Parveen Kumar Lehana / Supervisor: Prof. P. C. Pandey, “Spectral mapping using multivariate polynomial modeling for voice conversion”, *Ph.D.Thesis*, Department of Electrical Engineering, IIT Bombay, Jan. 2013.

ABSTRACT

Voice conversion involves modification of the speech of a source speaker to make it perceptually similar to that of a target speaker, using a mapping derived from speech material spoken by them. The thesis presents a technique for modifying spectral characteristics using a single mapping obtained by multivariate polynomial modeling of the relation between the acoustic spaces of the source and the target speakers. Each parameter for generating the target speech is modeled as a multivariate polynomial function of the parameters of the source speech. The set of these functions is obtained from the time aligned source and target feature vectors. Voice conversion of the source speech signal is carried out by applying the estimated mapping for modification of spectral characteristics along with pitch and time scaling. A pitch-synchronous implementation of harmonic plus noise model (HNM) is used as the analysis-synthesis platform for voice conversion.

Out of the various polynomial models investigated, multivariate quadratic model (MQM) was found to be most suited. The method was evaluated for voice conversion for four speaker pairs (male-male, female-female, male-female, and female-male) using parallel speech data for training. Removal of redundant feature vectors from the training data, using a distance threshold, was found to improve the estimation of transformation functions. Voice conversion using MQM and GMM (with 64 mixture components) using the same training and test data resulted in almost similar output quality and decrease in target-transformed distance, but MQM needed a much shorter computation time for estimation of the mapping. Subjective evaluation showed that combination of spectral modification and pitch scaling resulted in transformed speech having good quality and almost the same identity as the target. Averaged across listeners, the scores for identification of transformed speech as the target were 93% for same-gender conversion and 100% for cross-gender conversion. The proposed voice conversion system needs to be investigated using a larger number of speaker pairs, different types of speech material, and other speech analysis-synthesis platforms.