Human Pause Detection in Spontaneous Speech in an Online-Experiment

Jürgen Trouvain & Raphael Werner

Language Science and Technology, Saarland University, Saarbrücken (Germany)

1 Introduction

Functions of pauses:

- syntactic-prosodic breaks
- transitions in turn-taking
- hesitations
- emphasis

Cues for perception of pauses:

- silence
- inbreath noises
- hesitation particles
- phrase-final lengthening
- intonational boundaries
- voice quality changes
- intensity drops
- syntactic information

Focus of this pause detection study in spontaneous speech:

- reaction times
- main cues
- agreement on pause locations
- human vs. automatic detection

2 Experiment

Material from German dialogs:

- random selection of 160 sec speaking time from GECO [1]

Pause types (mean durations):

1. 32 w/ breath noise (843 ms)
2. 3 w/out breath noise (696 ms)
3. 3 w/ laughter (1563 ms)
4. 14 w/ hesitation (silent; lengthenings; fillers) (455 ms)

Stimuli (n=16):

- originals: 2 in each of 4 classes: 5, 10, 15, 50 sec
- copies of originals also manipulated (see Fig. 1):
  - breath noise replaced by silence (types 1+3)
  - silence (and potential fillers) completely removed (types 2+4)

Subjects (n=12):

- students from intro class to phonetics
- basic skills of annotation w/ Praat
- task: just listen and tap key when you hear a pause

Annotation of each pause:

- detection (yes/no)
- reaction time from silence onset (-500ms ≤ RT ≤ 1000ms)

Automatic pause detection with Praat script [2]

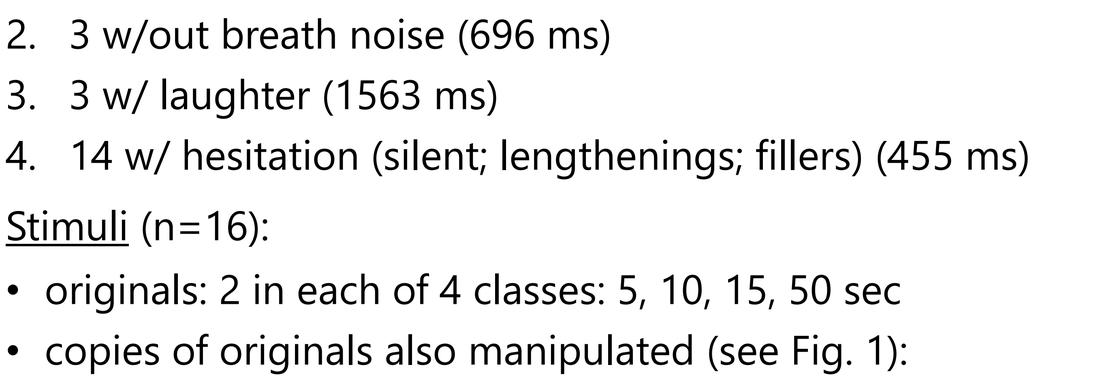

Fig. 1: Ex. of an original (left, c. 3 sec) and manipulation (right, silence removed)

3 Results

Differences between subjects:

- detection rate and reaction time (see Fig. 2)
- different strategies at work

Differences between pause types:

- pauses w/ breath noises detected by all subjects, also when breath noise replaced with silence
- detection of pauses w/ hesitation strongly varied, manipulated versions always lower detection rate
- 25% detection of manipulated pauses without any silence

Automatic detection:

- correct for all types of pauses except with removed silence
- problems with pauses containing laughter and hesitations like filler particles and lengthened syllables

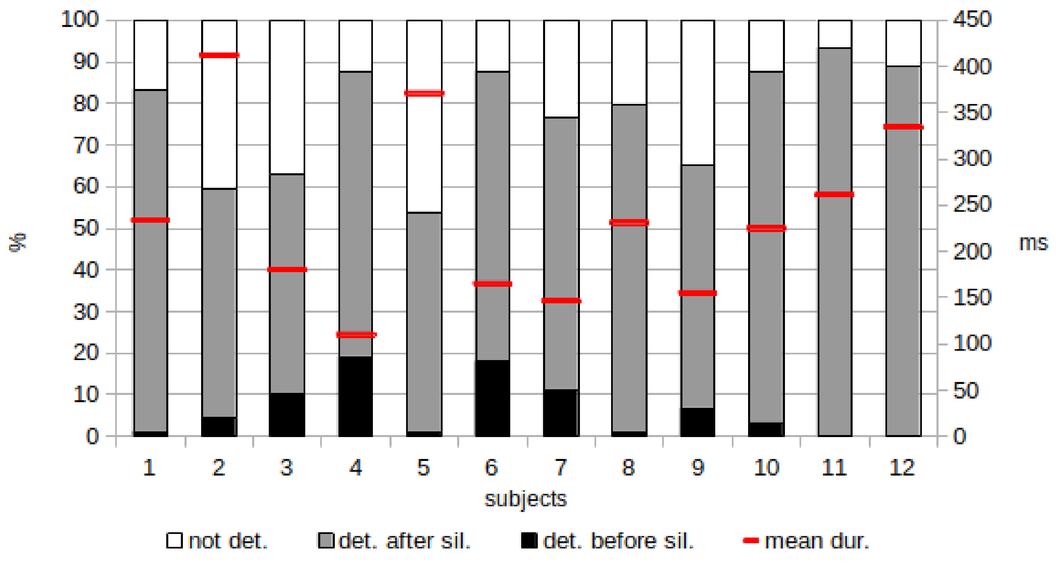

Fig. 2: Rate (in %) for detected pauses before silence onset (black), after silence onset (grey), and not detected (white) and reaction times (red, in ms)

4 Discussion

- Human detection of pauses not as easy as expected, individuals strongly differ in detection rate and reaction time
- Perceived pauses not necessarily need overt silence
  → pause in *perception* different to *production & acoustics*
- Fast pause detection required for using transitions in turn-taking and places for backchannelling *ideally before* silence
- Hesitation pauses with lower detection rate
  → unclear concept of "filled pause" in perception
- Human detection superior to automatic detection for pauses w/ removed silence, w/ hesitation, w/ laughter
- Experimental setup w/ skilled subjects feasible

Acknowledgements: This research was funded in part by DFG under grants TR 468/3-1 and MO 597/10-1. **References**: [1] Schweitzer & Lewandowski 2013. Convergence of articulation rate in spontaneous speech. Interspeech. [2] de Jong & Wempe 2009. Praat script to detect syllable nuclei and measure speech rate automatically. Behavior Research Methods, 41(2).