ABSTRACT

Visemes are visual counterpart of phoneme. In speech synthesis, a viseme was used as a small unit to generate an utterance of talking head model. Visemes depend on the pronunciation of each language. It is difficult to apply a viseme in cross language rule. Besides, viseme can be affected by facial muscle that actively contracted by the expression while talking. These result on visemes with very large combinations of forms. This makes the generation of a sequence of utterance for speech animation to use many visemes.

This study proposed the classification of viseme mapping. Visemes used in this study was limited on the consonant-vowel (CV) Indonesian's syllable pattern, which are combined with expression on lower face area. Classification was done using 19 crucial points as parameterized muscle and 1 reference points as the standard on normalization image process. Facial animation model are generated based on the classification result of visemes. The generation process used free form deformation (FFD) as deformation process and Bezier curve as the keyframe references for generating process.

This study grouped 315 combinations of visemes into 26 classes. Based on the distance criteria, facial animation which is generated to show the viseme movement, has achieved 92,3% realistic perception. It means that the viseme group as a result from this study has proven effectively to be used to produce a realistic perception of speech animation model.

Keyword: realistic speech model, *viseme* grouping, co-articulation, expression.