Phase 2 Week 3 - Speech Breathing Empathy Project

Nicolò Loddo

What happened: thoughts on the Study Design

- It would be fun to make it a guessing game with the possibility of continuing to guess
- It is also important to assess the quality of the synthesized speech with a MOS evaluation

Therefore the idea is:

- Demographics: age; sex; native language (maybe) -> 3 questions
- MOS evaluation on the quality -> 1 question

GUESS THE EMOTION QUIZ:

- 4 multichoice questions (4 choices)
- Possibility of continuing

REMOVED EVALUATIONS:

- No MOS on emotionality because emotionality will be assessed in the quiz; No evaluation of linguistic content.

P.S.: I started considering an Open Source implementation of FastSpeech2 as model.

The Quiz is described in the next slide.

Emotional Conditions: [possibly providing additional statistically significant results]

(as defined by James Russell's Circumplex Model through Arousal and Valence parameters;
adjective placement from Nagel et al.'s study:
www.researchgate.net/publication/45189833_Worms_in_Emotion_Visualizing_Powerful_Emotional_Music)

- High arousal, negative valence (Annoyed)
- High arousal, positive valence (Delighted/Excited)
- Low arousal, negative valence (Sad)
- Low arousal, positive valence (Relaxed/Serene)

Speech Features Conditions: [actually analysed in RQ]

- Without breathing
- Without filled pauses
- Without pitch contour
- Full features

Quiz parameters:

- Randomly extracted Sentences (with same linguistic emotional content, as recognized by ER model), never the same sentence, to not provide a comparison baseline
- Always randomly extracted Emotion Condition, can happen to be the same
- 1 question per Feature Condition
- If the subject continues with the quiz: randomly extracted Feature Condition (max 4 more)

*Number of conditions: 4x4=**16**. The conditions at issue in RQ are though only **4**: the Speech Features Conditions.*
*For our maximum 8 questions I need **8** chosen sentences with same linguistic emotional content.*
*I have to synthesize them across all conditions, for a total of: **128** recordings.*