Des filtres auditifs cochléaires aux filtres auditifs sociaux

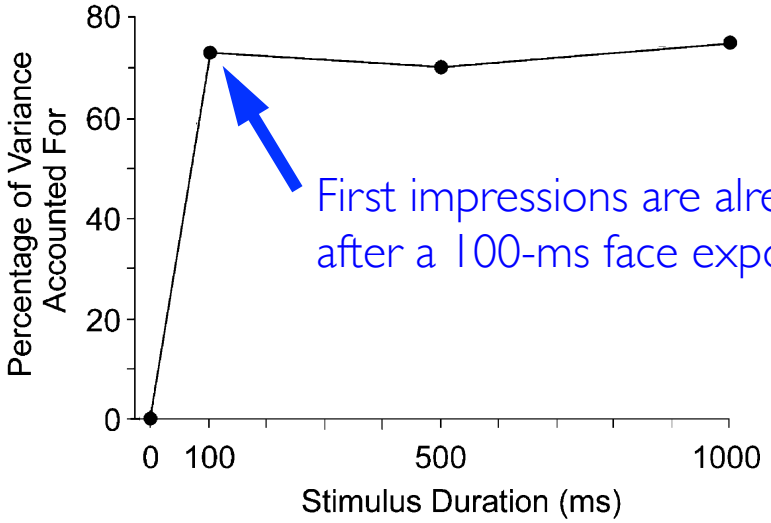
Emmanuel Ponsot - Ircam & ENS (Paris)

15 Juin 2017, Journées Perception Sonore (Brest)

Nous interprétons spontanément nos stimulations sensorielles.

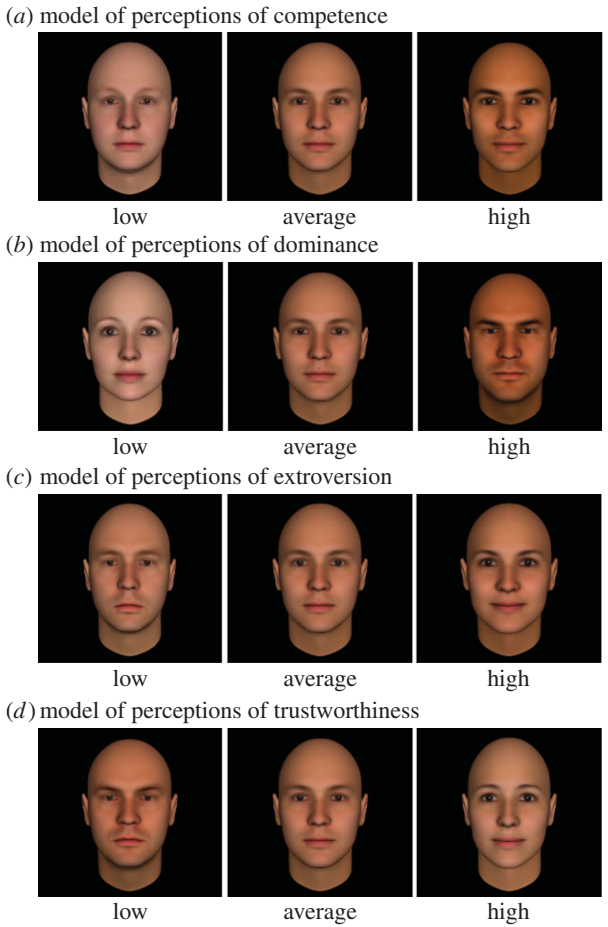
Nous nous formons instantanément des représentations sociales de nos interlocuteurs, notamment grâce aux expressions de leur visage et de leur voix

High-level social inferences from faces



Willis, J., & Todorov, A., *Psychological science* (2006)

A model that learnt how humans make social inferences from faces



Adolphs, R. et al., *Phil. Trans. R. Soc. B* (2016)

High-level social inferences from speech

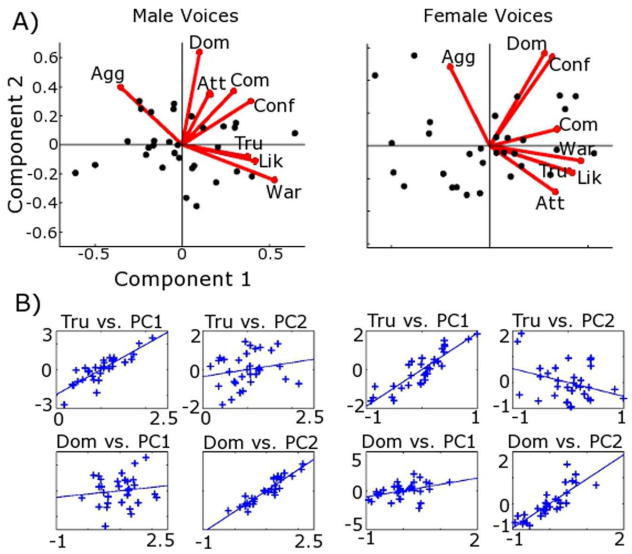
OPEN ACCESS Freely available online

PLOS ONE

How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices

Phil McAleer[1]*, Alexander Todorov[2], Pascal Belin[1,3,4,5]

1 School of Psychology, College of Science and Engineering, University of Glasgow, Glasgow, United Kingdom, **2** Department of Psychology, Princeton University, Princeton, New Jersey, United States of America, **3** Voice Neurocognition Laboratory, Institute of Neuroscience and Psychology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, **4** Département de Psychologie, Université de Montréal, Montréal, Quebec, Canada, **5** Institut des Neurosciences de La Timone, Université Aix-Marseille, Marseille, France



→ A robust code to infer first-impressions from the acoustical features of speech

McAleer, P., Todorov, A., & Belin, P. *PloS one* (2014)

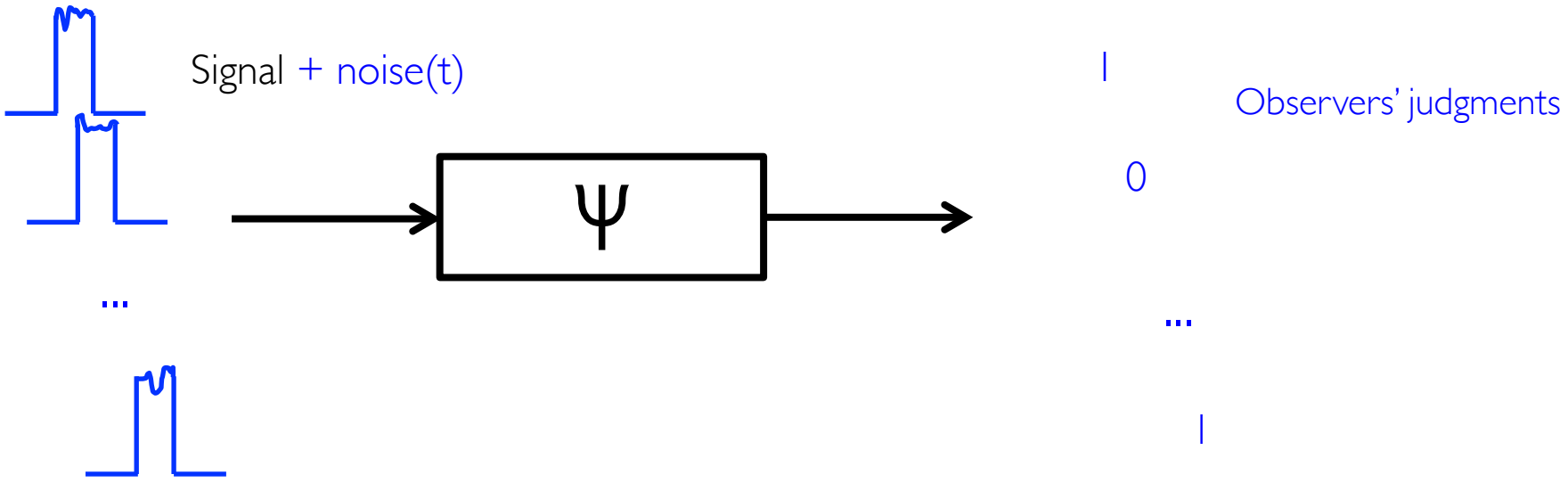
Probing the processing of complex auditory signals in high-level judgments

- The prosodic information (pitch, loudness, timing, …) conveyed by speech signals is crucial for social interaction
  - Pitch is the dimension that conveys most of the information about a speaker's traits, social and emotional states.
    - Social inferences mainly rely on pitch *dynamics* (i.e. intonation)

→How are social judgments inferred from dynamic pitch patterns?

Reverse-correlation constitutes a powerful approach to examine such question

Psychophysical reverse-correlation



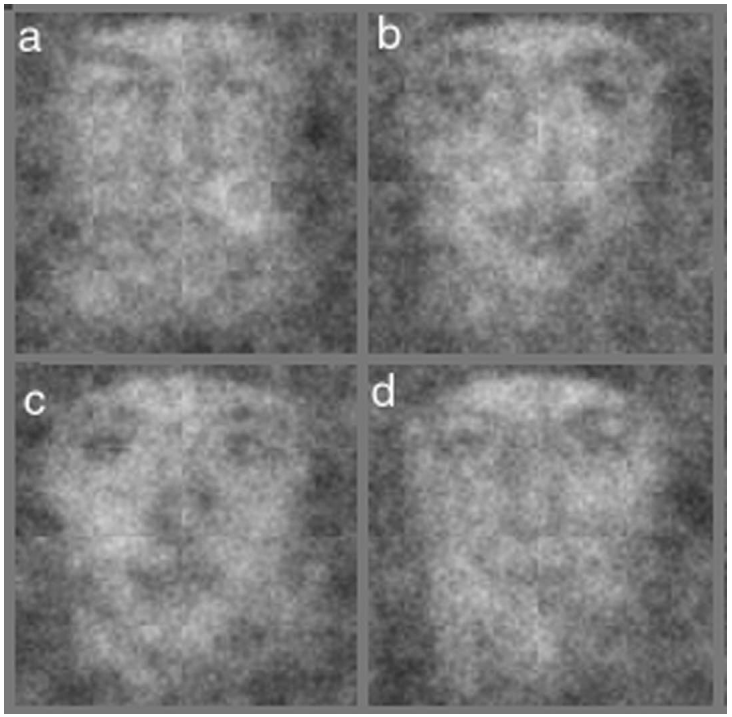
"Reverse-correlation": correlation between noise characteristics and responses to infer the functional properties of the system

Reverse-correlation in high-level vision

Internal templates of emotions on faces

Signal

Signal + Noise



Observers

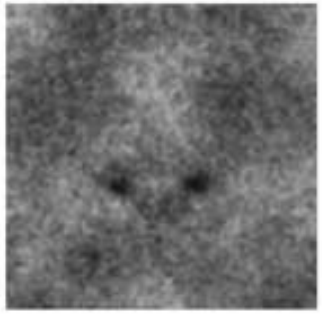
- Happy / Unhappy ?
- Female / Male ?

Mangini, M. C., & Biederman, I., *Cognitive Science* (2004)

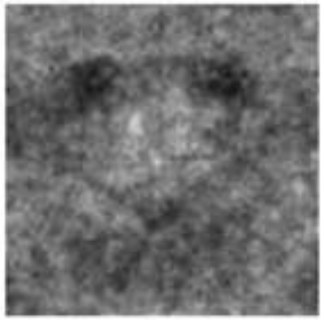
Signal

$(\Sigma\ h)/n_h - (\Sigma\ u)/n_u$

$(\Sigma\ f)/n_f - (\Sigma\ m)/n_m$



+ / - Signal

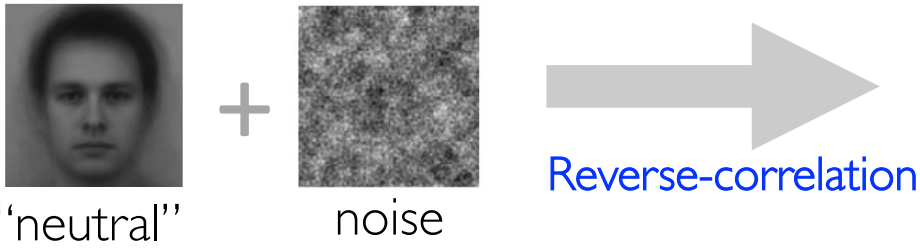


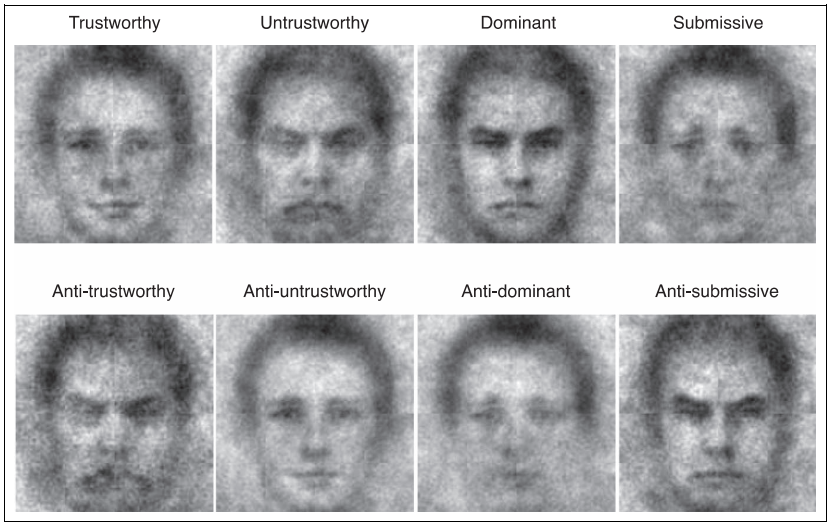
+ / - Signal

Happy / Unhappy

Female / Male

Mental representations of social faces (in the US)

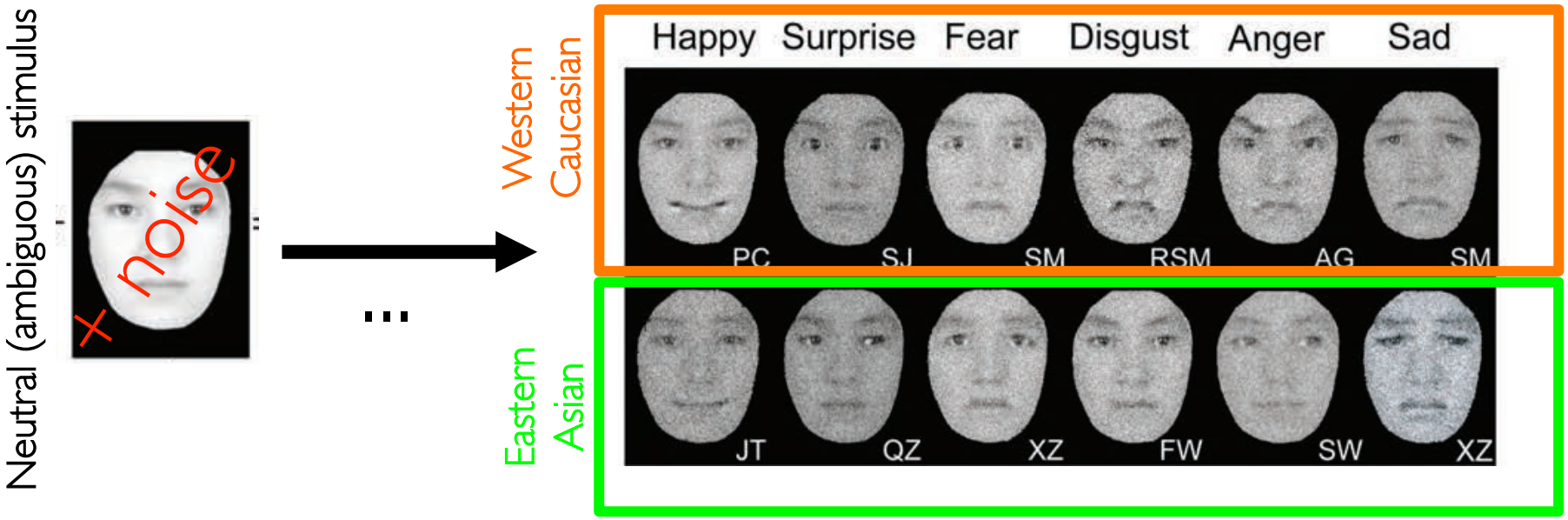




Reverse correlating social faces reveals internal templates

(Dotsh & Todorov, *Social Psychological and Personality Science* 2012)

Cultural differences of emotions



Jack et al., *JEP General*, 2012

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. Current Biology, 19(18), 1543-1548.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. Journal of Experimental Psychology: General, 141(1), 19.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. Proceedings of the National Academy of Sciences, 109(19), 7241-7244.

Jack, R. E., Garrod, O. G., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Current biology, 24(2), 187-192.

...

Reverse-correlation in audition

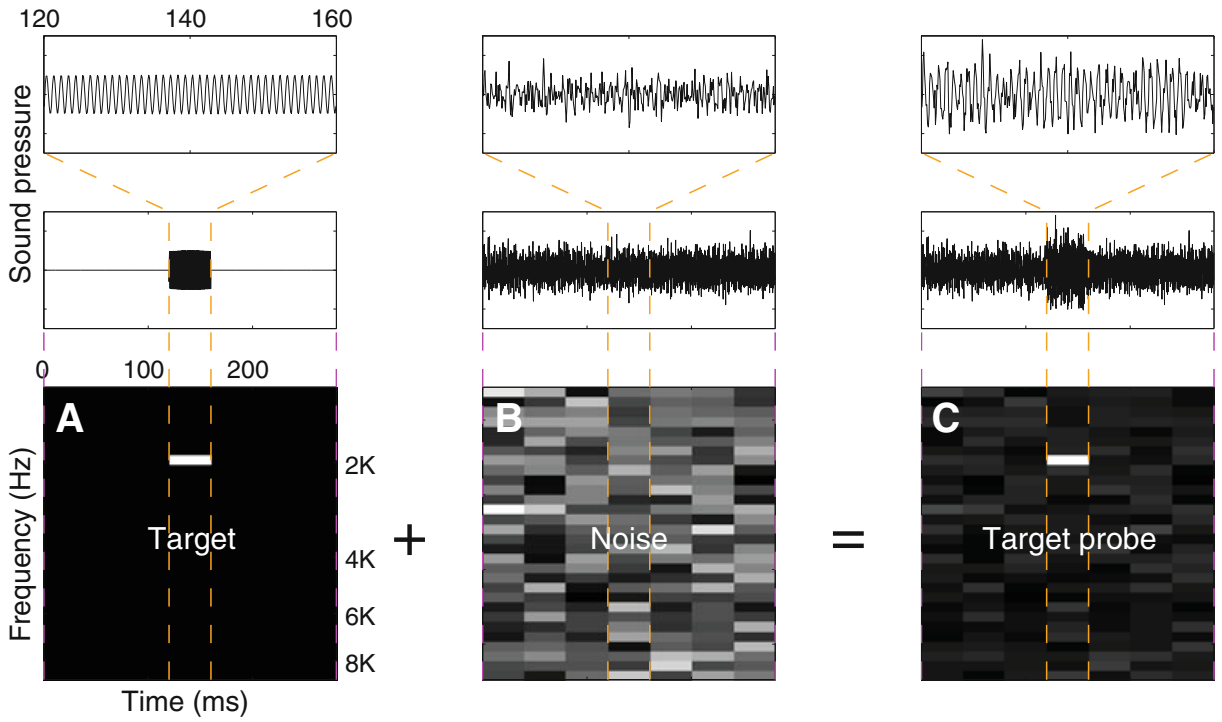
Historically: with low-level stimuli

- Both in auditory and visual domains to explore low-level mechanisms using (very) basic stimuli (since the seminal work of Ahumada et al. in the 70's)

- Various examples can be shown, ranging from linear to second-order nonlinear analyses

- Models can be used to simulate human behavior, serving as a basis for comparing groups' / observers' processing differences

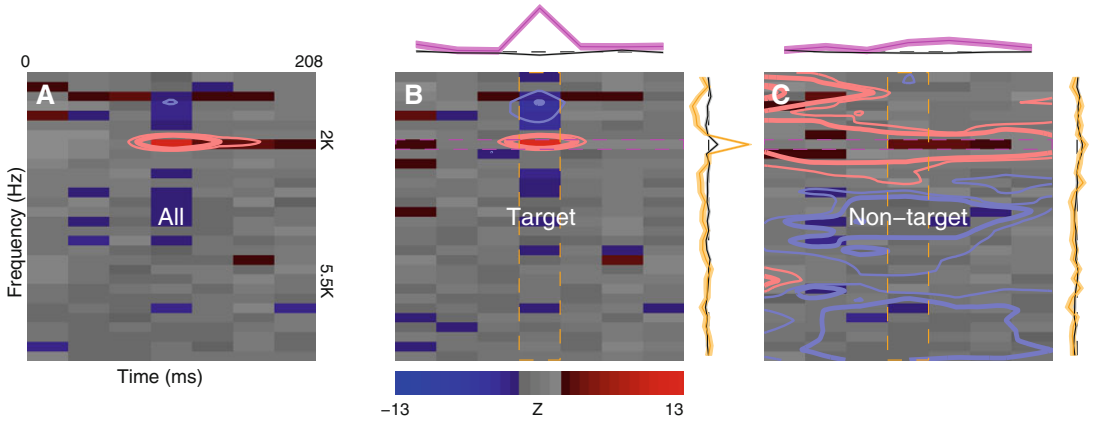
Pure Tone Detection Task



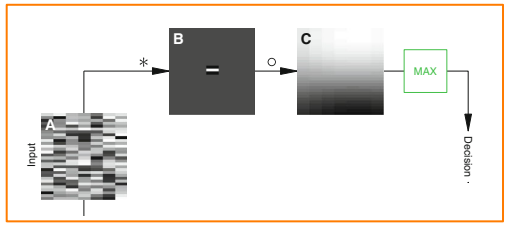
Joosten, E.R. M., & Neri, P., *Biological Cybernetics* (2012)

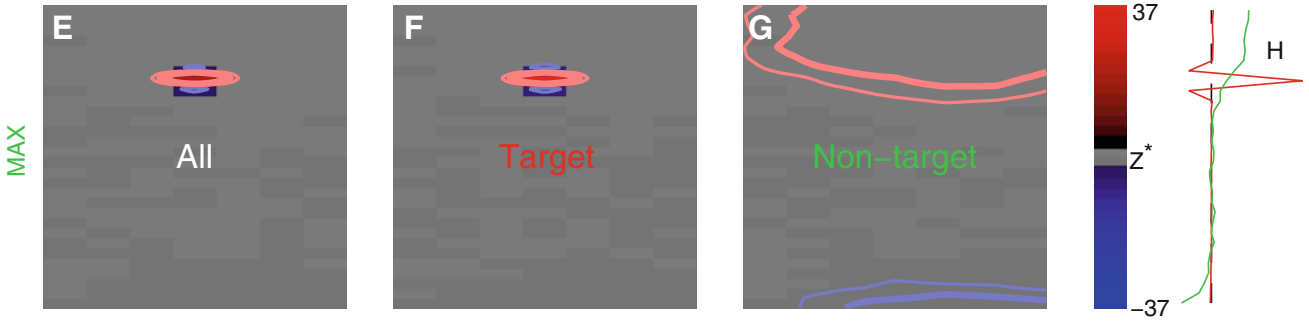
Pure Tone Detection Task

Human observer



MAX uncertainty model





Joosten, E.R. M., & Neri, P., *Biological Cybernetics* (2012)

Toward high-level audio stimuli?

- Only a few very recent studies in the auditory domain

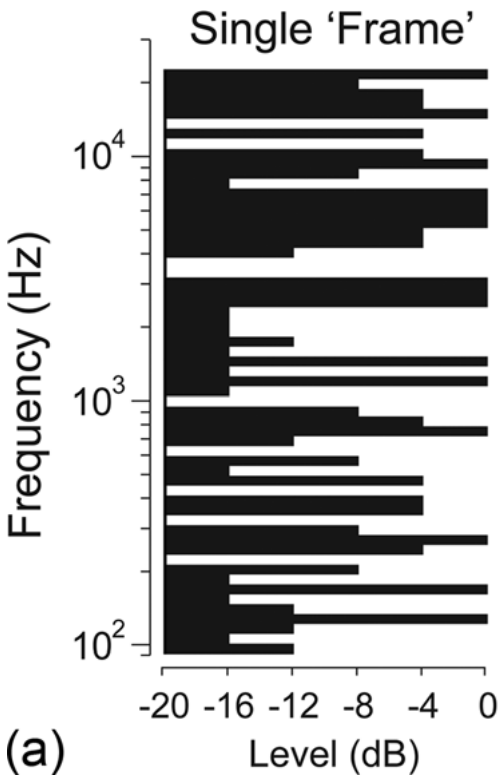
Workshop

Reverse-correlation for high-level auditory cognition

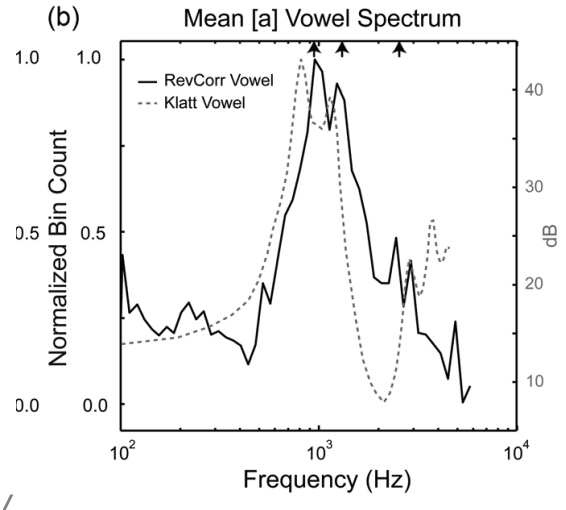
March 22nd & 23rd, 2017
Ircam (Paris) – STMS Lab

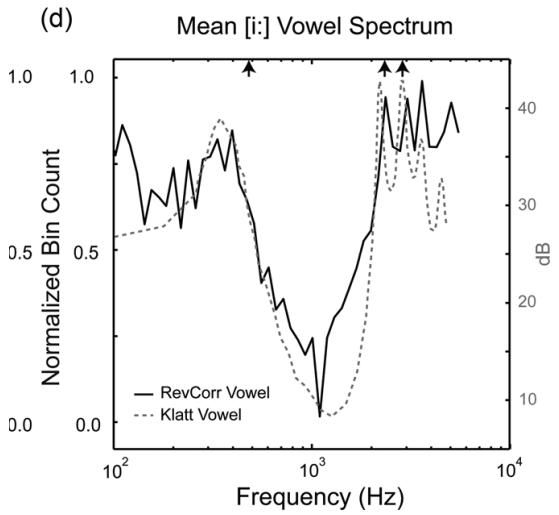
http://cream.ircam.fr/

Vowel Mental Representations



[a] / [i:] ? Press a key



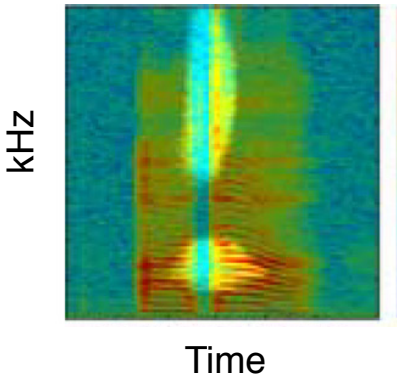


Brimijoin, W. O. et al., *JASA*, 2013

"Bubbles" for speech intelligibility

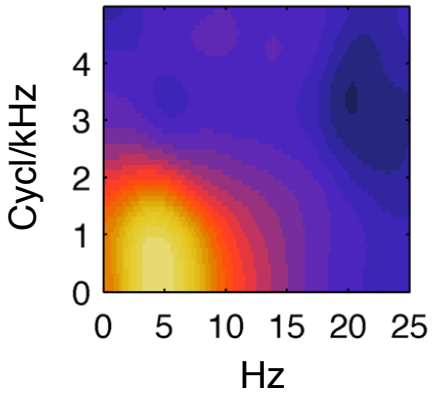
- Speech embedded in noise containing wholes reveals the « minimal window » for intelligibity in the spectro-temporal structure

Mandel, M., et al, *JASA*, 2016



- Speech filtered in its modulation power spectrum (MPS)

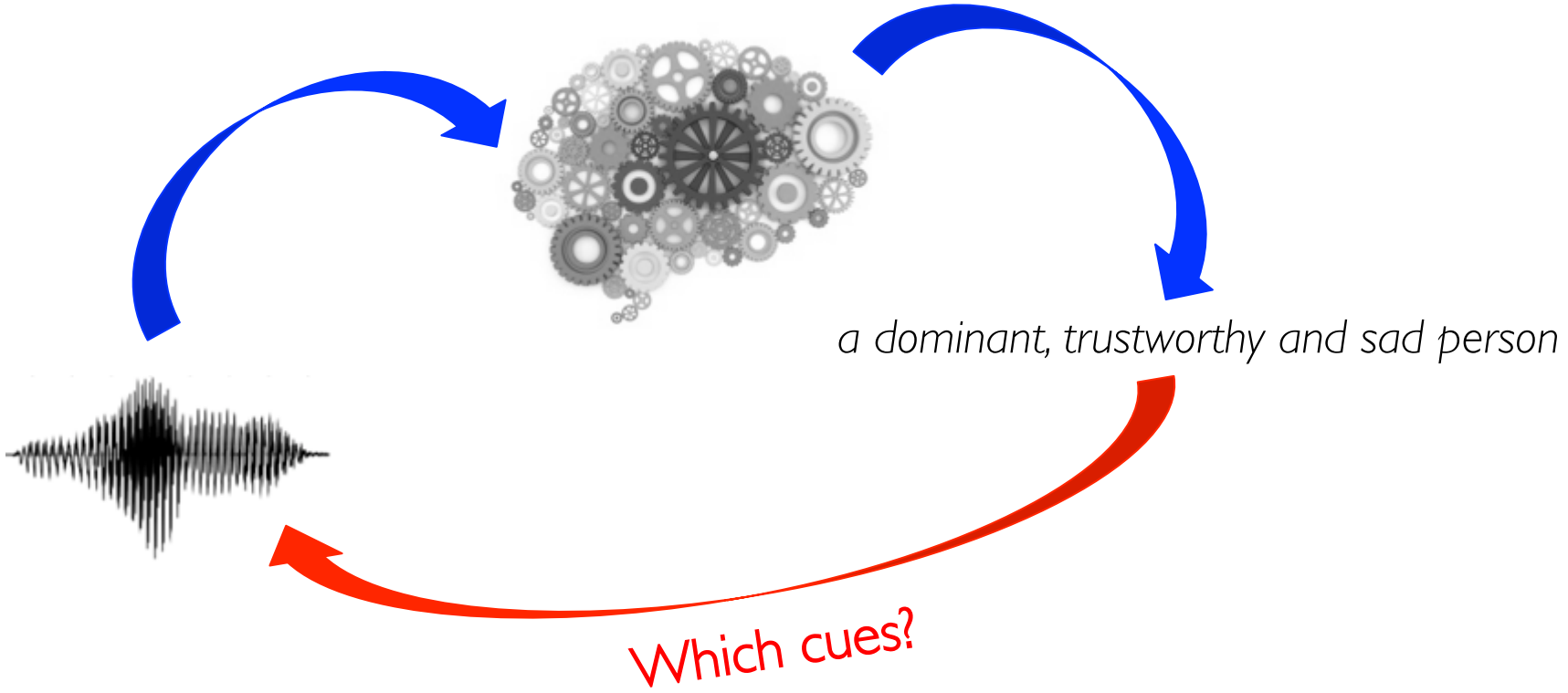
Venezia, J. H., et al., *JASA*., 2016



Using reverse-correlation to uncover social inferences from speech?


BELMONDO
ANCONINA
ITINÉRAIRE
D'UN ENFANT GÂTÉ
filmé par LELOUCH
ITINÉRAIRE D'UN ENFANT GÂTÉ ÉCRIT ET FILMÉ PAR LELOUCH
LIO • MARIE-SOPHIE L. • BEATRICE AGENIN
MICHEL BEAUNÉ • PIERRE VERNIER • JEAN-PHILIPPE CHATRIER et DANIEL GELIN
FRANCIS LAI

Uncovering social inferences from speech?

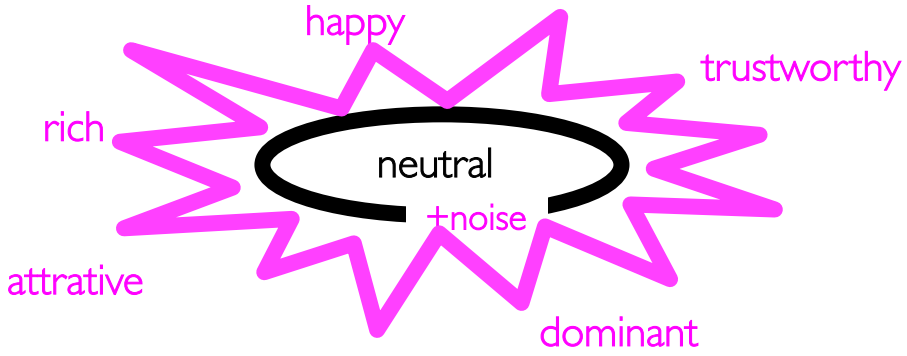


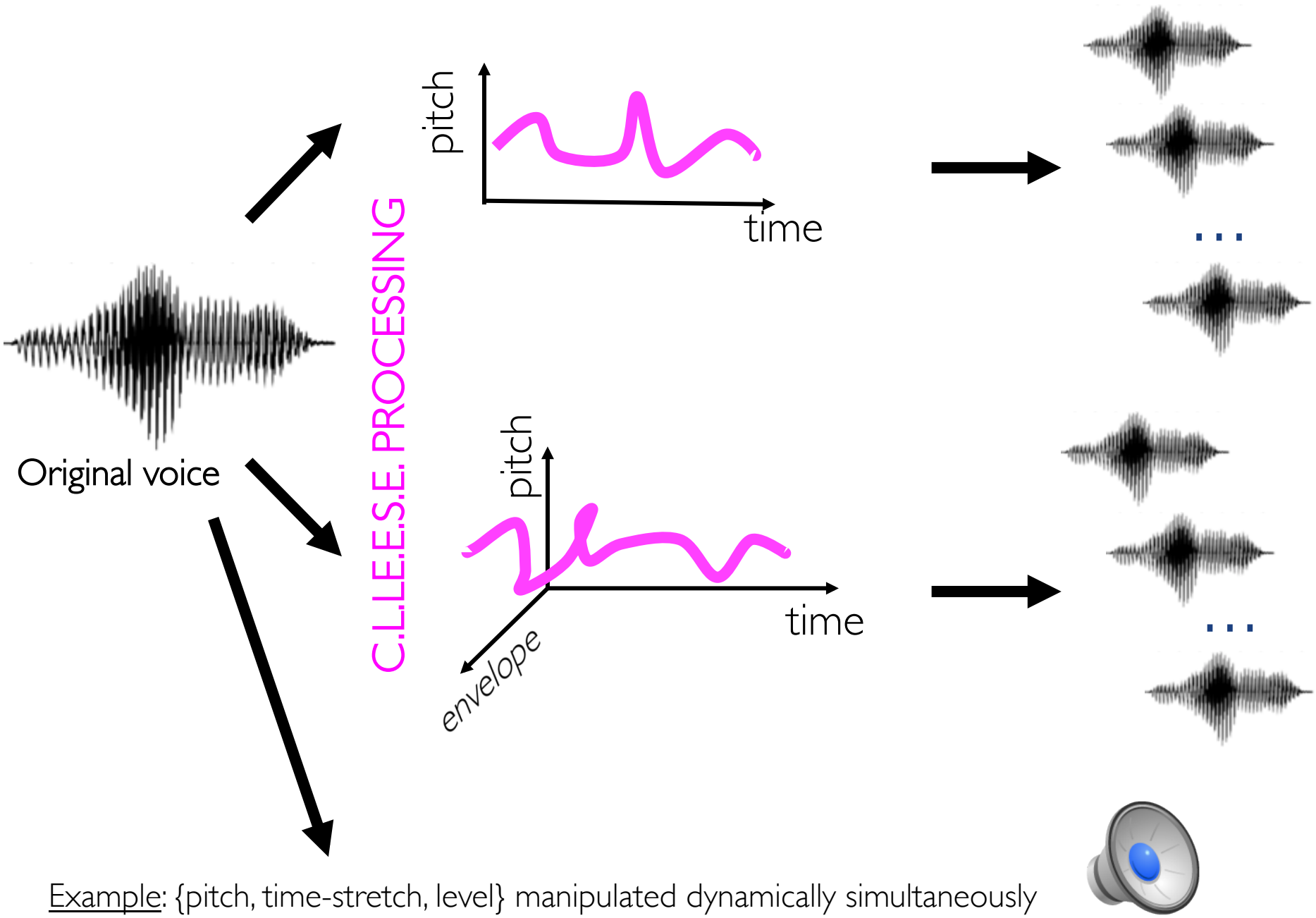
- It is particularly challenging to access mental representations, as one needs to design experimental paradigms and stimuli that cover the whole range of representations human observers might be exposed to.

C.L.E.E.S.E.

Combinatorial Expressive Speech Engine

- A Matlab toolbox (open-access: cream.ircam.fr) that allows dynamic transformations of human voices on 5 dimensions
- The main perceptual space is manipulated directly; real-time dynamic, parametric, fluctuations in pitch, loudness, timbre, speed, and spectral envelope (i.e. *prosody*).
- It allows us to generate an infinite number of natural-sounding, expressive variations around any speech recording





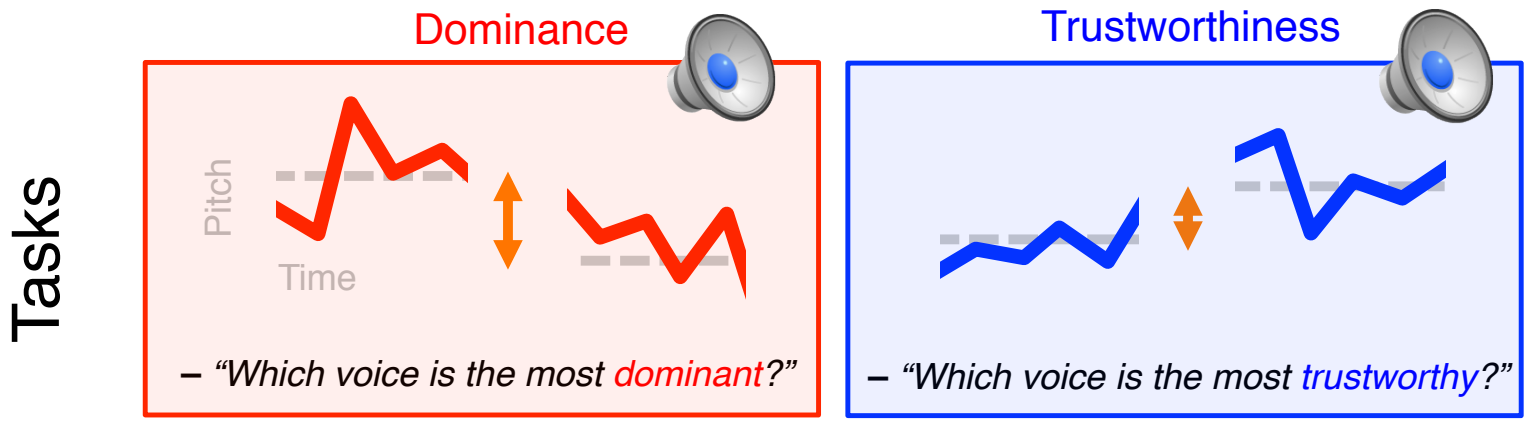
Example: {pitch, time-stretch, level} manipulated dynamically simultaneously

How pitch *dynamically* drives social judgments in speech

Research Questions & Experiments

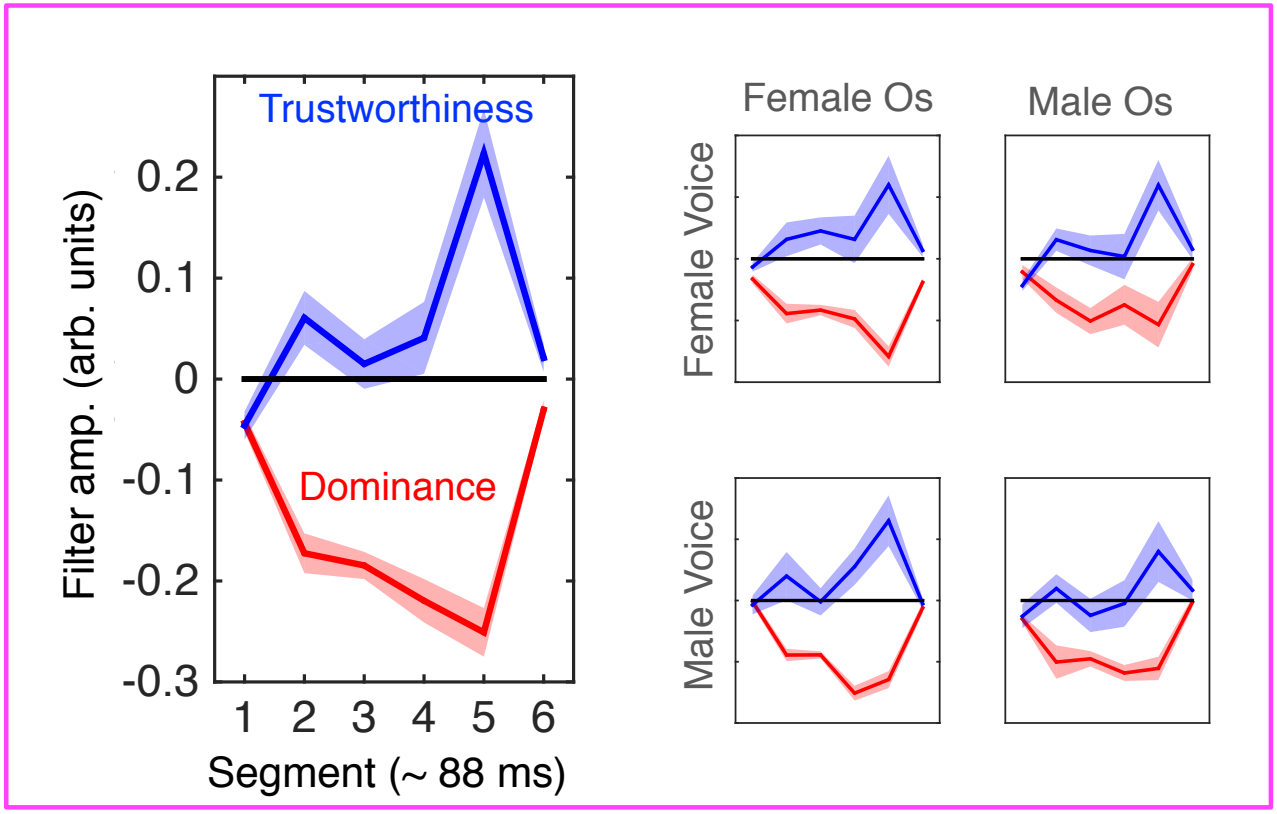
- What is the internal *pitch contour* of a stereotypical dominant / trustworthy voice?
- Are male and female temporal dynamics of processing similar?

→ *Psychophysical experiments to study social first-impressions on the word 'bonjour' (hello) using both a male and a female voice*

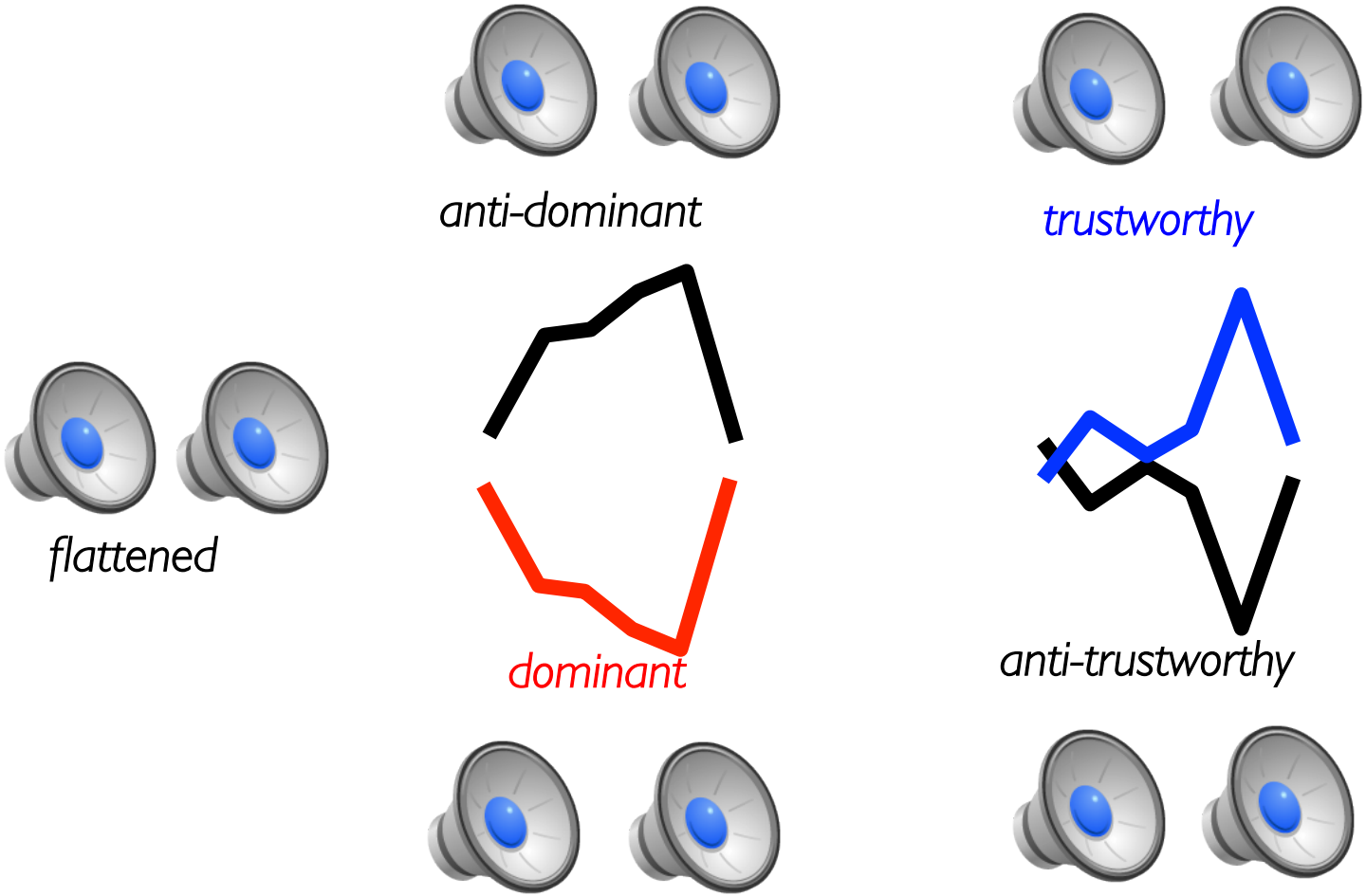


Pitch contour prototypes in judgments of social dominance and trustworthiness

Reverse-correlation



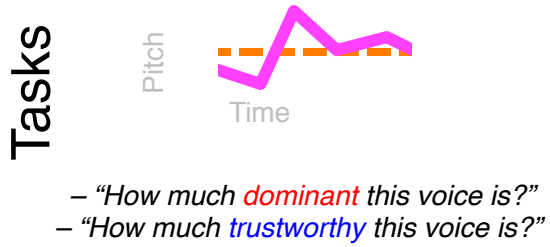
Mental pitch prototypes of dominance/trustworthiness



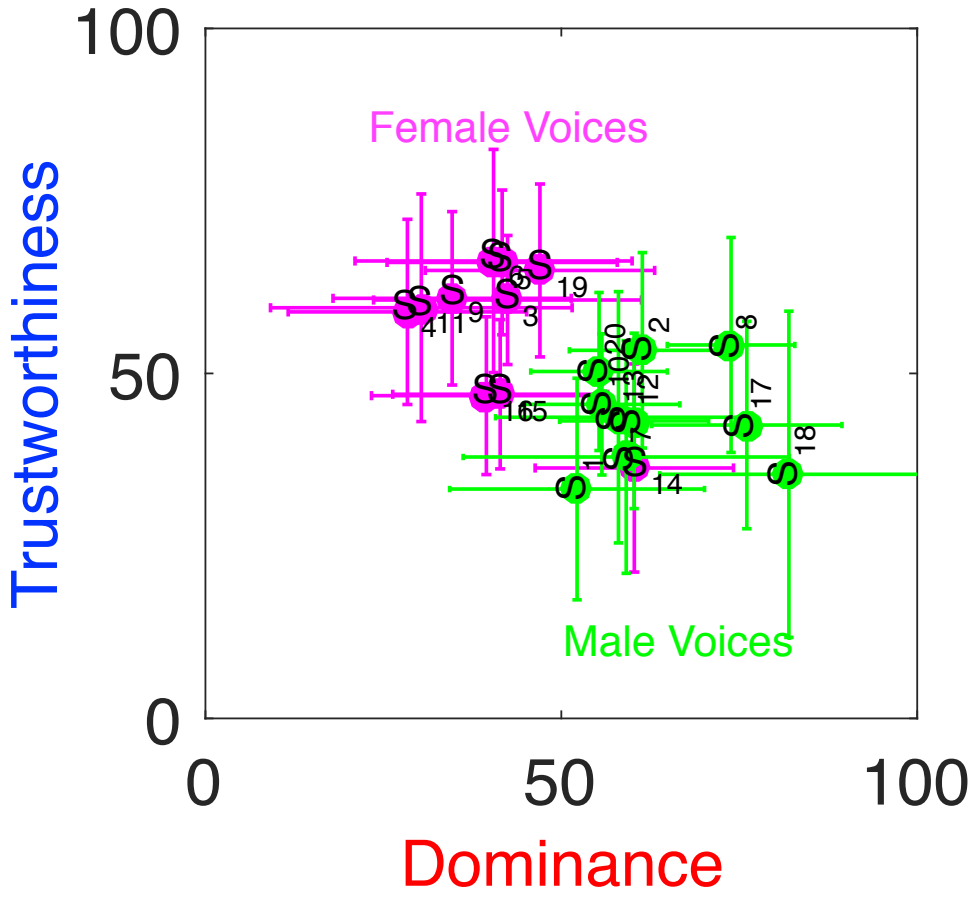
Generalization of these prototypes?

- Stimuli: 20 two-syllable utterances (10 'bonjour' and 10 novel words; from different male and female speakers)
- Thousands of different intonations of these words were randomly presented to novel observers, including pitch contour modifications using the (anti-)prototypes from Exp. 1

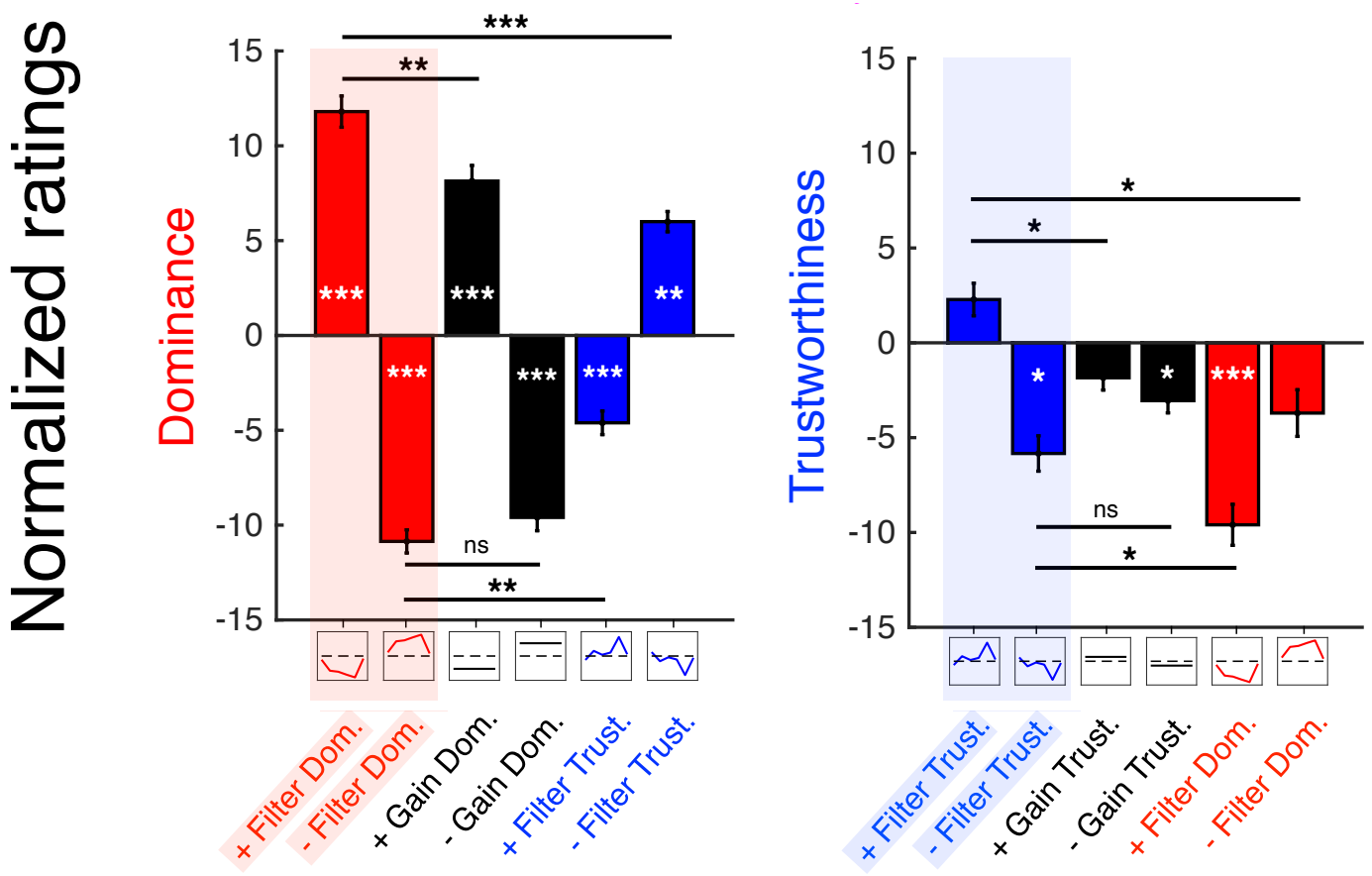
→ *Straightforward evaluation task on a Likert scale*



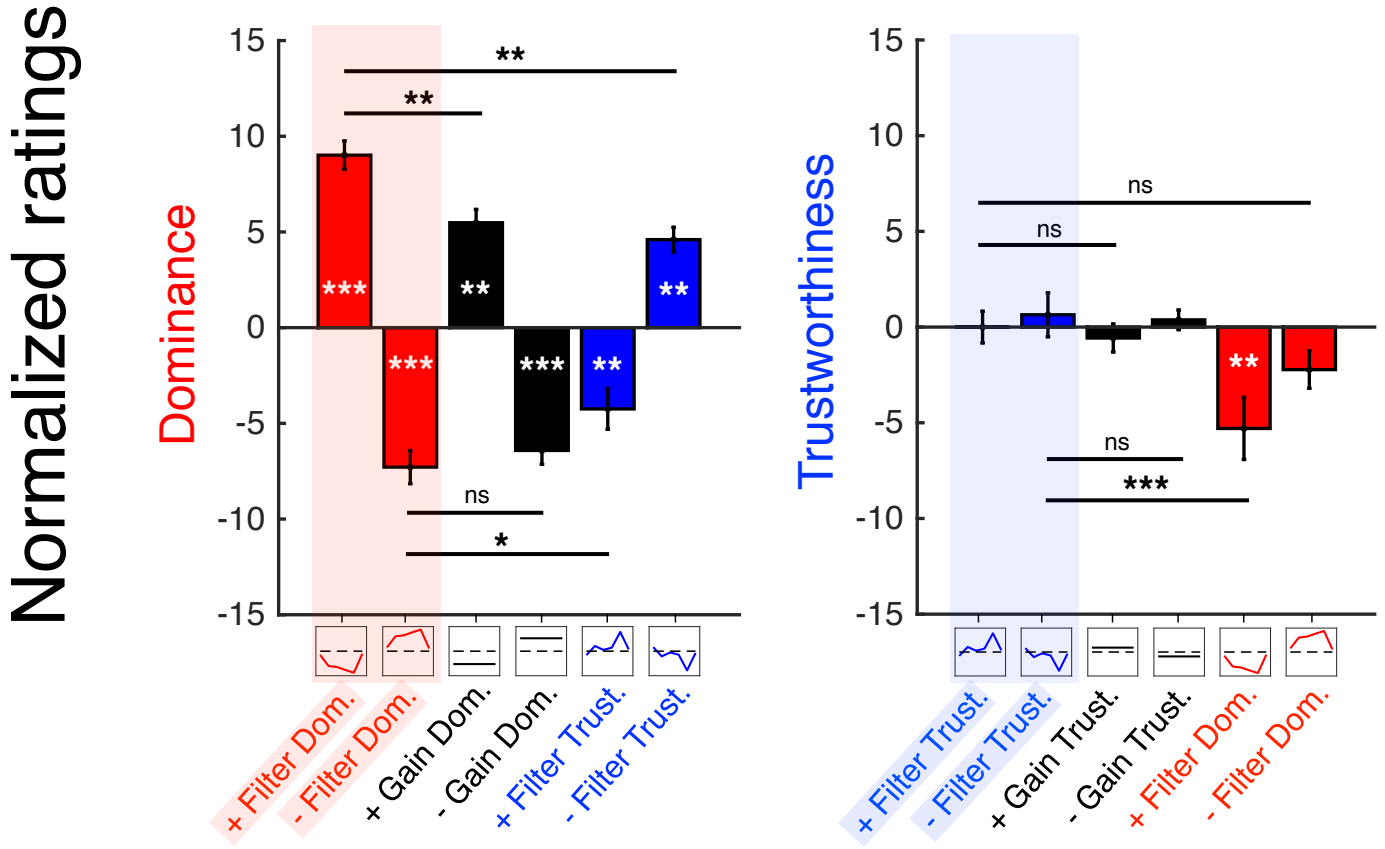
Pooled ratings



Effects of pitch contour changes on "bonjour"



Effects of pitch contour changes on novel 2-syllable utterances



Discussion

- We show *how* pitch contour dynamically drives dominance and trustworthiness in speech

- Strikingly similar prototypes across both speaker and listener gender suggests that humans have developed a common cross-gender *dynamic* code to go beyond the dimorphic characteristic of the voice

Potential applications

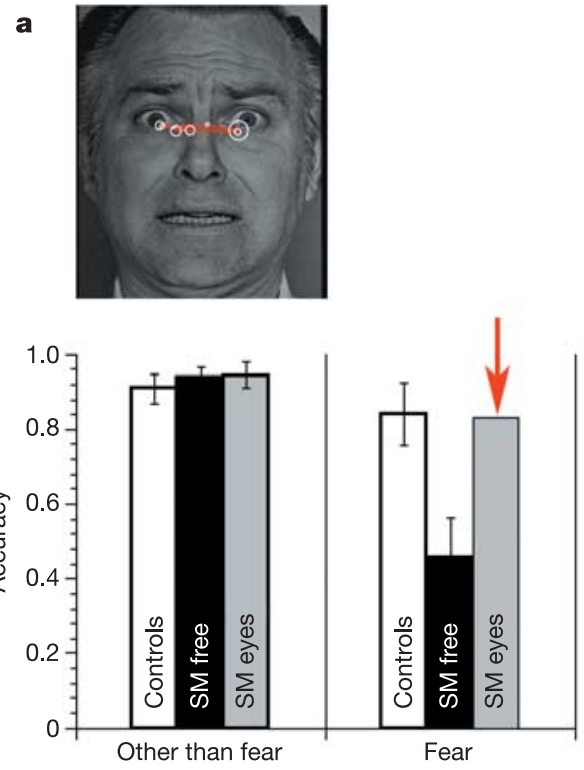
- A real-time vocal "social make-up" that could be the core of next audio algorithms in social signal processing

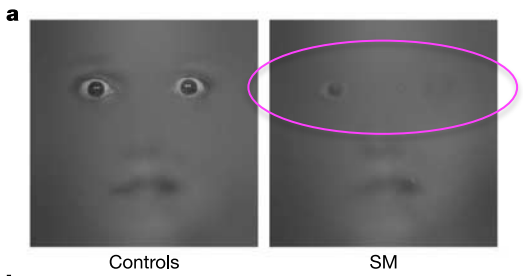
- Provide mechanistic accounts for people with auditory processing deficits → a step toward more targeted rehabilitation strategies.

    - Socially-relevant signal-processing strategies for cochlear-implant devices
    - Development of individually-shaped "speech therapies" for individuals suffering from dysprosody, such as depressive people, ASDs, schyzophrenian or in congenital amusia

Rehabilitation of SM, an amygdala-damaged patient

Eye-tracking measures



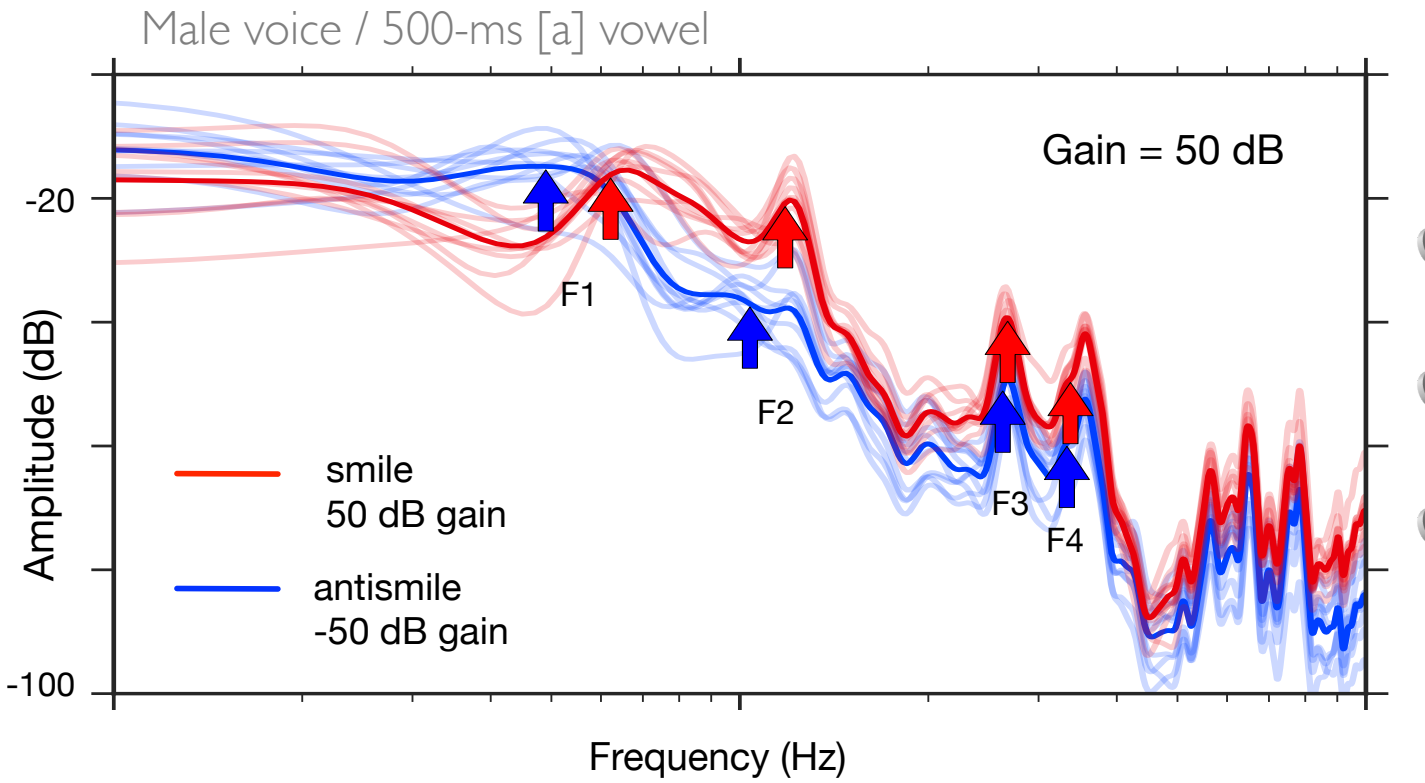


Bubbles measures

Adolphs, R., et al. *Nature*, 2005

Mental representations of smile in speech

colab w/ Pablo Arias (IRCAM/CNRS)



anti-smile

neutral

smile

int. noise ~ 1.7 ext. noise → similar to basic sensory tasks!

(Neri, *Psych. Bull. & Rev.*, 2010)

Conclusions

Human auditory processing has evolved to infer meaningful and relevant information from others' voice through robust filters

→ [voice transformation algorithms + reverse-correlation]
= an approach to uncover social auditory filtering

This work was done in collaboration with:

- Jean-Julien Aucouturier (IRCAM/CNRS)
- Pascal Belin (Univ. Aix-Marseille, France)
- Juan-José Burred (Independent Researcher)
- Pablo Arias (IRCAM/CNRS)

Thank you for your attention