

Covert Digital Manipulation of Vocal Emotion alters Speakers' Emotional States in a Congruent Direction¹

Radovan Dabetić

Computational Social Science with Images and Audio,
ETH Zürich

November 17, 2023

¹Jean-Julien Aucouturier et al., Science and Technology of Music and Sound (STMS UMR9912), CNRS/Institut de Recherche et Coordination en Acoustique et Musique (IRCAM)/Université Pierre et Marie Curie (UPMC), 74005 Paris, France

Outline

- 1 Question
- 2 Methods
 - Audio Manipulation
 - Experiments
- 3 Experiments

Vocal Emotion

Understand regulation of emotions

Vocal Emotion

Understand regulation of emotions
Emotional errors? → Feedback system?

Vocal Emotion

Understand regulation of emotions

Emotional errors? → Feedback system?

More easily accessible: vocal emotion

Idea for Experiment

Change self-perception

Altering voice while reading and measuring response in speech

Pitch-Shifting

Frequency of a sound [Hz]

Pitch-Shifting

Frequency of a sound [Hz]

Measure of relative change: cent

1 cent $\hat{=}$ frequency ratio of $2^{1/1200} \hat{=}$ 100th of a semitone

Inflection

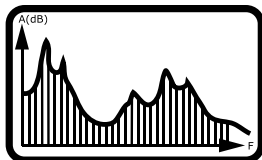
Inflection

Non-standard, here:

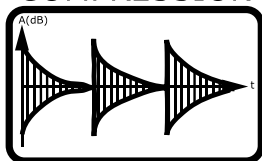
“ [...] process of rapidly modifying the pitch at the start of each utterance. [...] ”

Shelf-Filtering and Compression

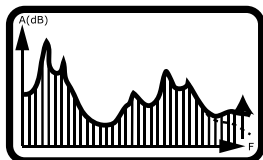
HIGH-SHELF FILTER



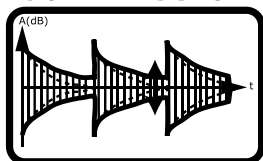
COMPRESSION



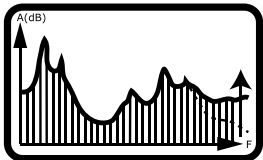
HIGH-SHELF FILTER



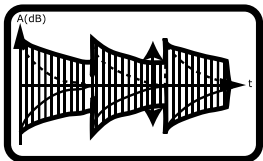
COMPRESSION



HIGH-SHELF FILTER

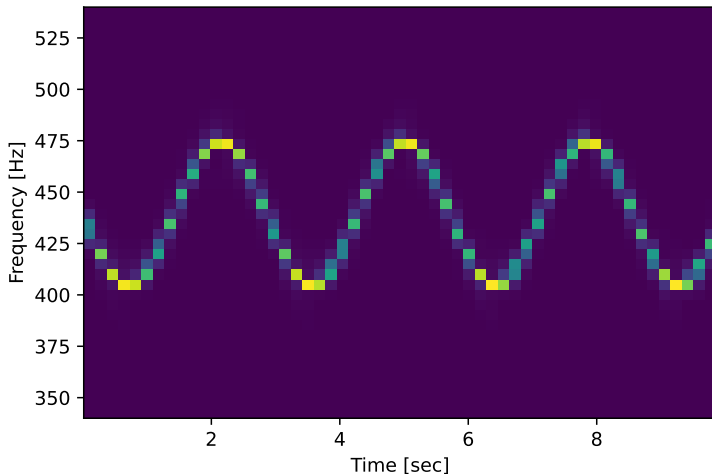


COMPRESSION

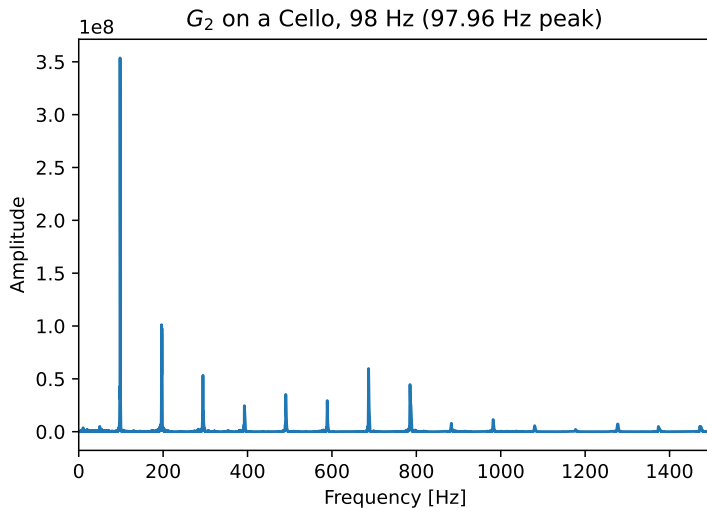


Vibrato

Periodic variation of pitch



Formant Shifting



Audio Manipulations

	Pitch	Inflection	Compression	Shelf-Filtering	Formant-shifting	Vibrato
Happy	+25c	-50c for 400ms	upward	+10dB per octave > 8kHz		

Audio Manipulations

	Pitch	Inflection	Compression	Shelf-Filtering	Formant-shifting	Vibrato
Happy	+25c	-50c for 400ms	upward	+10dB per octave > 8kHz		
Sad	-30c			-10dB per octave > 8kHz	yes	

Audio Manipulations

	Pitch	Inflection	Compression	Shelf-Filtering	Formant-shifting	Vibrato
Happy	+25c	-50c for 400ms	upward	+10dB per octave > 8kHz		
Sad	-30c			-10dB per octave > 8kHz	yes	
Afraid		+120c for 150ms				15c at 8.5Hz

Audio Manipulations

	Pitch	Inflection	Compression	Shelf-Filtering	Formant-shifting	Vibrato
Happy	+25c	-50c for 400ms	upward	+10dB per octave > 8kHz		
Sad	-30c			-10dB per octave > 8kHz	yes	
Afraid		+120c for 150ms				15c at 8.5Hz
Tensed	+100c	+150c for 150ms		+10dB per octave > 8kHz		

Text

“The Second Bakery Attack” from *The Elephant Vanishes* by Haruki Murakami (Original in Japanese パン屋再襲撃).

“[...] The text was judged emotionally neutral in all conditions [...]

Text

“The Second Bakery Attack” from *The Elephant Vanishes* by Haruki Murakami (Original in Japanese パン屋再襲撃).

“[...] The text was judged emotionally neutral in all conditions [...]”

“Sanity checked” the audio manipulations (in both French and Japanese):

- *Hardly* distinguishable from natural speech
- Correct emotion is conveyed

Experiment 1

French psychology students

(Gradually increasingly) Manipulated voice (excluding “tensed”, exclusive to experiment 2)

Evaluation by

- Measuring skin conductance level
- Detection questionnaire
- Mood scales (adjectives)

Result 1

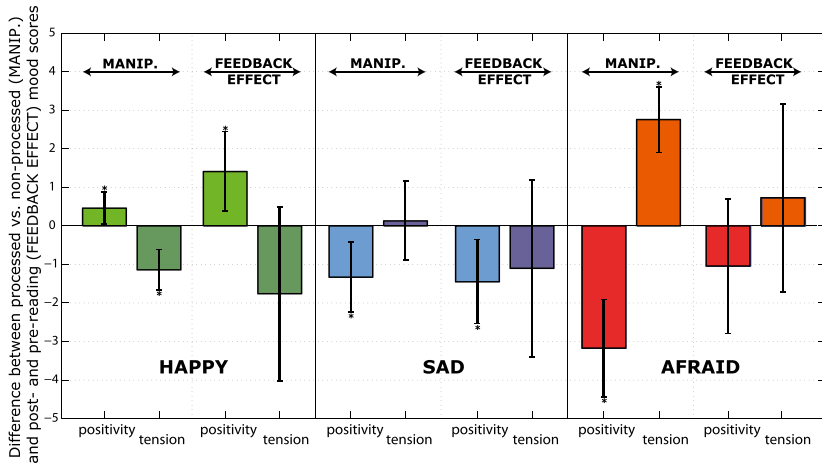
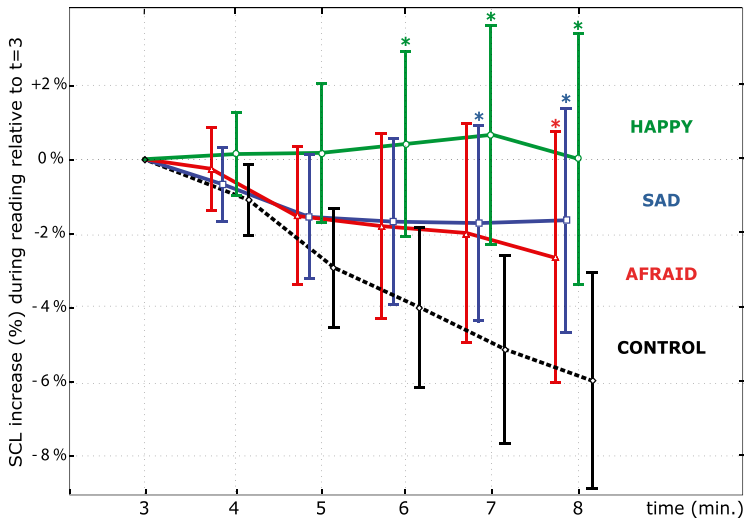


Figure: Positivity: happy/optimistic/sad
Tension: unsettled/anxious/relaxed

Result 1



Experiment 2

French psychology students

Same as before

Goal: confirm results and measure compensation/adaptation

Experiment 2

French psychology students

Same as before

Goal: confirm results and measure compensation/adaptation

Result: No significant compensation/adaptation

Discussion

Significance:

- Therapeutic applications (mood disorders)
- Enhance Karaoke, live singing, online conversations ...

Discussion

Significance:

- Therapeutic applications (mood disorders)
- Enhance Karaoke, live singing, online conversations ...

Questions:

- Different languages?