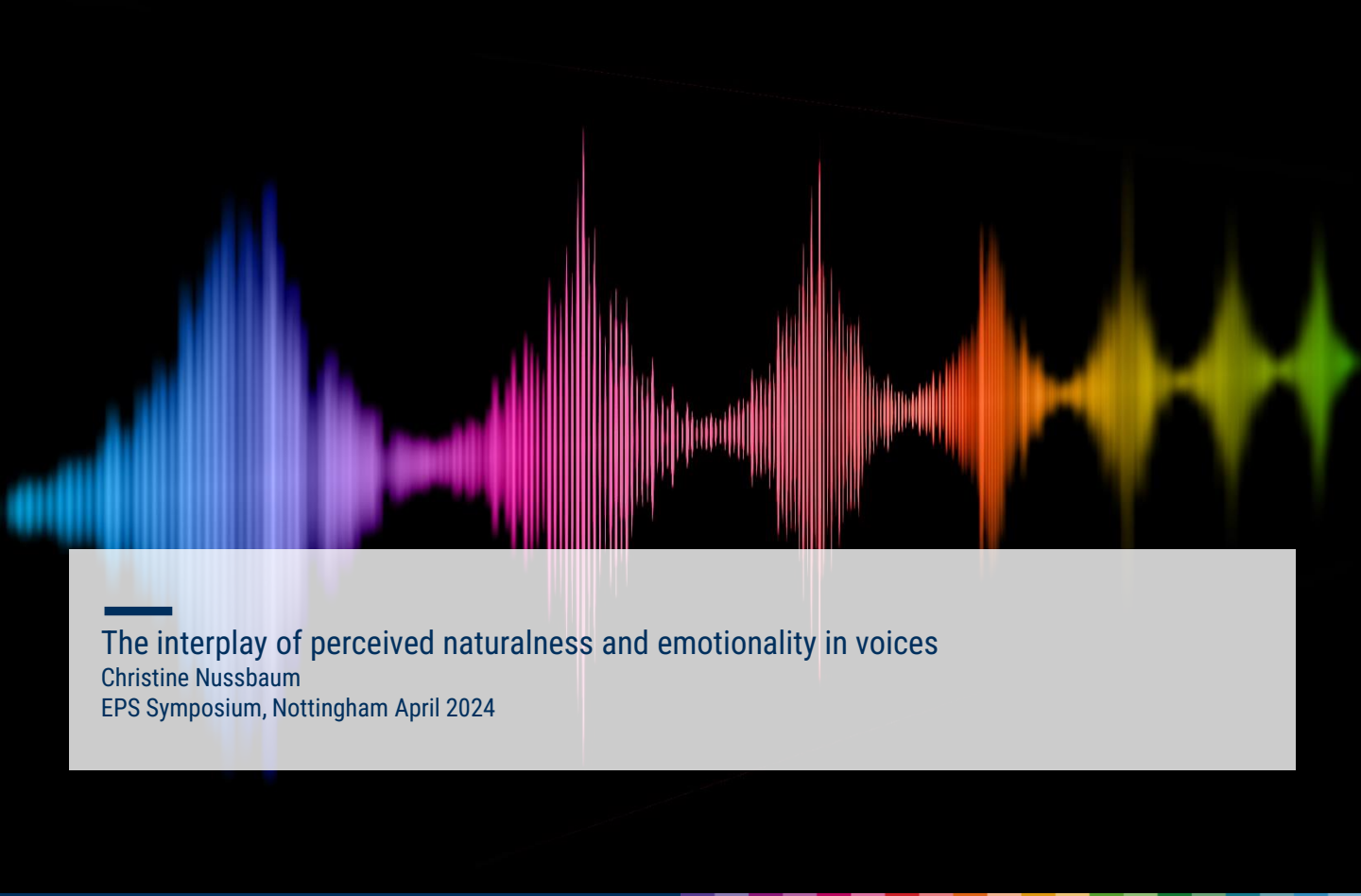


The interplay of perceived naturalness and emotionality in voices

Christine Nussbaum

EPS Symposium, Nottingham April 2024

Person Perception from Voices (PPV)

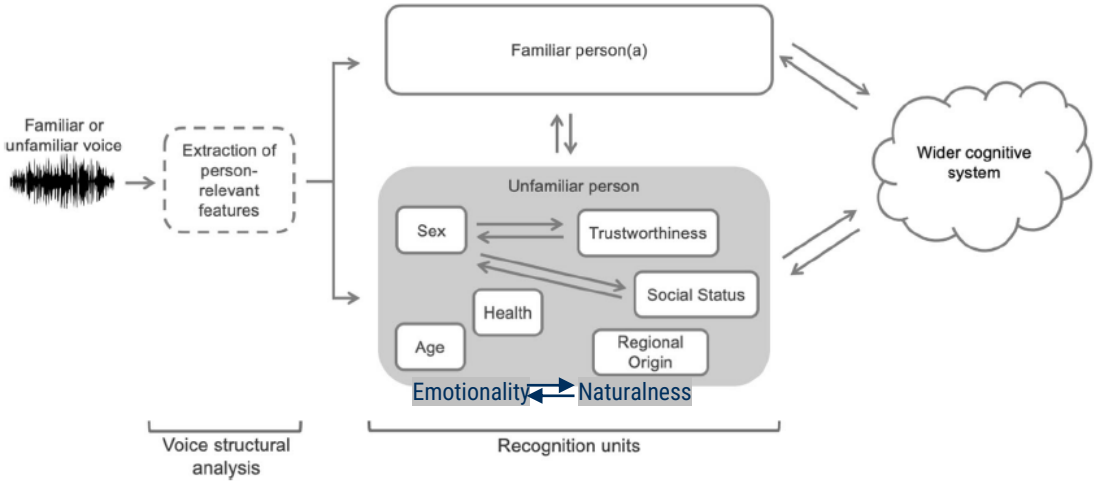


Fig. 3 A hierarchical model of person perception from voices (PPV). The model illustrates how different person characteristics may be recognised and may interact and inform each other during person perception from voices. Boxes with a solid outline describe a recognised percept and the box with a dashed outline describes a computational process. Brackets at the bottom of the figure loosely map the processing stages of the PPV to Belin et al.'s hierarchical model of voice identity perception[10]. Not all possible interactions between percepts are shown.

[Lavan & McGettigan 2023]

Vocal Emotions

- Speakers express vocal emotions through acoustic cues
  - **Fundamental frequency (F0)**
  - **Timbre**
  - Amplitude
  - Speech rate
- Listeners recognize vocal emotions above chance

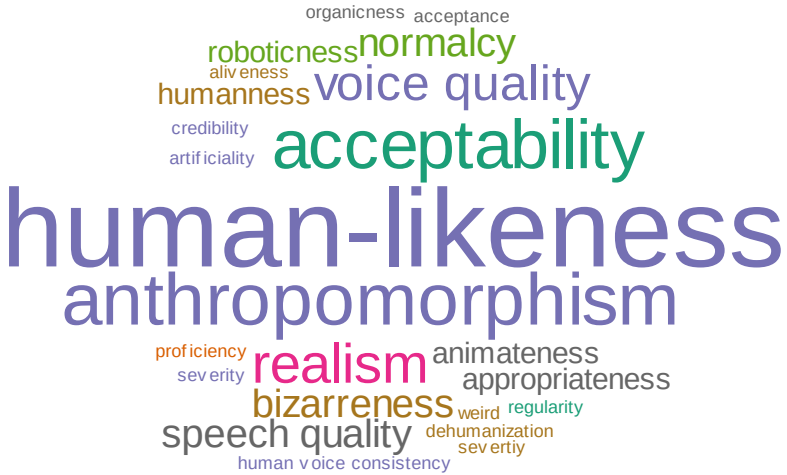
➢ multiple efforts to link emotions to acoustic cues

➢ profound interindividual differences

[Scherer 2018; Juslin & Laukka 2003]

Voice Naturalness

- pathological human voices
- manipulated human voices
- synthesized/artificial voices



"Naturalness was defined as conforming to the listener's standards of rate, rhythm, intonation, and stress patterning […]" [e.g. Yorkston et al. 1990]

"Natural speech is the speech most closely perceived as a human voice" [e.g. Mawalim et al. 2022]

"By naturalness, we understand the voice stimulus to be perceived as a ***plausible outcome of the human speech production system****."* [Nussbaum et al. 2023]

Voice Naturalness

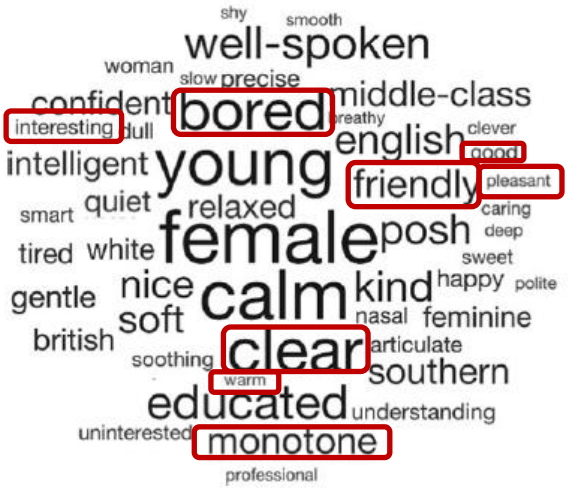


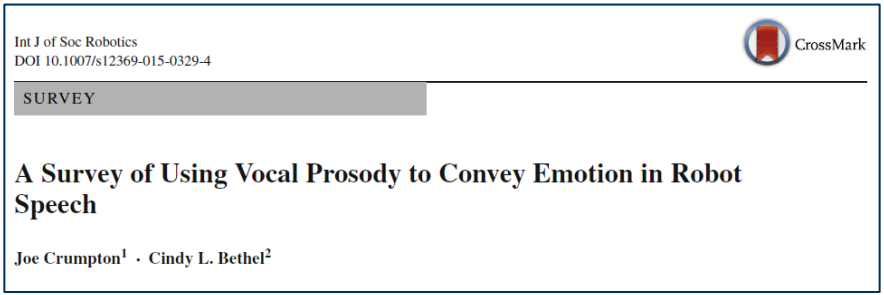
Fig. 2 Word cloud illustrating the multivariate nature of person perception from unfamiliar voices. [Lavan & McGettigan, 2023; Lavan 2023]

[e.g. Kühne et al. 2020, Rodero 2017, Rodero & Lucas 2023]

Voice Naturalness and Quality of Life

"Impairments in speech naturalness can lead to communication partners perceiving the affected individuals as unhappy, cold, withdrawn, introverted, or bored. These false perceptions can interrupt participation in regular life roles, leading to loss of employment and independence. Thus, impaired speech naturalness can result in social isolation, reduced quality of life, and depression." [Stepp & Voijtech 2019]

The interplay of perceived naturalness and emotionality in voices



Int J of Soc Robotics
DOI 10.1007/s12369-015-0329-4

CrossMark

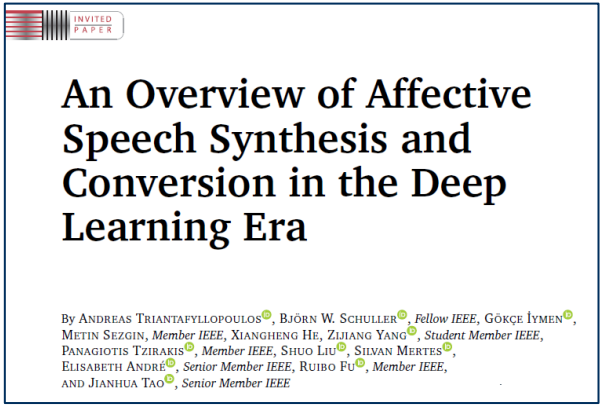
SURVEY

A Survey of Using Vocal Prosody to Convey Emotion in Robot Speech

Joe Crumpton[1] · Cindy L. Bethel[2]

2015

„The communication of emotions through changes in vocal prosody is one way to make synthesized speech sound more natural."



INVITED PAPER

An Overview of Affective Speech Synthesis and Conversion in the Deep Learning Era

By Andreas Triantafyllopoulos, Björn W. Schuller, *Fellow IEEE*, Gökçe İymen, Metin Sezgin, *Member IEEE*, Xiangheng He, Zijiang Yang, *Student Member IEEE*, Panagiotis Tzirakis, *Member IEEE*, Shuo Liu, Silvan Mertes, Elisabeth André, *Senior Member IEEE*, Ruibo Fu, *Member IEEE*, and Jianhua Tao, *Senior Member IEEE*

2023

The Interplay of Perceived Naturalness and Emotionality in Voices

Emotion perception is robust against unnatural voice features

- Emotional caricatures are better recognized despite being less natural than unaltered utterances [Ko et al. 2023]
- Emotions expressed by participants with autism were better recognized, while rated as being less natural [Hubbard et al. 2017]
- Rating of naturalness unrelated to ratings of emotionality
  [Nussbaum et al. 2023]

Naturalness affects/disrupts emotion perception

- Emotion recognition worse in text-to-speech generated voices
  [Ko et al. 2023]
- Human voices rated as more expressive compared to synthesized ones
  [Cabral et al. 2017]
- Reduction of naturalness in voices affects emotion recognition and electrophysiological correlates, but NOT valence and arousal ratings
  [Duville et al. 2022]

➢ Acoustic manipulation that affects perceived naturalness can be a threat to ecological validity!

Unwanted Interference by Voice Naturalness

COGNITION AND EMOTION
https://doi.org/10.1080/02699931.2023.2200920

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Check for updates

Perceived naturalness of emotional voice morphs

Christine Nussbaum [a,b], Manuel Pöhlmann [a], Helene Kreysa [a,b] and Stefan R. Schweinberger [a,b,c]

[a]Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Germany; [b]Voice Research Unit, Friedrich Schiller University, Jena, Germany; [c]Swiss Center for Affective Sciences, University of Geneva, Switzerland

2023

Unwanted Interference by Voice Naturalness

SPECIAL SECTION: EMOTION IN THE VOICE

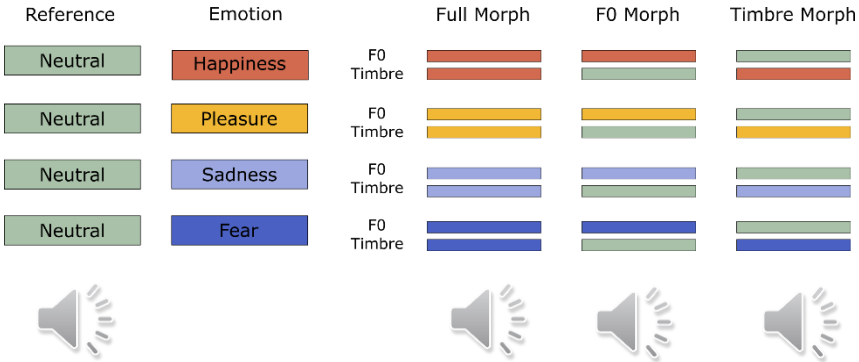
emotion review

Emotion Review
Vol. 13, No. 1 (January 2021) 12–24
© The Author(s) 2020
ISSN 1754-0739
DOI: 10.1177/1754073920934544
https://journals.sagepub.com/home/emr

Beyond Correlation: Acoustic Transformation Methods for the Experimental Study of Emotional Voice and Speech

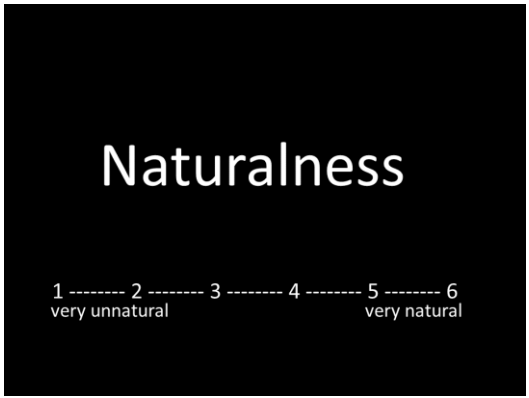
Pablo Arias*, Laura Rachman*, Marco Liuni and Jean-Julien Aucouturier
STMS UMR9912, IRCAM/CNRS/Sorbonne Université, France

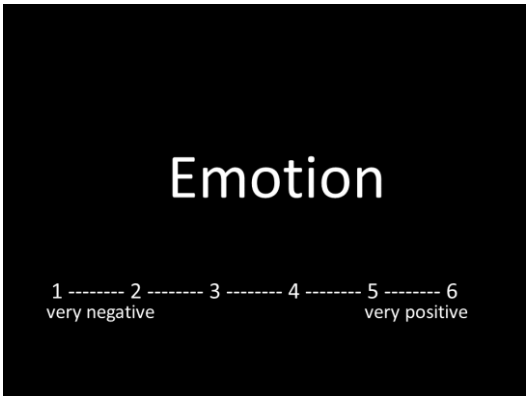
2021



parameter-specific voice morphing

Rating Study:





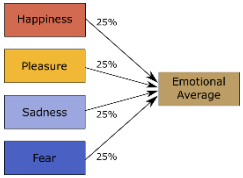
Stimuli:

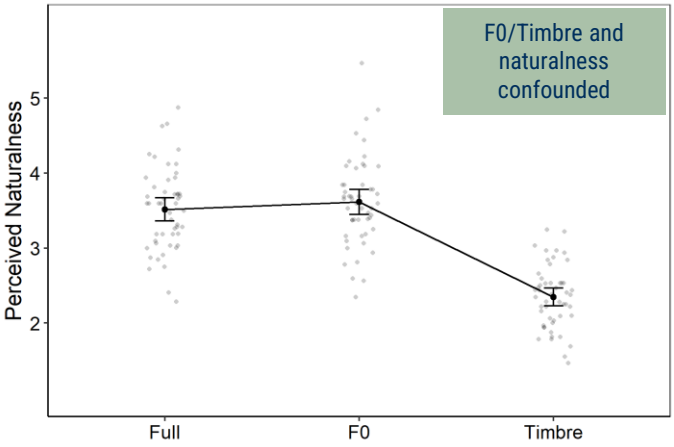
8 (speakers) x 3 (pseudowords) x 4 (emotions) x 3 (morphing conditions) x 2 reference types + 48 reference (8 speakers x 3 pseudowords x 2 reference types) = 624 stimuli; split between-subject per pseudoword

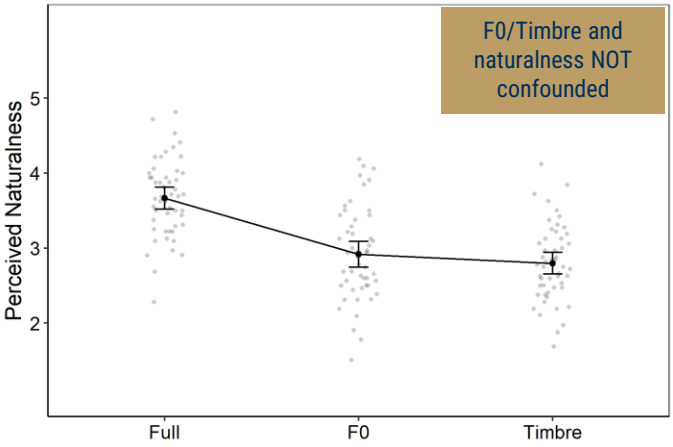
Sample:

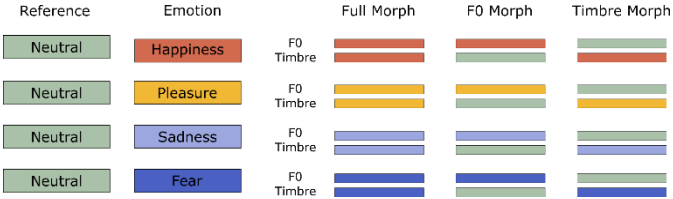
51 participants, 40 females, 11 males,
aged 19 to 31 years [M = 21.49; Mdn = 21; SD = 2.64],
with 16/18/17 per pseudoword

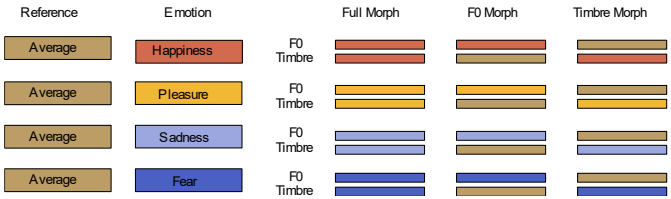
Result 1: Voice Morphing affects Voice Naturalness





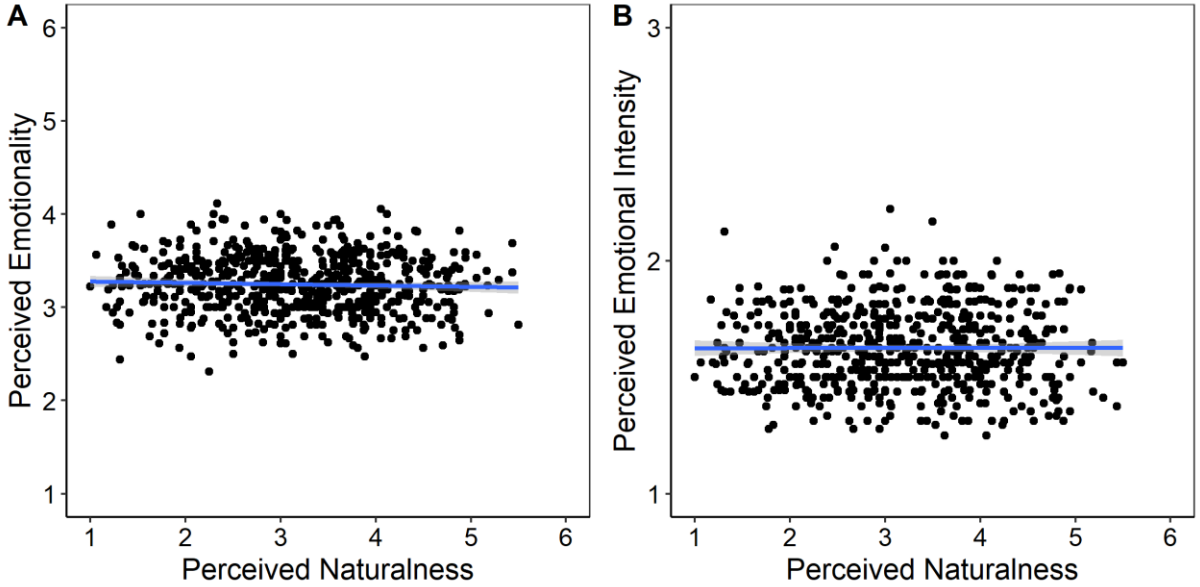






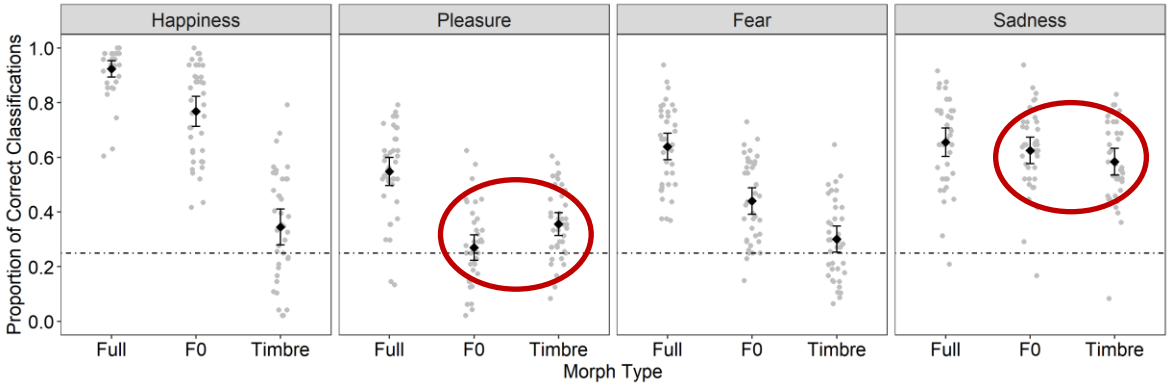
Result 2: No Relationship between Naturalness and Emotionality

Mean ratings of perceived naturalness, emotionality (A), and emotional intensity (B).



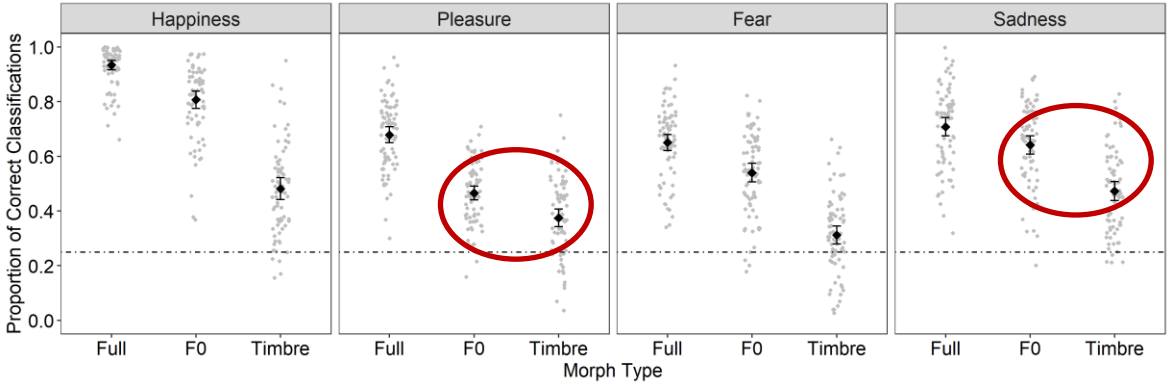
[Nussbaum et al. 2023, Cognition and Emotion]

Naturalness may affect Emotion Recognition Performance



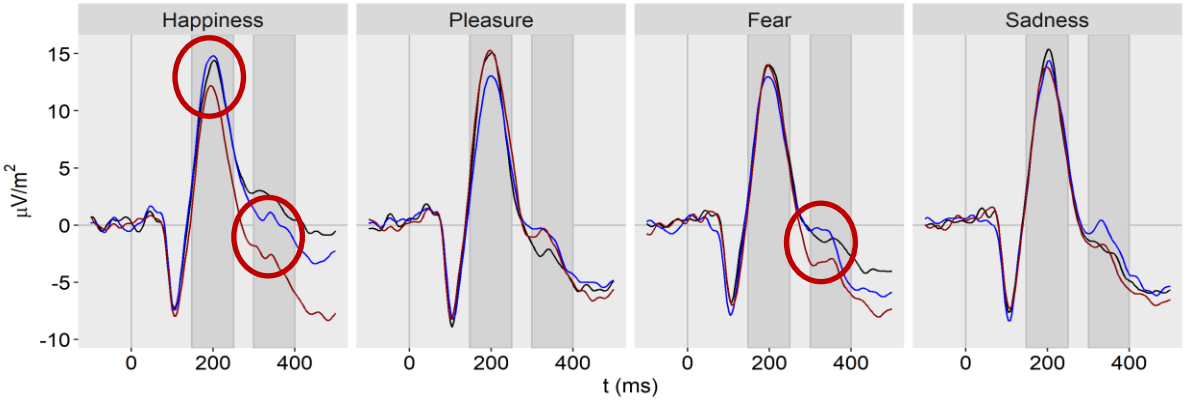
F0/Timbre and naturalness confounded

[Nussbaum et al. 2022, SCAN]



F0/Timbre and naturalness NOT confounded

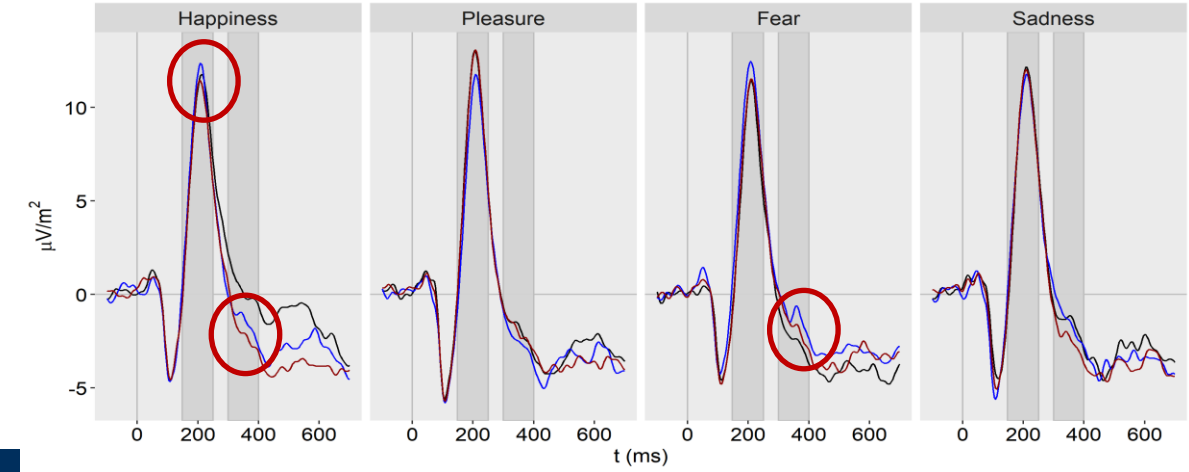
[Nussbaum et al. 2023, BJOP]



F0/Timbre and naturalness confounded

Morph Type
Full
F0
Timbre

[Nussbaum et al. 2022, SCAN]



F0/Timbre and naturalness NOT confounded

Morph Type
Full
F0
Timbre

[Nussbaum et al. 2023, Brain Sciences]

Summary

Listeners form an impression about the naturalness of a voice

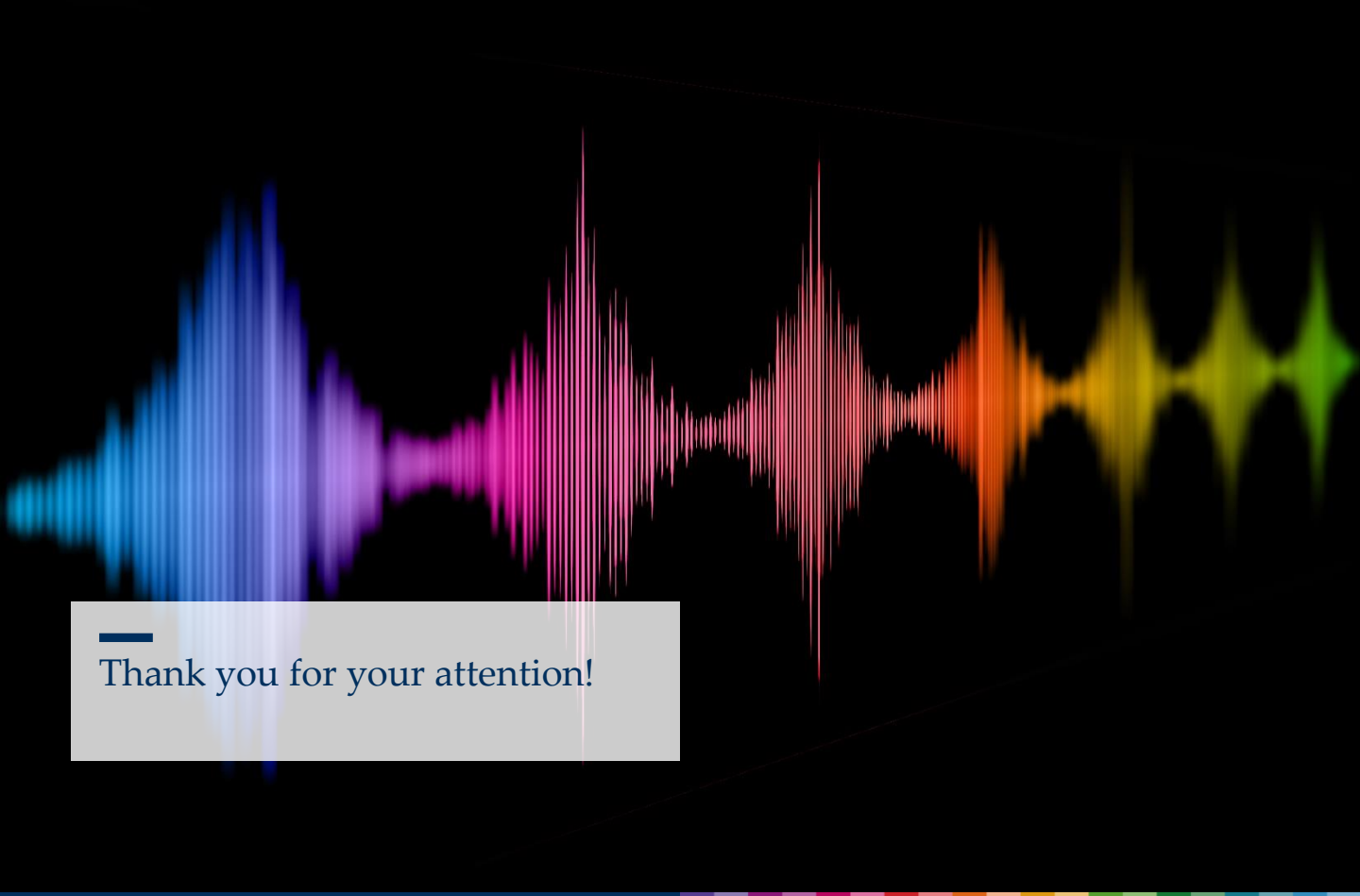
Naturalness interacts with the perception of other person characteristics

There is a complex interplay between perceived naturalness and emotionality

Reduced naturalness can reduce ecological validity of emotional stimuli

Naturalness may not affect ratings of perceived emotionality, but emotion recognition and the associated electrophysiological correlates

… there is a lot of work to do!



Thank you for your attention!