

On Temporal Organization of Spontaneous Estonian: Preliminary Analysis Results of Lecture Speech

Einar Meister¹ and Pärtel Lippus²

¹Laboratory of Phonetics and Speech Technology
Institute of Cybernetics at Tallinn University of Technology
Akadeemia tee 21, 12618 Tallinn, Estonia
einar@ioc.ee

²Department of Estonian and Finno-Ugric Linguistics
University of Tartu
Ülikooli 18, 50090 Tartu, Estonia
partel@murre.ut.ee

Motivation:

- spontaneous speech differs from read speech in many aspects
- descriptions of Estonian prosody are based mainly on laboratory speech
- most studies on Estonian conversational speech do not involve phonetic analysis
- adequate temporal models are necessary for automatic speech recognition of spontaneous speech as well as for synthesis of different speech styles

Goals:

- to study temporal structure of spontaneous Estonian involving methods of acoustic-phonetic analysis
- to analyze different variations of spontaneous speech including lecture speech, monologues, dialogues, interviews, etc.
- to classify and characterize different prosodic units and boundaries
- to provide new knowledge for the development of computational prosody models

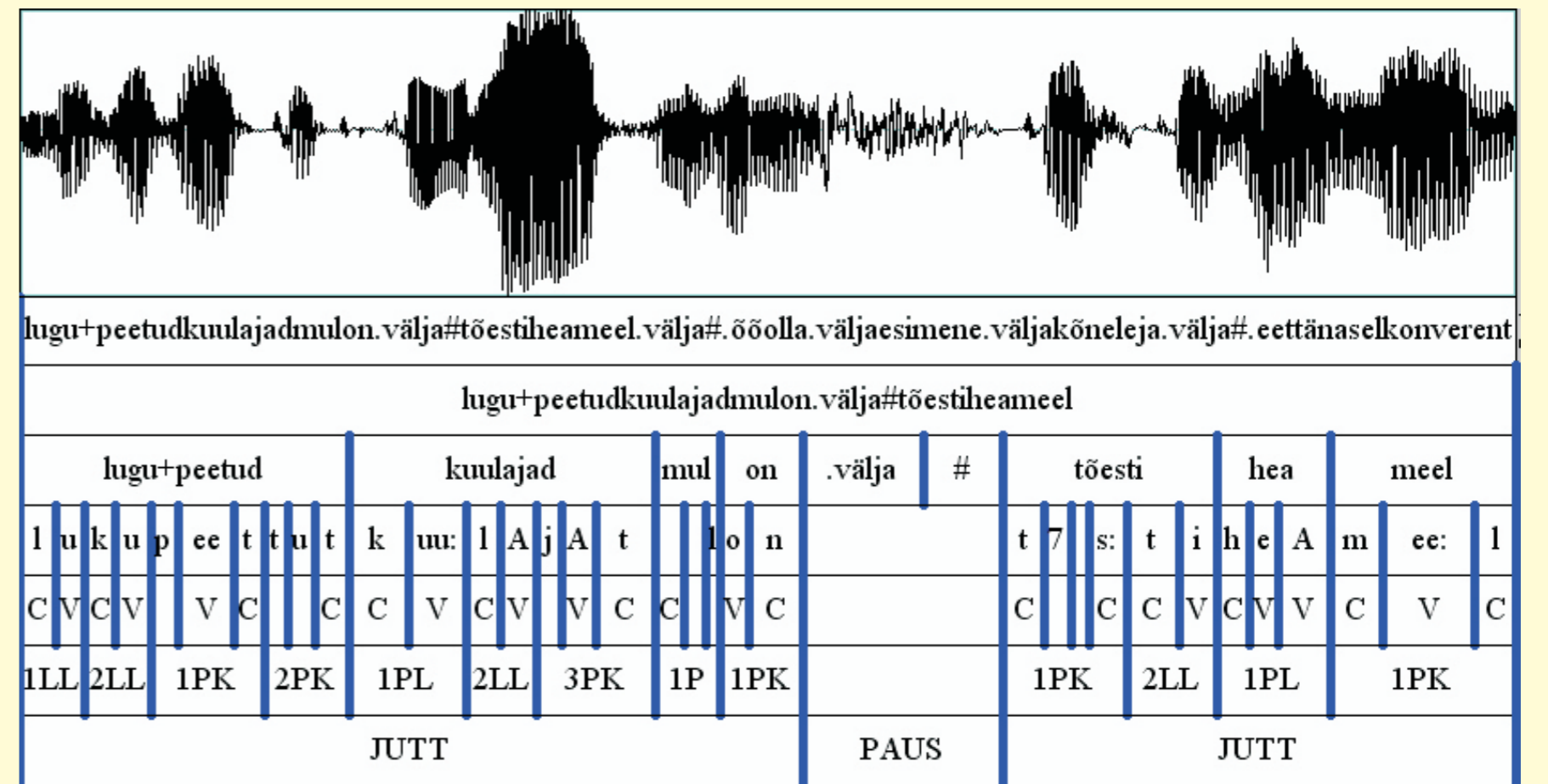


Figure 1. Segmentation example in PRAAT.

Spontaneous vs. read speech:

- hesitations
- repeating words
- self-repairs
- fillers
- incomplete utterances
- lengthened speech units
- different types of pauses
- mismatch of prosodic and grammatical units
- etc.

Disfluencies in spontaneous speech:

- cognitive - problems in speech planning process
- communicative - depending on communication situation
- physiological - breathing

Theoretical framework:

- Sociolinguistics (Labov, Chafe, Biber, etc.)
- Conversation Analysis (Sacks, Schegloff, Jefferson, Wooffitt, etc.)
- Intonational Analysis (Pierrehumbert, Beckman, etc.)

Diverse units in spontaneous speech:

- speech act
- idea unit
- information unit
- intonation unit, intonational phrase, tone group
- prosodic unit, prosodic phrase, prosodic group, prosodic utterance
- breath group
- etc

Previous research on spontaneous Estonian and temporal characteristics:

- conversational analysis of spoken Estonian, dialogue modeling (Hennoste, Koit, Gerassimenko, Rääbis, Kasterpalu, Strandson, Valdisoo, Treumuth, ...)
- temporal analysis and modeling of speech units and pauses in read speech for text-to-speech synthesis (Mihkla, Pajupuu, Kerge, Kuusik, Fishel, ...)
- word-level temporal characteristics of read speech (Lehiste, Liiv, Eek, ...)
- some temporal aspects of spontaneous Estonian (Krull, ...)

Corpora of spontaneous Estonian:

- Corpus of Spoken Estonian (Tartu University, on-going project since 1997) - mainly for conversational analysis and dialogue modelling, not suitable for acoustic-phonetic analysis
- Phonetic Corpus of Estonian Spontaneous Speech (Tartu University, project started in 2006) - mainly for acoustic-phonetic studies
- Database of Spontaneous Estonian (Institute of Cybernetics, project started in 2006) - mainly for automatic speech recognition

Speech material of the study:

- **lecture speech** ca 45 minutes recorded during public presentation
- **speaker:** male, Estonian, age of 48 years, standard speech
- **recording equipment:** portable digital recorder **Edirol R1** (16 bit, 44,1 kHz, linear, mono), head-mounted close-talking microphone **AKG C 444L** with preamplifier
- **presentation style:** PowerPoint slides
- **speaking style:** semi-spontaneous

Definition of major speech units and pauses:

- TOPIC - a time interval for presentation of a single PowerPoint-slide
- TOPIC PAUSE (TP) - a pause for slide change
 - BREATH GROUP (BG) - a chunk of speech between two successive inhalations
 - BREATH GROUP PAUSE (BGP) - a pause between BGs (inhalation)
 - PROSODIC GROUP (PG) - a speech chunk between two successive silent or filled pauses within BG
 - PROSODIC GROUP PAUSE (PGP) - a pause between PGs

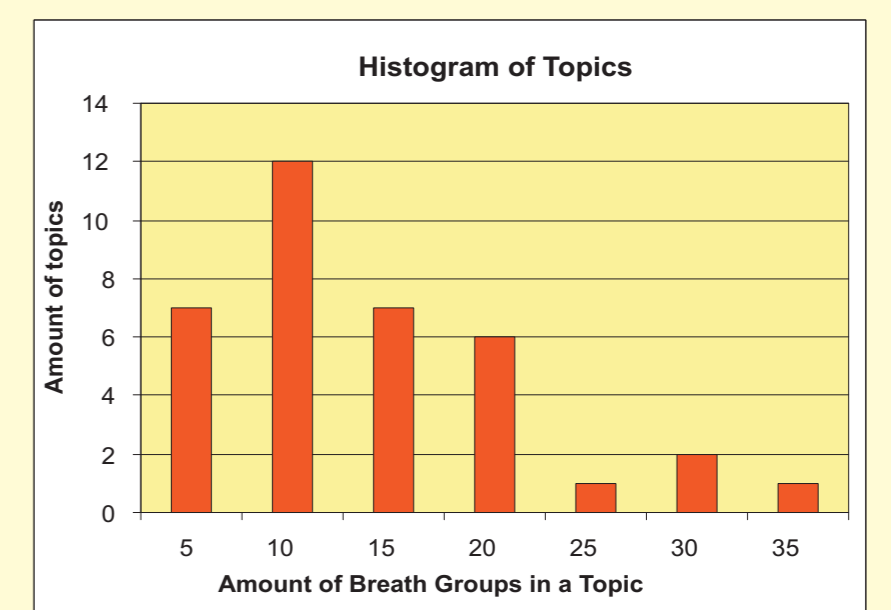
Segmentation levels:

- topic
 - breath group (BG)
 - prosodic group (PG)
 - word
 - syllable
 - phoneme

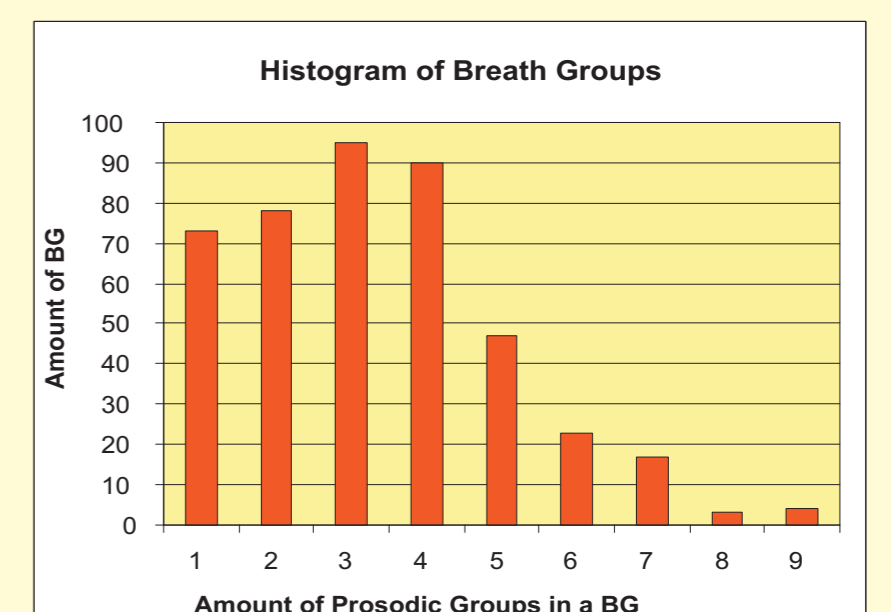
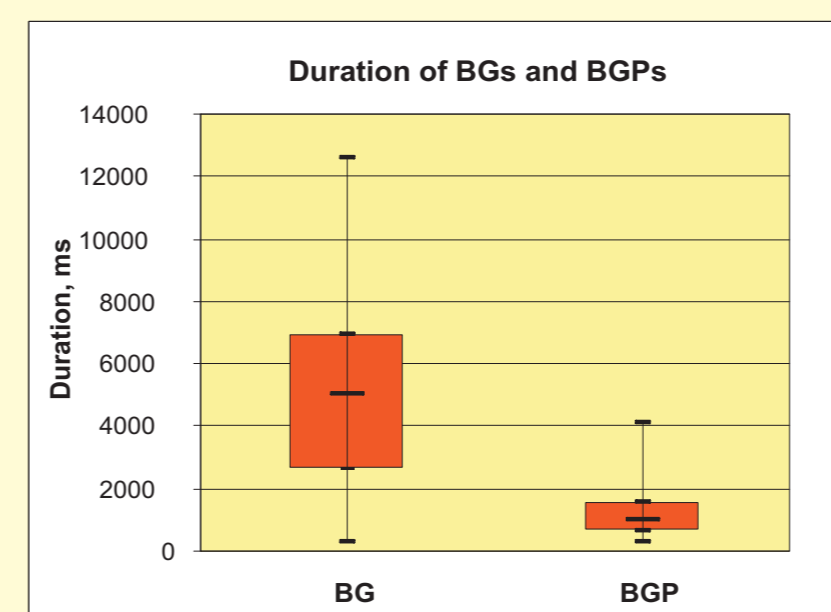
PRELIMINARY FINDINGS

TOPICS:

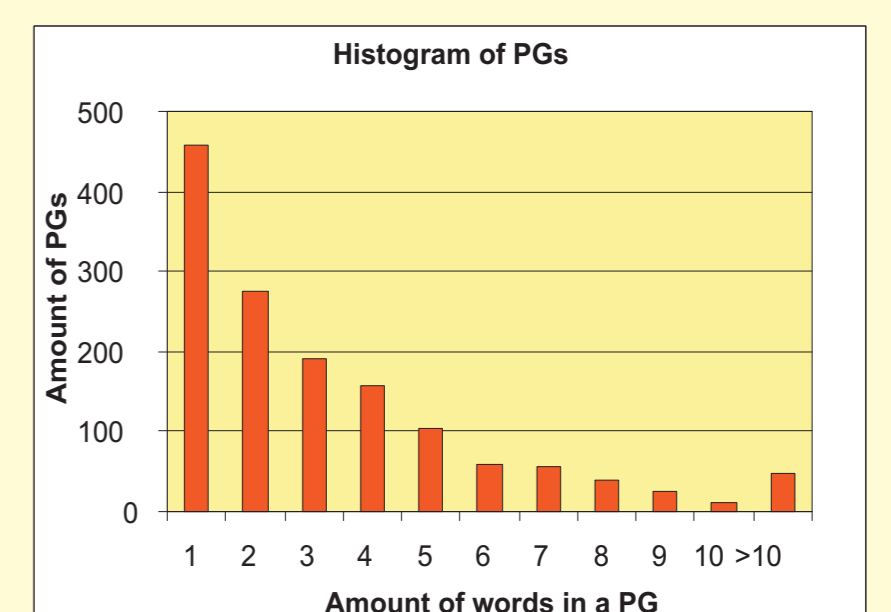
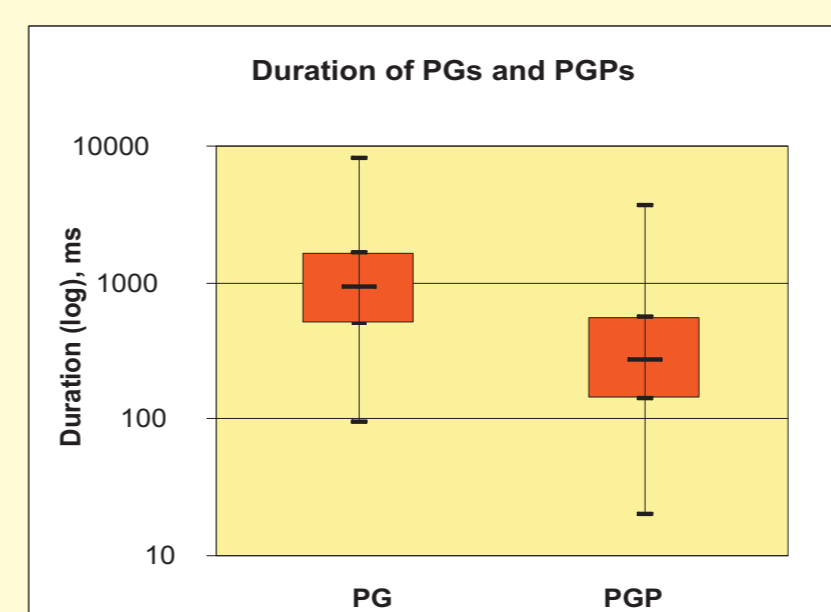
- 36 topics
- min duration: 13 sec
- average/median duration: 72/55 sec
- max duration: 210 sec
- average/median duration of topic pause (TP): 3.2/3.1 sec



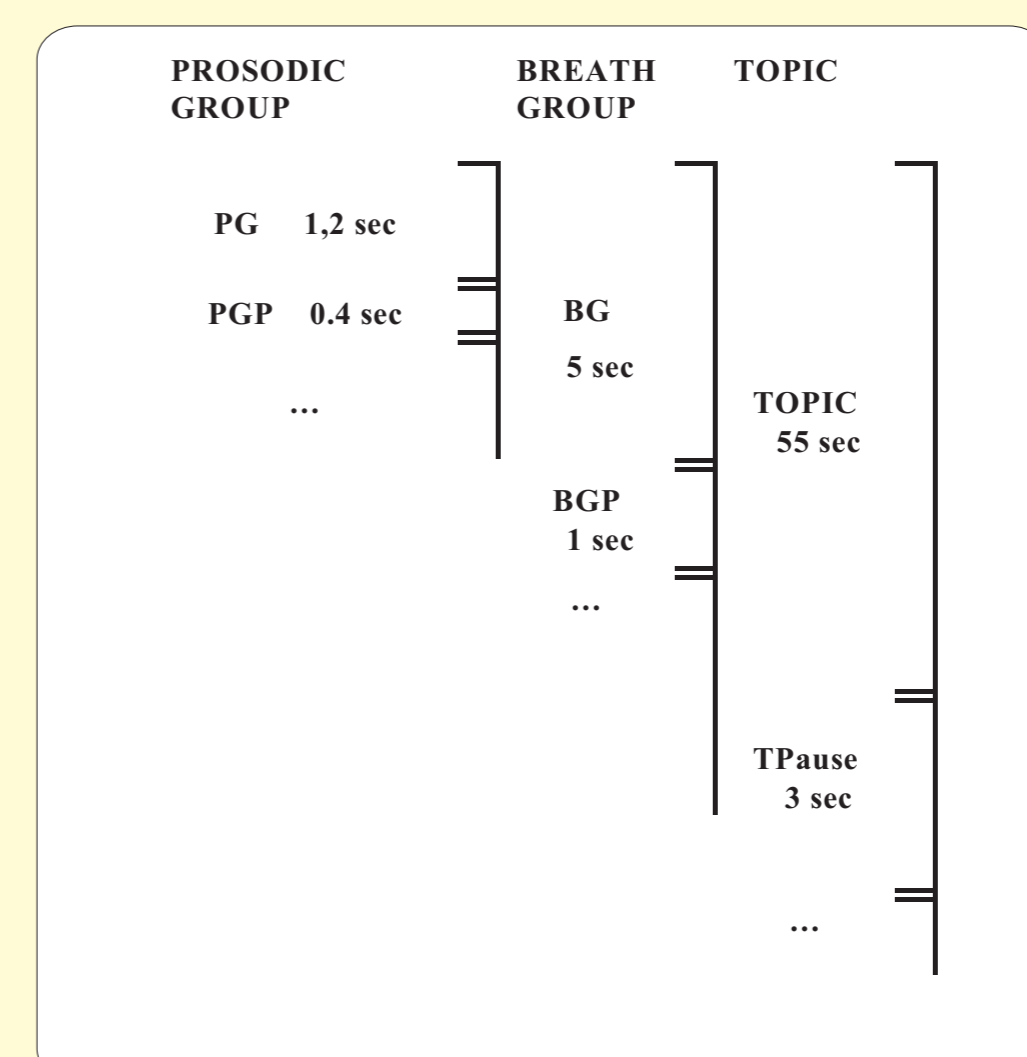
BREATH GROUP



PROSODIC GROUP



A Model of Temporal Structure



Summary and future work

- global picture of temporal organisation of lecture speech
- single-speaker data can not be generalised
- further analysis of acoustic characteristics of prosodic boundaries and groups
- study of lexical and syntactic structure of prosodic groups
- recording and analysis of new speakers

Acknowledgements

The study is supported by:
National Language Technology Programme (2006-2010)
Doctoral School of Linguistics and Language Technology

