Automatic computational models of acoustical category features: talking versus singing.
David B. Gerhard (Dept. of Computer Science, University of Regina, 3737 Wascana Parkway, Regina, SK Canada S4S 0A2, david.gerhard@uregina.ca)

The automatic discrimination between acoustical categories has been an increasingly interesting problem in the fields of computer listening, multimedia databases, and music information retrieval. A system is presented which automatically generates classification models, given a set of destination classes and a set of a-priori labeled acoustic events. Computational models are created using comparative probability density estimations. For the specific example presented, the destination classes are talking and singing. Individual feature models are evaluated using two measures: The Kologorov-Smirnov distance measures feature separation, and accuracy is measured using absolute and relative metrics. The system automatically segments the event set into a user-defined number (n) of development subsets, and runs a development cycle for each set, generating n separate systems, each of which is evaluated using the above metrics to improve overall system accuracy and to reduce inherent data skew from any one development subset. Multiple features for the same acoustical categories are then compared for underlying feature overlap using cross-correlation. Advantages of automated computational models include improved system development and testing, shortened development cycle, and automation of common system evaluation tasks. Numerical results are presented relating to the talking/singing classification problem.