
It has been shown that different vocal-tract shapes could produce acoustic signals with nearly identical values of the first three formant frequencies and amplitudes in the geometrical articulatory model (Atal et al., 1978). Also, we found that different articulatory positions could produce similar formant frequencies and bandwidths in spite of using the actual human articulatory-acoustic data, which was designed from a set of speech consisting of three Japanese vowel sequences (e.g. /auii/). Here, the articulatory positions were measured using an electromagnetic articulographic (EMA) system and the acoustic data consisted of the first three formant frequencies and bandwidths. In many previous studies, reducing redundancy in acoustic-to-articulatory inversion has been performed by using articulatory dynamic constraints. We found, in the actual data, that the redundancy was effectively reduced by using formant data and fundamental frequency (F0) compared to formant data only. In line with this result, we developed a method of estimating articulatory positions from speech acoustics using a stochastic speech production model, in which F0 values were also taken into account. This model could statistically represent dynamical articulatory behavior and non-linear relationship between the articulatory and acoustic data. Result from this method showed that the estimation error of articulatory positions was effectively decreased by using F0 values. In particular, the use of F0 values was efficient for decreasing the estimation error when context-independent model was used.