

Use case LOQ1: Generation of emotional speech

Use case LOQ1a: annotation of speech acts

In this example, we want to insert pre-recorded utterances into a corpus based speech synthesis system. Most of these prompts are frequently used expressions that have a pragmatic function such as greetings, thanks, regrets, disapprovals, apologies, compliments, etc. Given their intrinsic nature, these sentences are read in an expressive way, and if appropriately used in the unit selection mechanism, convey an emotional intention in the generated speech.

We have to group these expressions into linguistic categories and describe them in terms of emotional intensity. For instance “Good morning!” could be read in different ways: it could be happy, excited, or even sad. Moreover, given the emotional style, there could be different levels of intensity that could be described quantitatively using a range of values between 0 and 1.

The input text (or part of it) of the synthesis system should be annotated specifying the emotional style as well as the level of activation. The system will look for the pre-recorded expression in the speech database that best fits the annotated text.

Requirements:

- Scope of the emotion annotation: : input text in a Text-To-Speech system, speech database
- emotion description
 - emotion category
 - level of activation
- other: description of the speech act category (greetings, thanks,...)

Use case LOQ1b: annotation of paralinguistic events

We want to generate speech containing paralinguistics events such as laughs, sighs, pains or phenomena like these, in order to strengthen the expressive effect of the generated speech. These elements have to be described in terms of event category and of the emotion which they refer to. Also in this case it could be useful to describe quantitatively the effort of these events (for instance there could be “weak” laughs or “exaggerated” laughs). The speech synthesis engine should appropriately select the best speech event from the database, given an effective annotation for it in the text that has to be synthesized. These events could be inserted at a particular point in the sentence or could be generated following certain criteria.

Requirements:

- Scope of the emotion annotation: input text in a Text-To-Speech system, speech database
- emotion description
 - emotion category
 - level of activation
- other: description of the paralinguistic category (laughs, sighs,...)

Use case LOQ2: Annotation of emotional speech databases

We want to annotate a speech database containing emotional phrases. This material is used to extract prosodic models that will be used to appropriately select acoustic units in a corpus based speech synthesis system. The database consists of short sentences that are recorded from many

speakers that read the scripts simulating certain emotional styles. Actually, each sentence is read in different emotional styles (e.g. sad, happy, angry, etc.) and a neutral style is also considered as the baseline. We want to study the acoustic correlations of the considered emotional styles in order to extract simple rules that account for the variation of some acoustic parameters. To achieve this, we need to annotate the speech data, taking into account the intensity of the relative emotion and the level of valence.

Requirements:

- Scope of the emotion annotation: utterance
- emotion description:
 - emotion category
 - a scale to annotate the emotional dimensions: valence, activation and power