR 5216, Grenoble)

https://sites.google.com/site/julienmeyerlab/publications

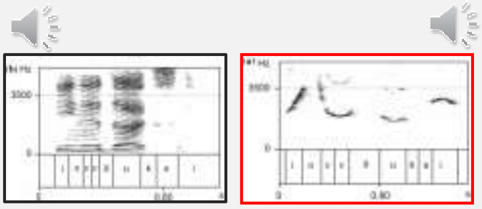
Meyer J (2016) “Whistled Languages”Reveal How the Brain Processes Information. Scientific American. 22 Nov 2016 Nature Publishing Group

Laure Dentel

fanny.meunier@unice.fr (BCL, UMR 7320, Nice)

Whistling languages

- Whistled adaptation of the spoken language: similar complexity in terms of syntax and vocabulary.
- Use the modulations of the whistle instead of those of the vibrations of the vocal cords.



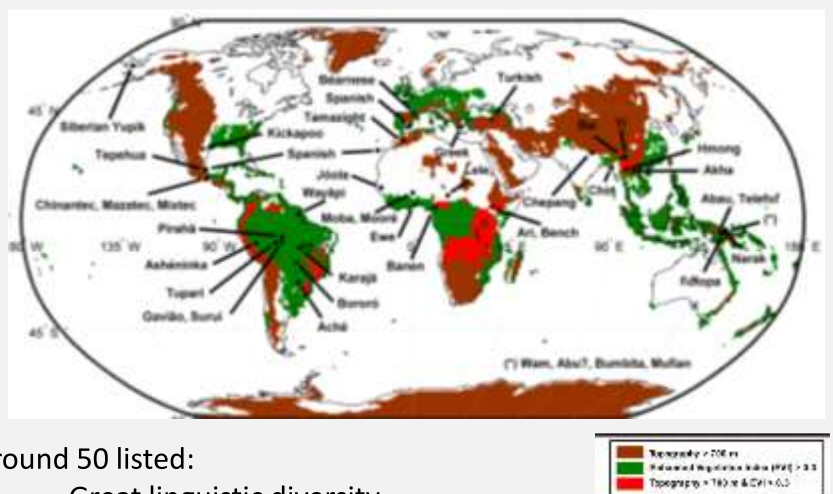
Same sentence : voice & whistle (Grec: “ειναι ενταξει’ (meaning ‘all right’) – emulation of the quality of spoken vowels and consonants (formant-based whistling))

➡ Drastic signal reduction

➡ Selection by the whistler of acoustic cues

Great interest to find the acoustic cues needed to understand speech.

WHISTLED SPEECH – WORLD LANGUAGES



Around 50 listed:

Great linguistic diversity,

Spread around the world

Essentially in:

mountainous area or area with dense vegetation

Diversity and typology of whistled speech
Different coping strategies

Moore & Meyer, 2015

Tonal Languages

Type 1: pitch transposition

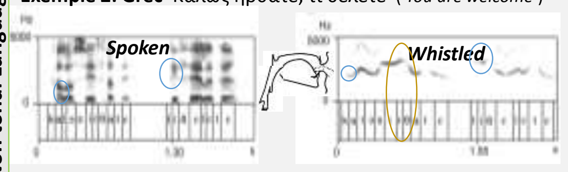
Example 1: Gavião 'jaá pa-víjí-á' ('Let's go swimming')

Recognition
Tonal languages
Stereotyped phrases

Non tonal Languages

Type 2: Formantic transformation/synthesis

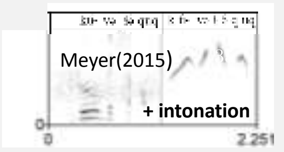
Exemple 2: Grec 'Καλώς ήρθατε, τι θέλετε' ('*You are welcome*')



Non Tonal Languages
- Sentences 80% (Busnel, 1970)
- Words 70 % (Meyer, 2005)
- VCV 60% (Rialland, 2005)

Exemple 3: Siberian Yupik

Type 3: intermadiate



For Language Science

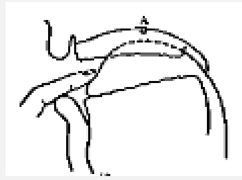
Alternative approach of the Phonetics / Phonology interface

Different types of adaptation of speech signal

- Allows longer distance communication
- Blend into the sounds of the forest.

Different technics

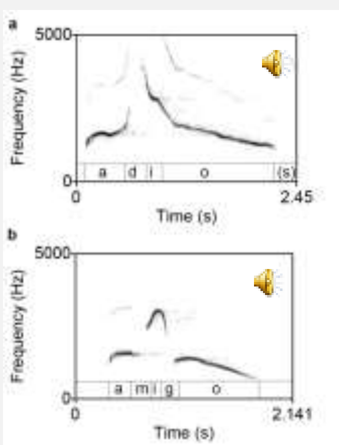
Whistling



Drawing from radiography, André Classe.

- The whistling technique does not involve the vocal cords.
- The intensity of the whistling is modulated by the shape of the mouth (close for each vowel of the form in the spoken language).
- The lips remain fixed, rigid and cover the teeth. The acoustic impedance of the lips does not vary.
- But the posterior two thirds of the tongue can rise or fall, thus varying the volume of the oral cavity and its resonance with the possibility of tuning it.

WHISTLED SPEECH



Drastic constraints on source/articulation

Complex frequencies of the voice

Phonetic reduction

simple melody

Untrained listeners => non intelligible

<u>But if trained:</u>

- High intelligibility scores

90% non stereotyped sentences (Busnel 70)

70 % isolated words (Meyer 2005, PhD)

=> Informational skeleton of speech

acoustico-phonetic ‘primitives’?

Questions:

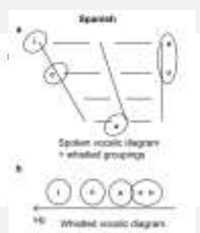
- Can a native and naive speaker correctly categorize whistled vowels?
- Is learning fast?
- Does categorization dependent on the language of the speaker?
  - If so, to what extent?

Participants

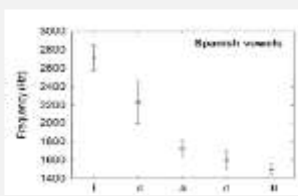
- 1 reference Spanish whistler
- 20 naïve Spanish speakers - language of stimuli
- 20 naïve French speakers
- 19 naive Chinese speakers - tonal language

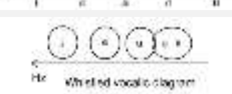
EXP. VOWEL CATEGORISATION

Meyer, Dentel, Meunier, 2017



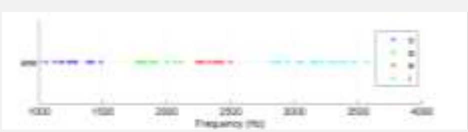
The whistle of the vocalic system
Vocal segments ⇔ whistled at frequency intervals
The distribution of these vocalic intervals follow the same pattern in all the non tonal languages already studied (except for /u/ often due to rarity).





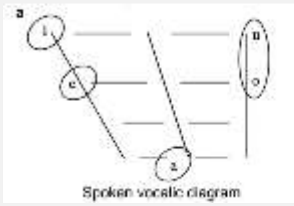
Stimuli:

Spanish whistled vowels extracted from sentences (nuclei without consonant modulations)

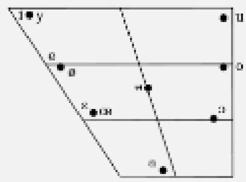


- Frequencies: from 1 kHz to 3.7 kHz
- Lenghts: quasi-stationairy part (85 ms to 1 s), n.s. effect
- Amplitudes: normalized on the max

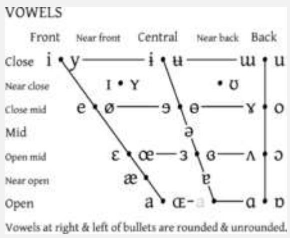
INFLUENCE OF THE NATIVE VOCALIC SYSTEM



Spanish

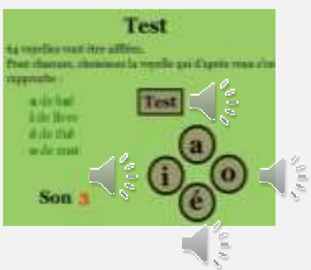


French



Chinese

Experiement on vowel categorisation



Task

- Vowel categorisation
- Forced choice between 4 categories of vowels (4-AFC)
- 3 Phases (1 = 3, 2 = test)

Procedure

- Phase 1: 40 whistled vowels (10 for each vowel with 4 different productions)
- Phase 2: 64 whistled vowels (16 for each vowel with 16 different productions)
- Phase 3 = Phase 1

=> 144 stimuli

Results: Spanish whistler Reference

Phase 2 (Meyer, 2008)

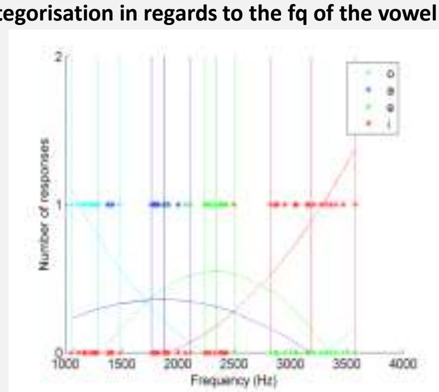
- 87.5 % of correct answers
- Vowel categorisations
  « substantial » agreement (kappa=0.78)

Categorisation in regards to the fq of the vowel played

Answered vowels - %

Played Vowels	« o »	« a »	« e »	« i »
/o/	*87.50*	**12.50**	0	0
/a/	**6,25**	*75*	**18.75**	0
/e/	0	**6.25**	*87.50*	**6.25**
/i/	0	0	**0**	*100*

- /i/>/o/=/e/>/a/

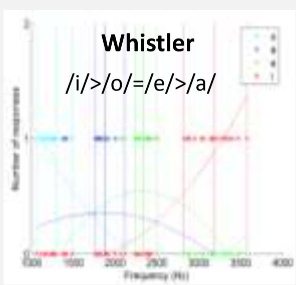


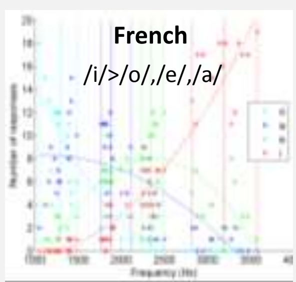
Estimated curves of the answers averaged by polynomial interpolations of the second order. The maxima of the estimated curves of whistler's answers are always within 5% of variance of the range of variation of the vowels.

Results

87.5 %

« substantial » agreement (kappa=0.78)





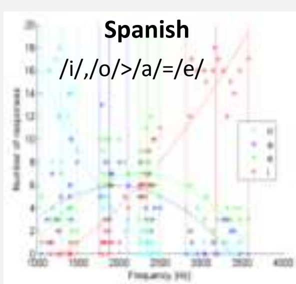
55 %

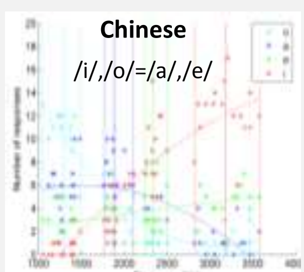
« moderate » agreement (kappa= 0.4)

53 %

« moderate » agreement (kappa=0.37)

Learning between phases 1-3= 3.4%





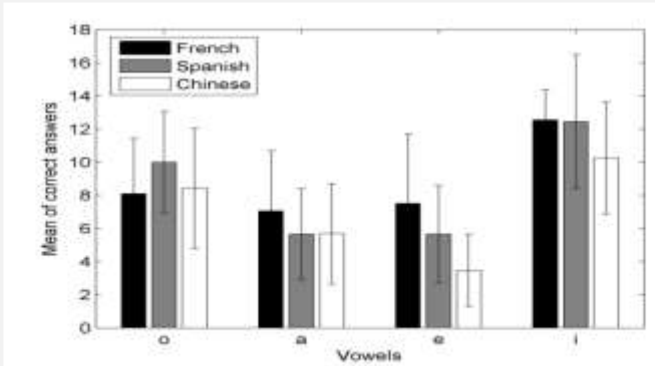
43,5 %

« low fair » agreement (kappa=0.25)

No learning between phases 1-3

Results

Correct answers

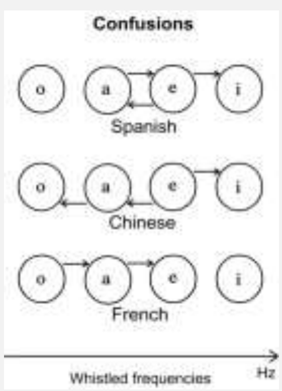


No clear effect of density.

Importantly, the vowels /i, e, a, o/ also exist in French with similar or close pronunciations (Calliope, 1989).
In Chinese /e/ appears only in some very specific and quite rare contexts (syllable-finally after a palatal glide or in diphthongs before [i̯] (Lee and Zee, 2003, Duanmu 2007)).

47

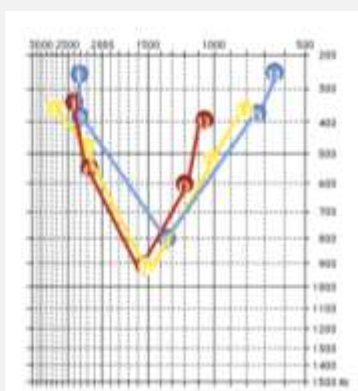
Results



Multiple t-tests with Bonferroni correction

Vocal systems:
- position
- scope
- density

Blue: French
Red: Chinese
Yellow: Spanish



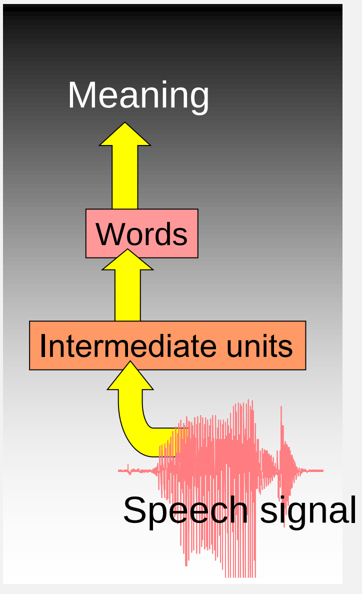
/i/ very extreme in French

48

Conclusions

- Spanish and French natives have performances above chance: they managed to extract the information necessary for categorization.
- But no clear superiority of the natives of the whistled tongue.
- Confusions suggest differences depending on the native language.

⇒The task is not purely acoustic
⇒Influence of native vocal space



If phonemes are not the intermediate units then we should find:

1) Continuous sensitivity to fine-grained detail in the signal.

2) That fine grained detail actually affects higher level (lexical) processes and that information is retained and used to improve word recognition.

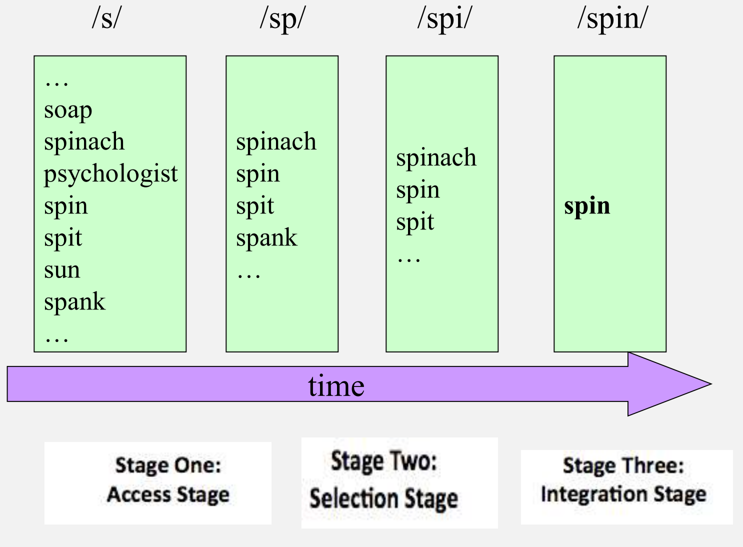
Is sensitivity to subphonemic differences gradient?

Need:

- tiny acoustic gradations
- online, temporal word recognition task

Does activation for lexical competitors during online recognition reflect continuous details?

Cohort model (Marslen-Wilson et al., 1987, 1997)

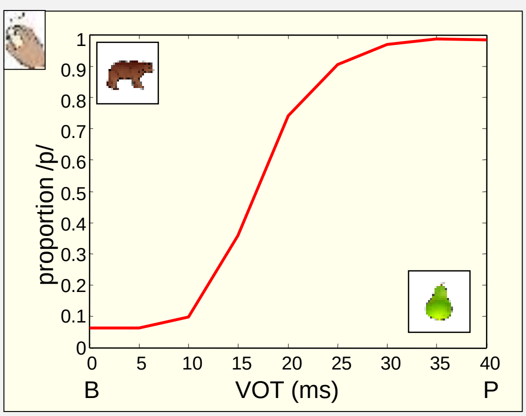


VISUAL WORD PARADIGM

Repeat 1080 times

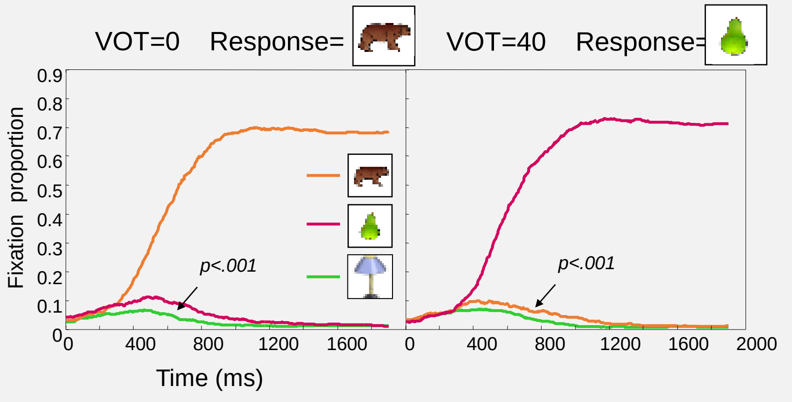
McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. Cognition, 86, B33–B42.

Identification Results



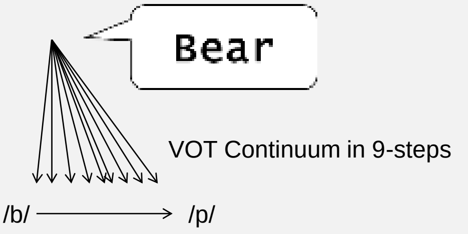
High agreement across subjects and items for category boundary.

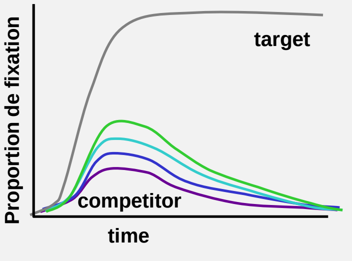
By subject: 17.25 +/- 1.33ms
By item: 17.24 +/- 1.24ms



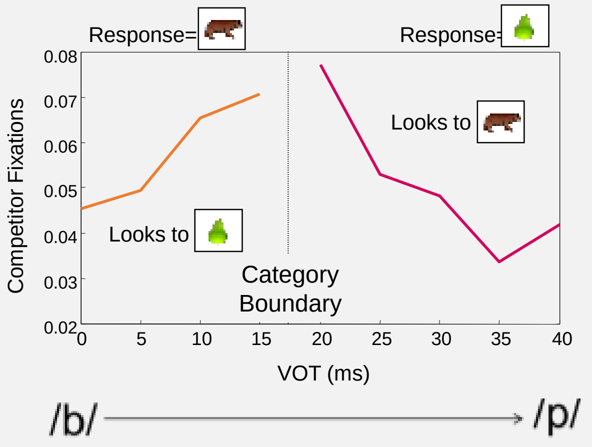
More looks to competitor than unrelated items.

VISUAL WORD PARADIGM





How often was the subject looking at the “pear”?



Lexical activation is sensitive to continuous acoustic details.

- Subphonemic acoustic differences in VOT have gradient effect on lexical activation.
  - Refutes strong forms of categorical perception
- Fine-grained information in the signal is not discarded prior to lexical activation.
- Extends to vowels, l/r, d/g, b/w, s/z *(McMurray & Tanenhaus et al., 2008, 2009)*
- Does not work with phoneme decision task *(McMurray, Aslin, Tanenhaus, Spivey & Subik, 2008)*

Could that help for homophonic sequences ?

« One to many problem »

- There are no clear word boundaries in spoken language.
- Due to elision, some spoken utterances in French are phonemically identical (e.g., l'amie 'the friend' vs. la mie 'the crumb', both [lami]).
- Correct segmentation into discrete word units is necessary for comprehension.
- There are, some acoustic differences between members of ambiguous sequences that could be exploit by listeners.

Phonemically identical pairs

30 phonemically identical pairs

- clear intonational differences between the two sequences.
- rise in fundamental frequency beginning at the left edge of the first content word syllable.

C'est la fiche ('it's the sheet') | **C'est l'affiche** ('it's the poster') ⇒ [selafiʃ]

Acoustic analyses

	C'est la fiche	C'est l'affiche
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

	C'est la fiche	C'est l'affiche
F2	[a] = 1834 Hz	[a] = 1808 Hz

⇒ More front articulation for the determiner *la*

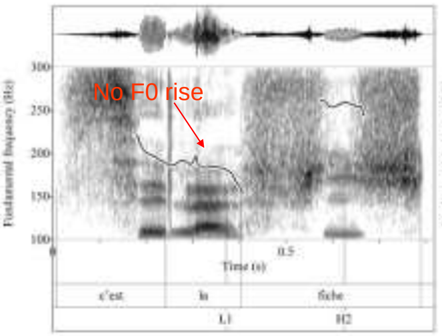
	C'est la fiche	C'est l'affiche
F0	[a] = 179 Hz	[a] = 195 Hz

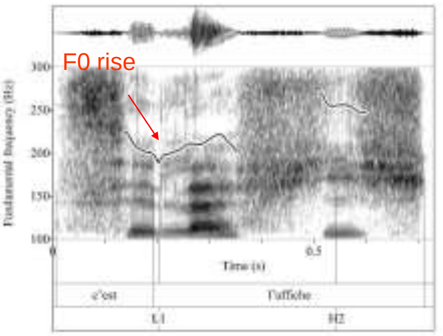
⇒ F0 rise at the beginning of content words

Intonational cue

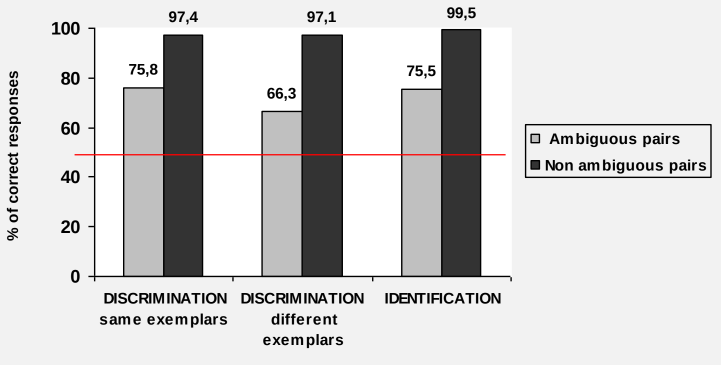
[selafiʃ]

C'est la fiche ('it's the sheet') **C'est l'affiche** ('it's the poster')





RESULTS

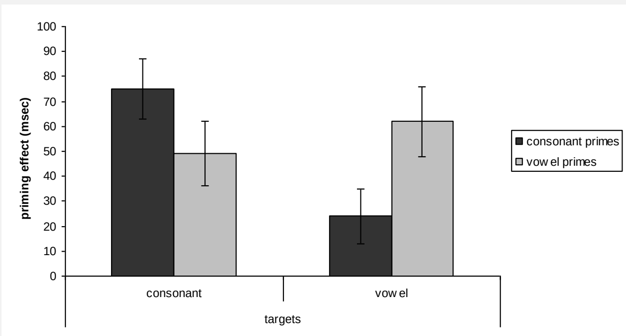


Listeners can retrieve the correct segmentation in such utterances: they can discriminate between and identify them

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

Spinelli et al. (2007).

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.

Automatic fine-grained speech segmentation despite intra-speaker variability: an electroencephalography study.

Stéphane Pota, Elsa Spinelli, Léo Varnet, Michel Hoen, Fanny Meunier (2017, sub.)

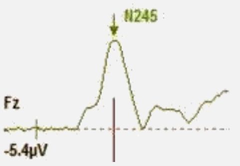
- Investigation the electrophysiological correlates of subphonemic perception in homophonic nominal utterances, such as *l'amie* 'the friend' vs. *la mie* 'the crumb', both [lami].

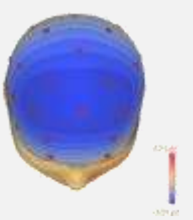
- We examined Mismatch Negativities (MMN) with a variant of the oddball paradigm while healthy participants heard syllables (Experiment 1) and words (Experiment 2) in an experimental set up that preserved natural intra-speaker variability.

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

- detection of unexpected changes in some aspects of a regular continuous auditory stream
- oddball paradigm: one rare sound (deviant) occurs in a series of frequent stimuli (standards).
- fronto-central negative wave peaking between 100 and 300 ms after the deviance onset.
- 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)





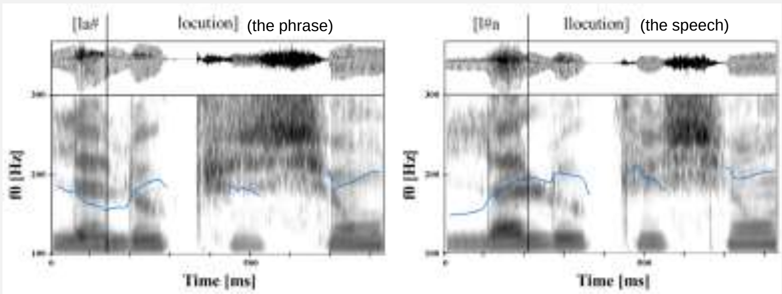
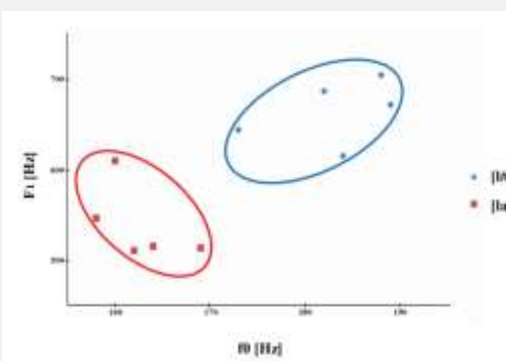


Figure 1. Spectrogram and waveform of one token of [la#locution] (at the left) and one token of [l#allocution] (at the right).
=> rise in fundamental frequency at the beginning of the content word syllable



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

5

- Procedure :

		Bloc 1	Bloc 2
Expérience CV (16 sujets)	**Standard**	L'A	LA
	Deviant	LA	L'A
Expérience MOT (16 sujets)	**Standard**	L'ALLOCUTION	LA LOCUTION
	Deviant	LA LOCUTION	L'ALLOCUTION

- 1800 Stims : Std (p=.80), Dev (p=.20)

- EEG :
  - **Biosemi EEG Systeme 32 electrodes**
  - **SR : 2048 Hz – BP : [0.1 - 400 Hz]**
  - **Temporal window : [200-300ms]**

6

Results and Discussion

- no MMN was observed in either experiments 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.

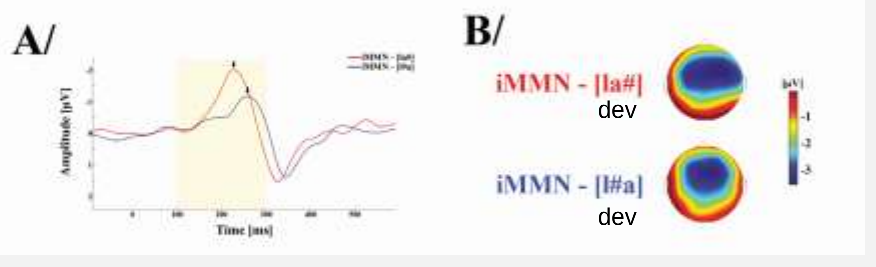
⇒intra-speaker variability was not considered deviant by the neural system = the MMN is a good tool for tracking the meaningful changes of acoustical stimulation.

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

- Under the homophone conditions, our results clearly revealed MMNs for both syllables and nominal sequences, despite the variability of standard tokens.

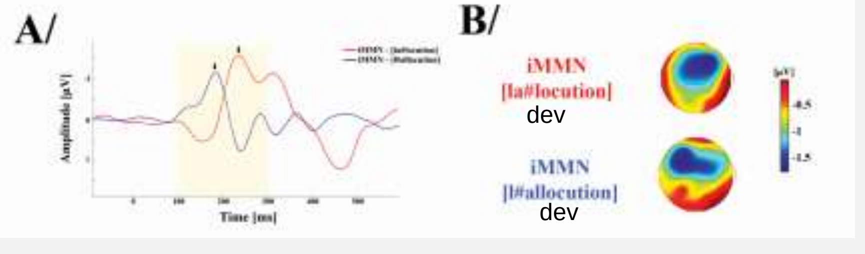
EEG data for ExpSyll.

electrode Fz (averaged mastoids reference).



EEG data for ExpWord

40 ms time-window centered on the amplitude peak



RESULTs (2)

Asymetric timing LA / L'A for the 2 experiements

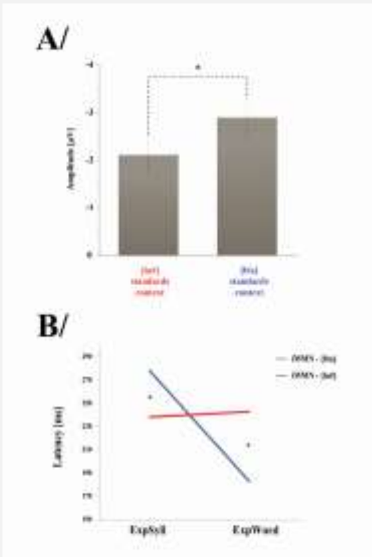


Figure 6.

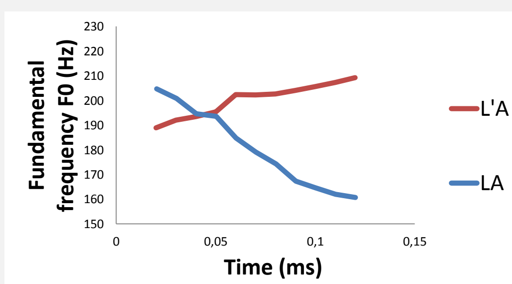
A/ Mean MMN amplitude (homophone as non-homophone) for each experimental block. At the left, the [la#] standards context, and at the right, the [l#a] standards context. The ExpSyll data are represented here. The same pattern was observed in ExpWord.

B/ Comparisons between the latency of the homophone iMMNs for both experiments.

Contrast LA – L'A

- *l'a* as a standard showed more negative MMNs than the blocks with *la* as standards in the 2 experiements:
  - The initial syllable duration and the mean first formant value on the vowel [a] showed greater variability in the different tokens of *la* compared to those of *l'a*.
  - *l'a*, which is less variable may produce a more refined sensory memory trace, allowing better deviant detection. This memory trace could lead to more precise predictions, thereby increasing sensitivity to deviants and amplifying MMNs.

Difference in the dynamic of the F0s



La : the determiner

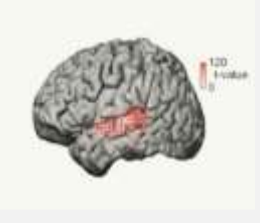
- The deviant *l'a* elicited a later MMN peak latency than its homophone *la* in ExpSyllable (+39 ms), while it appears earlier (-59 ms) in ExpWord. This result could be due to the syllable [la] in Expword stimuli being a determiner and corresponding to one of the most common definite articles in French ($\text{the}_{\text{feminine}}$). Therefore, the speech perception system is overtrained for its understanding, and the earliness of its detection could be caused by an effect of statistical learning associated with long-term exposure.
- In ExpWord, the MMN for the *l'allocution* deviant was clearly lateralized to frontal-left sites, contrary to what was observed for MMNs elicited by *la locution* or *l'a* deviants.
  - 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).

How the superior temporal gyrus (STG) that participates in high-order auditory processing of speech, is encoding phonetic information ?

Mesgarani et al. (2014)

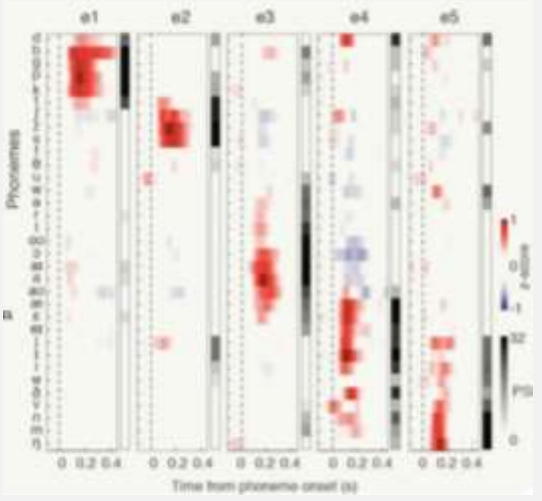
Many researchers have presumed that brain cells in the STG would respond to phonemes.

6 participants listening to 500 English sentences produced by 400 different speakers.



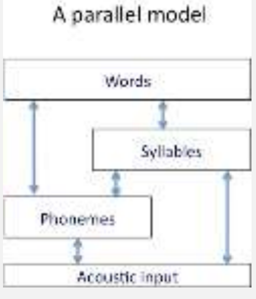
High-density direct cortical surface recordings in humans while they listened to natural, continuous speech to reveal the STG representation of the entire English phonetic inventory.

No invariant, local selectivity to single phonemes.
Multidimensional feature space for encoding the acoustic parameters of speech sounds.
Phonetic features defined by distinct acoustic cues for manner of articulation were the strongest determinants of selectivity, whereas place-of-articulation cues were less discriminable.
These findings demonstrate the acoustic-phonetic representation of speech in human STG.

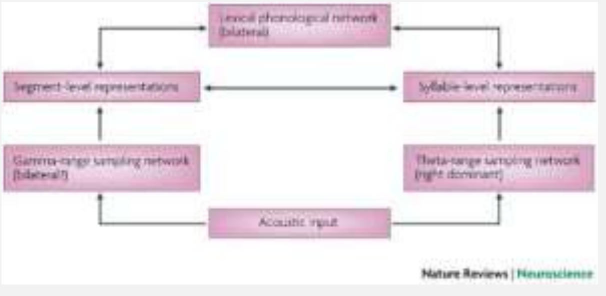


For examples: e1 respons to /p/, /t/, /k/, /b/, /d/ and /g/ => stops,
While e2 to sibilant (/s/, /ʃ/, /z/), made by directing a stream of air with the tongue towards the sharp edge of the teeth.

Adapted from
Hickok, 2009



Hickok & Poeppel
Nature Reviews
Neuroscience, 2007



One pathway samples acoustic input at a relatively fast rate (gamma range : 30-70 Hz) that is appropriate for resolving segment-level information, and may be instantiated in both hemispheres.
The other pathway samples acoustic input at a slower rate (theta range: 4-8 Hz) that is appropriate for resolving syllable level information, and may be more strongly represented in the right hemisphere.
These pathways interact, both within and between hemispheres, yet each appears capable of separately activating lexical phonological networks.

Summary

- 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 (the more stable ?) seem to be used early to constrain lexical acess and other cues don't seem to be used early.

- Which acousic cues are used ?

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

Thanks for your attention