A corpus-based analysis of back-channel vocalizations

Sathish Pammi and Marc Schröder
DFKI GmbH

Outline

- Introduction
- Motivation
- Previous work
- Our attempt
- Method for database collection
- Methods for Annotation
- Results
- Ongoing work
- Summary

Introduction

- Back-channel vocalizations play an important role in communicating listener intentions while the other person has the turn or other is talking.
- Back-channels transmit :
  - Messages like 'I am listening' or 'I am with you'
    - Examples: um, hu-huh
  - Listener affective states like excited, bored, confused, surprised and so on
    - Examples: wow! , laughter

Motivation

- Synthesis of back-channel vocalizations is one of the focused research areas to improve emotionally colored conversational speech synthesis
- It includes different research questions like:
  - where to synthesize?
  - what to synthesize?
  - what kind of acoustic properties have to be obeyed to communicate different affective states in different situations?

Previous work

- Attempts made in this Area:
  - The importance of affect bursts as a feedback in a conversation was investigated (Schröder et al, 2006) through listening tests
  - Nigel Ward and Wataru Tsukahara (2000) had developed some rules to generate back-channel responses in a conversation and investigated how to use low pitch regions as cues for back-channel responses.

Our attempt

- To improve Interactive speech synthesis, we have to study:
  - The analysis of different back-channel vocalizations
  - Identification of distinguishable types among back-channel vocalizations
  - Their acoustic properties
  - Affective states behind them
- In this work, we report:
  - A method for collecting back-channel vocalizations
  - Ongoing research work on annotation and data analysis

Method for database collection

- Dialog speech recorded when a professional German actor was engaged in a conversation with student assistants
- Recordings were made in sessions of about 20 minutes each.
- Instructions were given to the actor to keep the conversation live as long as he can act like a specific character among Spike(aggressive),Obadiah(gloomy), Poppy(happy) and Prudence(neutral).

Method for database collection-2

- Our student assistants, acting as dialog partners, tried to keep the actor in listening mode for maximum amount of time
- The speakers were sitting in separate rooms but saw each other through a glass wall.
- Each speaker was listening his/her dialog partner's speech using headphones.
- Each speaker's voice was recorded on a separate channel.

Examples

- Prudence 
- Poppy 
- Spike 
- Obadiah 

Methods for Annotation

- Many non-verbal vocalizations in the corpus belong to three broad categories: back-channel, affective and laughter vocalizations
- ABL – annotation schema
  - A stands for Affective
  - B stands for Back-channel
  - L for Laughter
- The corpus was annotated according to ABL-schema using Praat software.
- 'floor' or 'turn' annotated with a silence detector algorithm by processing two channels separately

Instructions given to annotators

- ABL – Annotation scheme

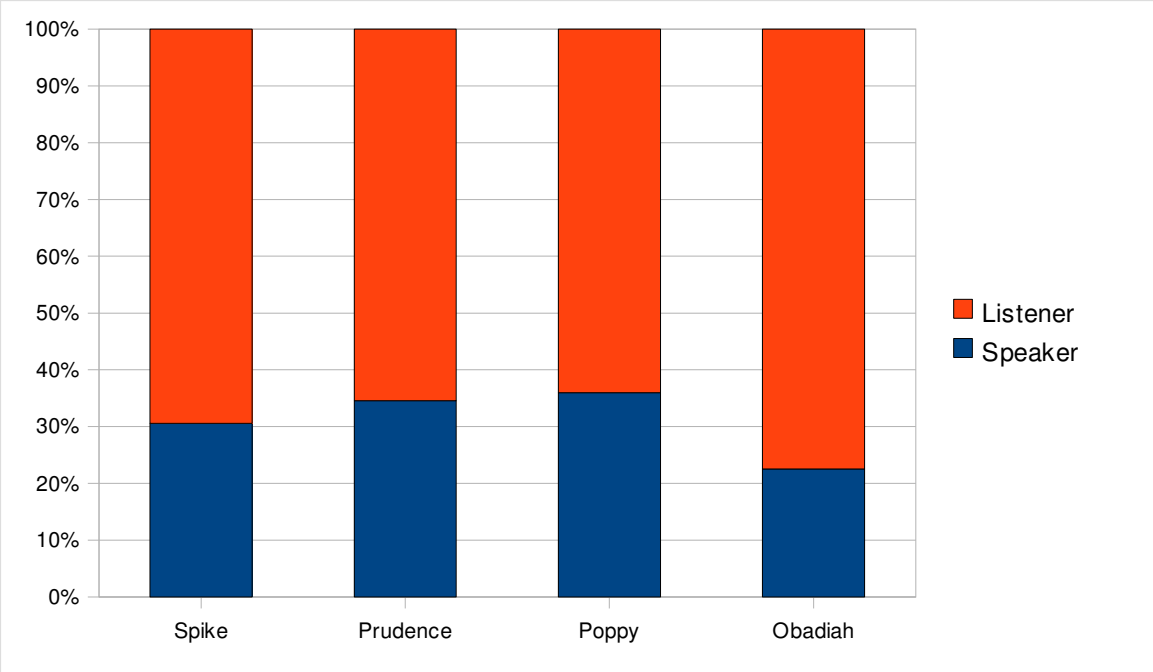
  A – communicates Affective states

  B – communicates something when the actor is in listener mode.

  L – when the non-verbal vocalization is perceived like Laughter.

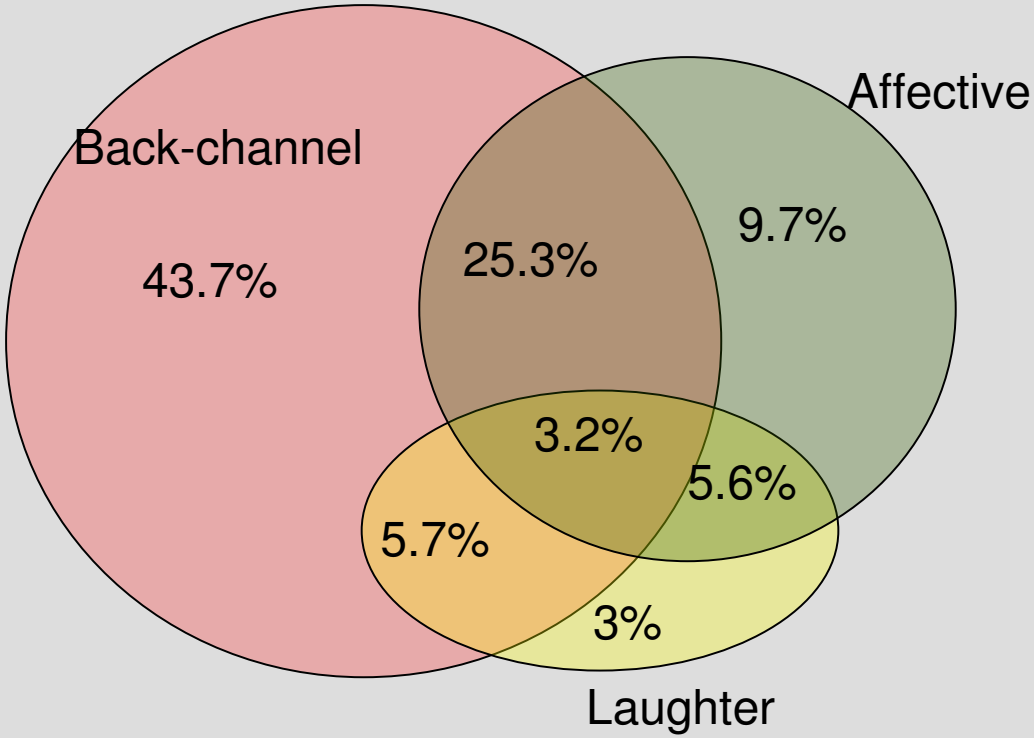
Results - 1

- Dialog speech ~ 6 hours
- Found 1175 non-verbal vocalizations



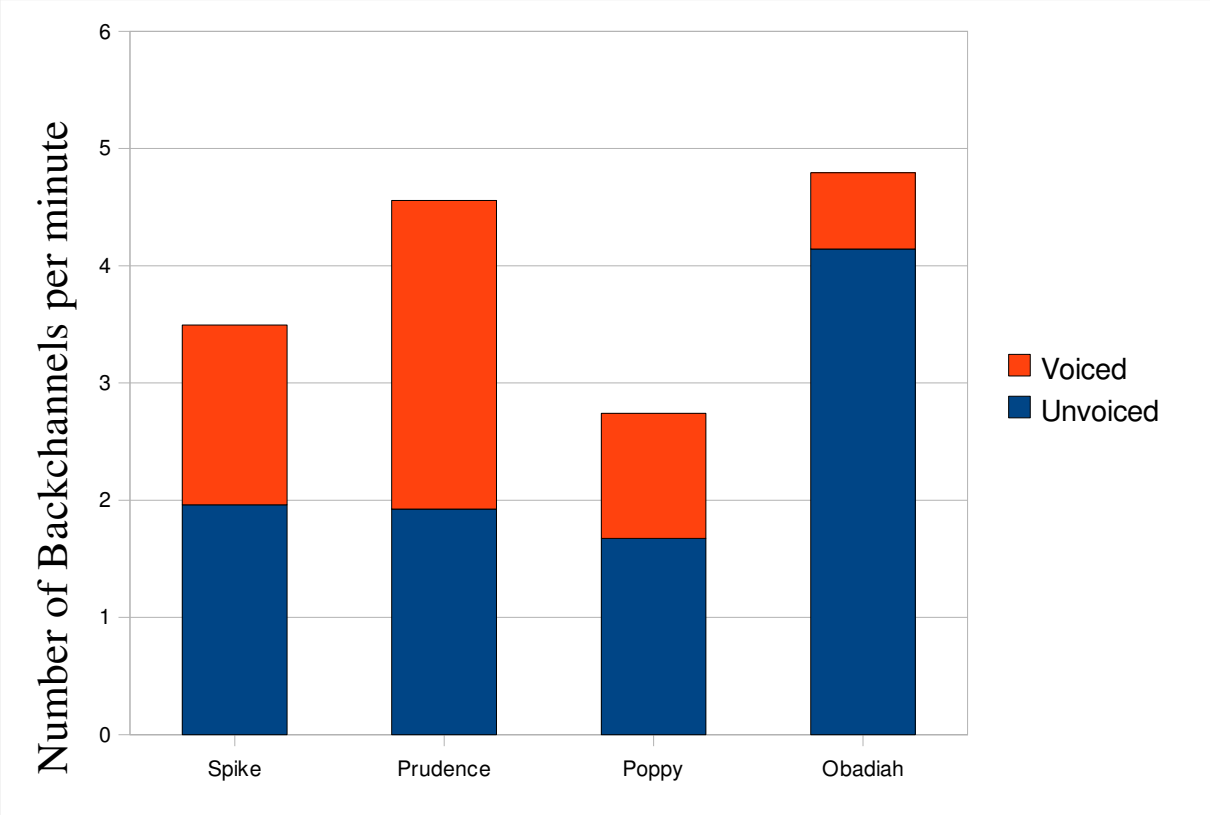
Percentage of Speaker/Listener mode of Actor in conversation

Results - 2



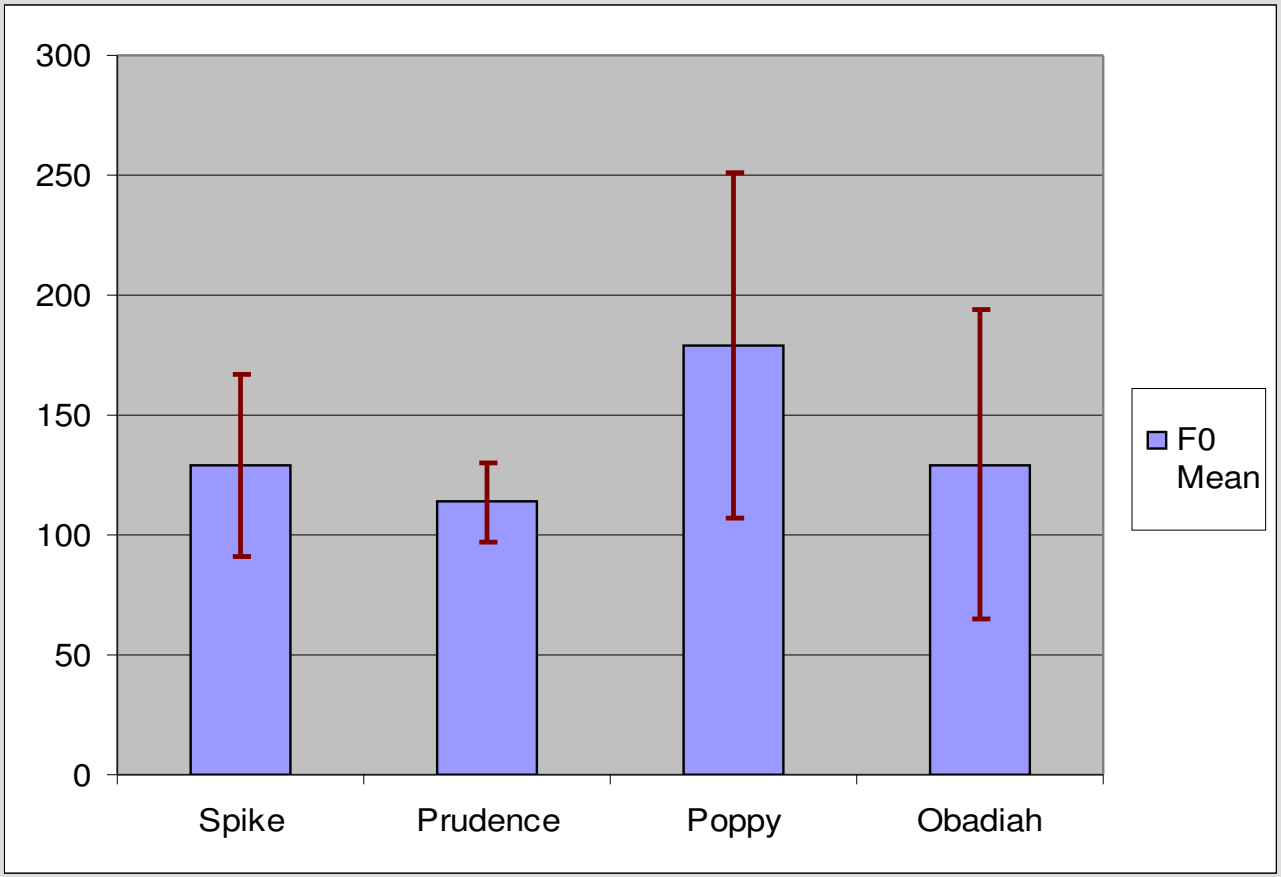
Distribution of non-verbal vocalizations

Results - 3



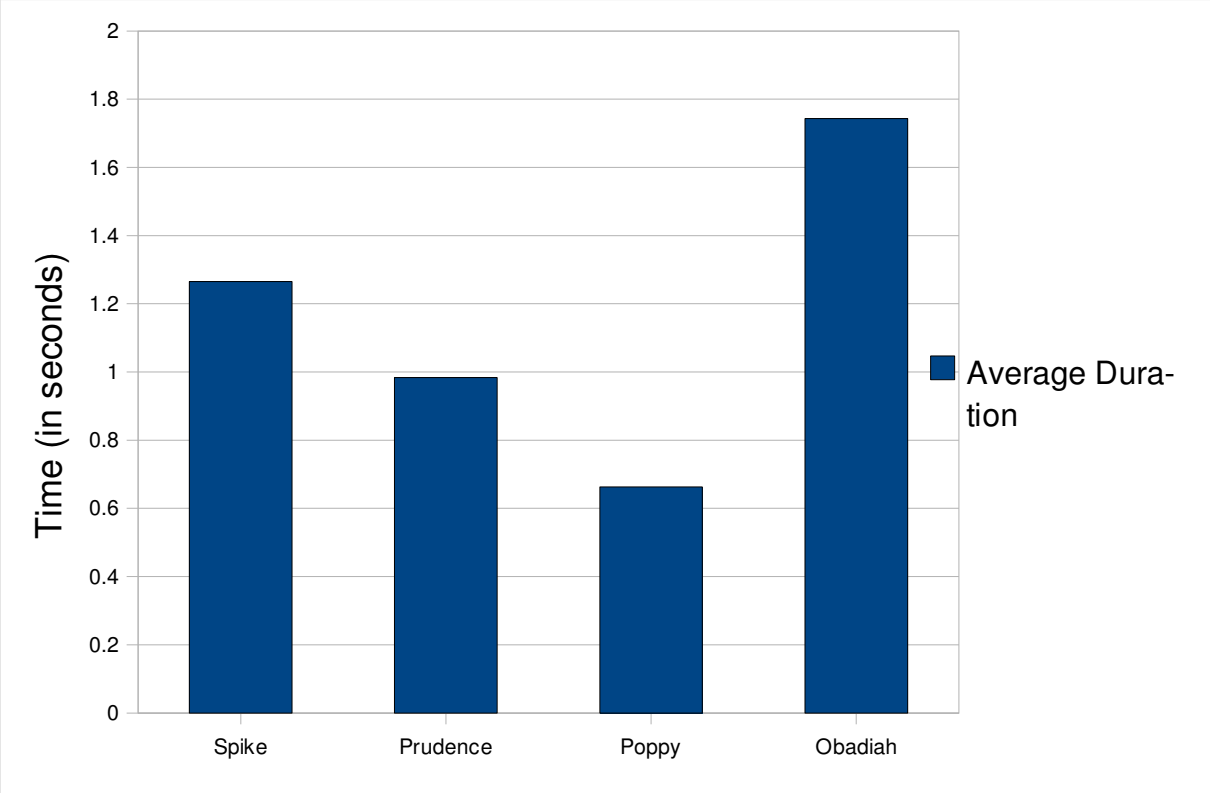
Average number of back-channels produced by different characters per minute and percentage of voiced-unvoiced vocalizations

Results - 4



Pitch (F0) mean and standard deviation of voiced back-channel vocalizations

Results - 6

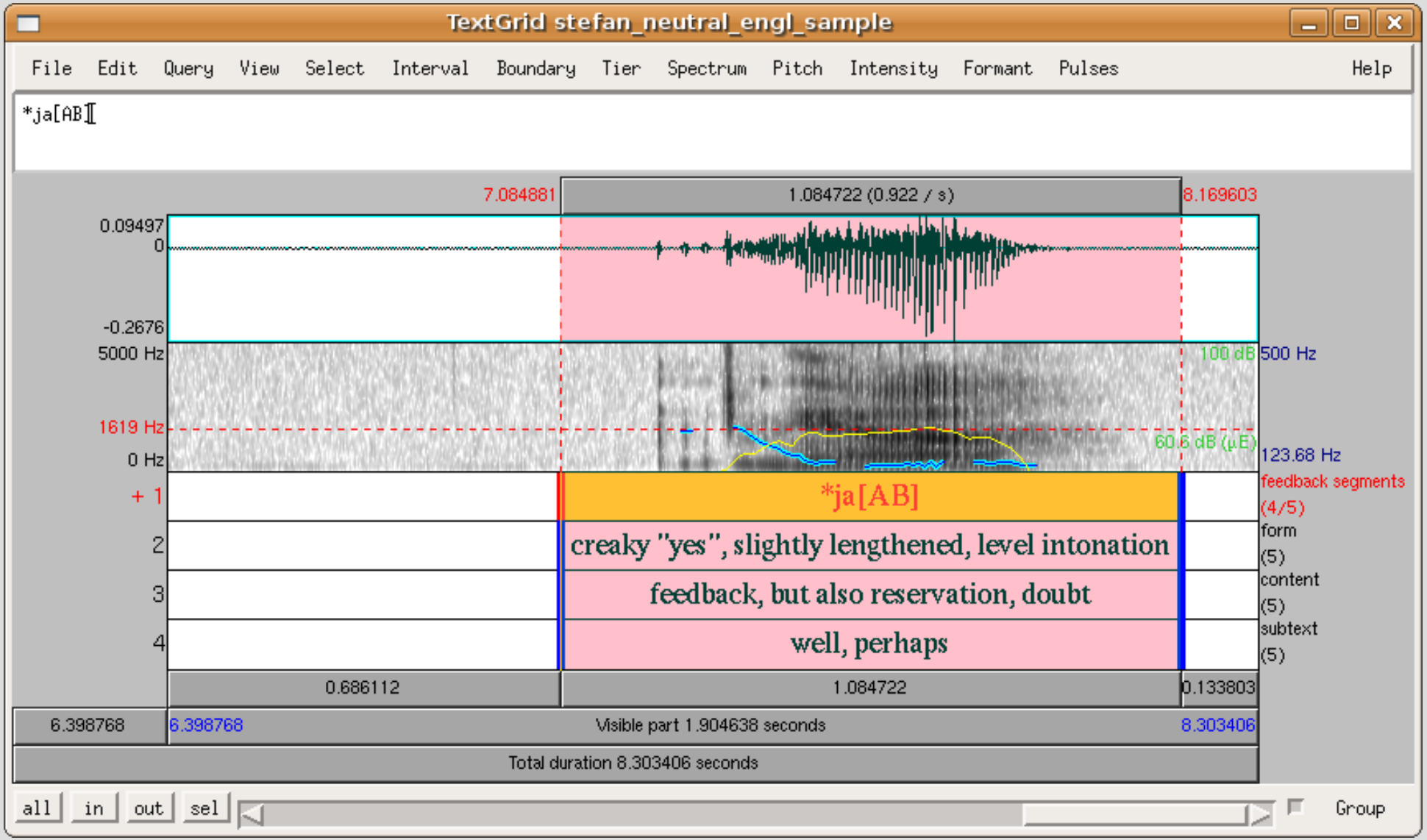


Average duration of back-channel vocalizations in different characters

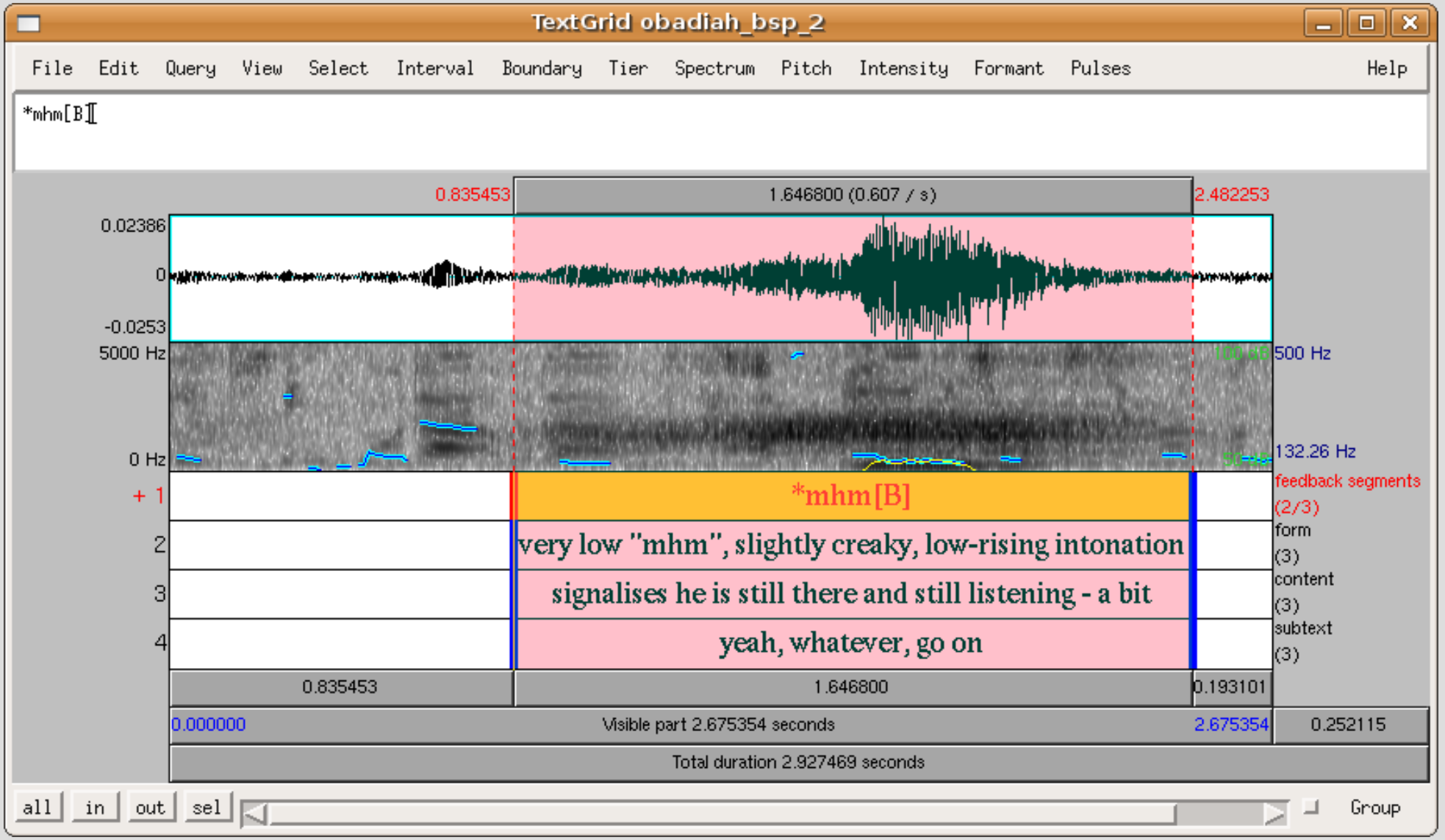
Ongoing work

- To understand better the structure of both behavior and function of non-verbal vocalizations:
  - Annotate non-verbals using informal descriptions
  - Informal descriptions in 3 levels:
    - Function
    - Behavior
    - Sub-texts
- These descriptions will help to understand the types of form and meaning of non-verbal vocalizations

Informal descriptions - 1



Informal descriptions - 2



Summary

- Useful observations in terms of speech synthesis
  - The gloomy character (Obadiah) produces an average of 4.8 back-channels per minute, most of them nasal sounds with long durations around 1.74 seconds
  - The happy character (Poppy) utters only 2.7 back-channel responses per minute, which are relatively short utterances around 0.66 seconds.
- We are trying to identify distinguishable types among back-channels in different levels using informal descriptions.

Questions?

Thank you!