Towards automatic speech processing and intonational labeling for different speaking styles

Helena Moniz, Fernando Batista, Ana Isabel Mata, Isabel Trancoso

1 Faculdade de Letras da Universidade de Lisboa
2 Engenharia de Sistemas e Computadores – Investigação e Desenvolvimento em Lisboa
3 Instituto Universitário de Lisboa (ISCTE-IUL)
4 Instituto Superior Técnico

• Goals
This work aims at presenting a framework which encompasses multilayered linguistic information, focusing on prosodic features (pitch, energy, and tempo patterns), and at contributing to the characterization of intonational patterns in European Portuguese, based on large data sets of speech representing different speaking styles.

• ASR architecture
• Audimus (Neto et al., 2008): broadcast news pipeline
• ASR system both in a fully automatic and in a forced alignment mode

• Corpora
• ALERT (broadcast news - 70h)
• LECTRA (university lectures - 31h)
• CPE-FACES (high-school presentations - 15h)
• CORAL (map-task dialogues - 9h)

• Creating the automatic prosodic description
• Pitch and energy extraction: Snack Sound Toolkit
• Adjusting phones: monophone and diphone units
• Marking syllable boundaries and stress: boundaries and stress with heuristics

• Adjusting word boundaries and silent pauses: informed by pitch/energy and lexica
• Testing on structural metadata events

• Framework usage
• Disfluencies
• Prosodic contrast marking from disfluency to repair of fluency
• Style effects in the production of disfluencies
• Punctuation and capitalization
• Detection of commas, full stops and question marks
• Set of prosodic features to discriminate between punctuation marks and disfluency boundaries
• Analysis of left dislocated structures in the scope of the Project COPAS – in progress
• 40 minutes with ToBI annotations

• Conclusions
The framework proved to be suitable for the analysis of structural metadata events, allowing for:
• empirical evidence regarding intonational patterns of disfluent events across different speaking styles
• discrimination between punctuation marks based on informative sets of prosodic features
• discrimination between sentence-like units and disfluency boundaries
• Step forward towards both fully automatic processing of multilayered linguistic description of spoken corpora and intonational labeling in European Portuguese
• Our main milestones concern the linguistic structured knowledge resulting from our data-driven approach and also the application of the framework on national ongoing projects.

• Acknowledgments
This work was supported by FCT – Fundação para a Ciência e a Tecnologia – under project PTDC/CLE-LIN/120017/2010.