

Toward Affective Speech-to-Speech Translation

Masato Akagi^(✉)

Japan Advanced Institute of Science and Technology, 1-1 Asahidai,
Nomi, Ishikawa 923-1292, Japan
akagi@jaist.ac.jp

Abstract. Speech-to-speech translation (S2ST) is the process by which a spoken utterance in one language is used to produce a spoken output in another language. The conventional approach to S2ST has focused on processing linguistic information only by directly translating the spoken utterance from the source language to the target language without taking into account para-linguistic and non-linguistic information such as the emotional states at play in the source language. This paper introduces activities of JAIST Acoustic Information Science Laboratory, Human Life Design Area, Japan Advanced Institute of Science and Technology that explore how to deal with para- and non-linguistic information among multiple languages, with a particular focus on speakers' emotional states, in S2ST applications called "affective S2ST." In our efforts to construct an effective system, we discuss (1) how to describe emotions in speech and how to model the perception/production of emotions and (2) the commonality and differences among multiple languages in the proposed model. We then use these discussions as context for (3) an examination of our "affective S2ST" system in operation.