Ph.D. in Information Technology: Thesis Defense

February 18th, 2019

Room PT1– 10.30 am

Sonia CENCESCHI– XXXI Cycle

"Speech analysis for Automatic Prosody Recognition"

Advisor: Prof. **Licia Sbattella**

Abstract:

This thesis presents a wide-ranging research work on prosody. Prosody is defined as the group of audio paralinguistic and suprasegmental clues involved in the communicative and understanding process of human speech. According to the main Universals in language, each Speech Act expresses common needs (e.g., talking about past or future events) similar for all humans, which are acoustically realized according to linguistic and phonotactics language related rules. At the same time, a spoken message can be uttered with a variable prosody because of countless factors as social context, emotions, intentions, rhetoric or spatial dislocation.
The work starts proposing a new descriptive model in order to analyze prosody complexity in a structured and orderly manner, within which the sound of an utterance is considered as the final product of many exogenous and endogenous influences referring to the speaker. An Italian recited-speech corpus and a psychoacoustic experiment were built in order to validate part of the model and to analyze the influence of semantics, phonotaxis and intonation on understanding processes. Results have been useful to defining the feature set to rely on the following parts of the work, regarding automatic recognition.
Two neural network architectures have been developed, both of them regarding the Italian language. The first concerns the recognition of statements, questions and exclamations (using both textual and sound inputs), while the second identifies the presence of corrective focus into utterances (sound inputs only).
A last section is focused on the semi-automatic characterization of prosody, laying the groundwork for further automatic recognition systems focused on prosodic skills. A monitoring protocol of expressivity and vocal qualities based on features extraction is then described, followed by practical applications to clinical, educational and forensics fields.
The main contributions of this thesis are the definition of a new multi-dimensional conceptual model describing prosodic forms, two NNs based architectures for structures and corrective focus detection, two new audio/textual corpuses composed by recited and read speech used to feed NNs, and a proposal for the semi-automatic analysis of some aspects of prosody and expressiveness.

PhD Committee:
Prof. **Barbara Pernici**, DEIB
Prof. **Giulia Bencini**, Universita' Cà Foscari
Prof. **María Inés Torres,** Universidad del Pais Vasco