

A procedure is demonstrated for learning to recover the relative positions of simulated articulators from speech signals generated by articulatory synthesis. The algorithm learns without supervision, that is, it does not require information about which articulator configurations created the acoustic information in the training set. The procedure consists of vector quantizing short time windows of a speech signal, then using multidimensional scaling to represent quantization codes that were temporally close in the encoded speech signal by nearby points in a continuity map. Since temporally close sounds must have been produced by similar articulator configurations, sounds which were produced by similar articulator positions should be represented close to each other in the continuity map. Continuity maps were made from parameters (the first three formant center frequencies) derived from acoustic signals produced by an articulatory synthesizer that could vary the height and degree of fronting of the tongue body. The procedure was evaluated by comparing estimated articulator positions with those used during synthesis. High rank-order correlations (0.95 to 0.99) were found between the estimated and actual articulator positions. Reasonable estimates of relative articulator positions were made using 32 categories of sound and the accuracy improved when more sound categories were used.