"Actors Challenge": Collecting data to study prosodic patterns & their mapping to meanings across.

Sia Vosh Sepanta, Natallia Chaiko & Roberto Zamparelli
University of Trento, CIMeC
siavosh.sepanta@unitn.it; natallia.chaiko@unitn.it;
roberto.zamparelli@unitn.it

The study of Affective Prosody, an umbrella term that includes emotional and attitudinal prosody (Ross, 2000), involves mapping prosodic patterns to the information structure encoded in text. Researchers working to establish such correspondence would benefit immensely from an annotated set of cross-linguistic data in which contexts are mapped onto utterances produced according to prosodic patterns appropriate for that context. While there are various multilingual/-modal data sets of labeled phrases, e.g. (Bagher Zadeh, 2020) used for Speech Emotion Recognition, and multilingual multisubject corpora of spoken language, the community is still missing a comprehensive multilingual oral dataset of minimal pairs uttered by a large number of speakers, suitable for studying emotional as well as attitudinal prosodic patterns across languages.

This need prompted us to implement a web-based game, *Actors Challenge* (AC), designed to collect prosody-meaning mappings in multiple languages in the form of recordings that are simultaneously generated and validated by the players. The game, which has just been launched in 4 languages (English, Italian, German and French), works as follows: To target emotional prosody, we prepare a series of target phrases that could be uttered in various contexts and are as neutral as possible in their affective value, such as "It's a cappuccino". Similarly, for the attitudinal prosody, the chosen targets should lend themselves to various topics of study, e.g., focus: "Kevin isn't drinking because he is unhappy." Next, we prepare a set of contexts in which the target could fit. These contexts give the background to understand how the target should be uttered: triggering emotions, resolving ambiguity, changing focus, etc. Participants alternate in two roles: *Auditioners*, where they record and submit their voices reciting the given target phrase in each context; *Casting Directors*, where they listen to the recordings of other players and are asked to choose the context they think the actor was aiming for, and rate the performance. Each player obtains an acting score based on how often their recordings were correctly matched, weighed by their rating. Research ideas benefiting from this dataset include studying the ambiguity or complexity of different prosodic patterns or attitudes, detecting mismatches between production and recognition, and evaluating the difference between metalinguistic directions and contextual cues.

References: • Bagher Zadeh, A. C. (2020). CMU-MOSEAS: A multimodal language dataset *(EMNLP), (pp. 1801–1812)*. • Ross, E. D. (2000). Affective prosody & the aprosodias. In M.-M. Mesulam, Principles of behavioral and cognitive neurology (pp. 316-331). Oxford University Press.