Naturalness perception of different synthesis methods for high frequencies

Participants

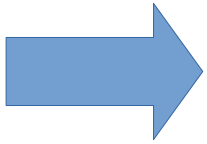
- Age 18-25
- Half males, half females
- French native speakers
- Tonal audiometry from 500 Hz to 12500 Hz
  - Exclude when hearing loss > 20 dB

Stimuli

- 5 vowels : /a/, /i/, /e/, /o/ and /u/
- 2 speakers : one female and one male
- 2 voice qualities : modal and pressed
- 3 different methods to generate high frequencies

Stimuli

- 3D TF from MM
- 1D : blending 0-4 kHz MM + 4-12 kHz 1D
- Bandwith extension of 0-4 kHz part of 3D stimuli



All stimuli are similar in 0-4 kHz

Experiments

- Expe 1 : Pairwise comparison of naturalness
- Expe 2 : Rating of naturalness from 0 to 100

Expe 1 : Pairwise comparison

- 3 types of pairs : 3D-1D, 3D-BWE and 1D-BWE
- 5 vowels X 2 genders X 2 voice qualities
  X 3 types of pairs X 2 moments
  = 120 pairs to evaluate
- Task : Evaluate which stimuli of each pair is more natural than the other (with option none?)

Expe 2 : Naturalness rating

- 5 vowels X 2 genders X 2 voice qualities
  X 3 models X 2 moments
  = 120 stimuli to evaluate
- Task : Rate the naturalness on a scale from '*not natural at all*' 0 to '*natural like real speech*' 100

Hypothesis 1st exp

- General hypothesis : the most realistic, the most natural
  - 3D more natural than 1D
  - 3D more natural than BWE
  - 1D more natural than BWE
- More HF in pressed voice should make difference more clearly perceived

Hypothesis 2d exp

- General hypothesis : the most realistic, the most natural
  - 3D more highly rated than 1D
  - 3D more highly rated than BWE
  - 1D more highly rated than BWE
- Some vowels should be perceived as more natural than others (result previous exp)