

# **An Approach Towards a Heart Beat Sound Information Retrieval System**

Ehsan Safar Khorasani\*, Shyamala Doraisamy, Azreen Azman and Masrah Azmi Murad

Department of Multimedia, Faculty of Computer Science, University of Putra Malaysia, Serdang, Selangor, Malaysia

Interpretation of heart sounds is a problematic and difficult skill that requires cardiology specialists. The diagnosis of heart disease from heart sound can differ between cardiologists and would require more detailed and expensive tests. However, heart disease diagnosis by heart beat is preferable and still widely used as the first step to diagnosis. Computer aided auscultation has emerged as a cost-effective technique to analyze and interpret the heart sounds. Digital Heart sound recordings with background noise, similarity among heart diseases, recording environment conditions, auscultation body points makes detection of heart diseases complicated. There are several methods for automated detection and classification of heart diseases and heart sound analysis that have been proposed. Some of them used Artificial Neural Network method for detection and classification of heart sounds. Another technique that it used for diagnosis the heart problem is Hidden Markov Model (HMM) that they suggest HMM for segmentation of heart sound recorded for clinical and classification purpose.

In this study we propose a feasible technique for developing a heart beat sound retrieval system using text based approaches useful towards automated heart disease detection. The audio format heart sound recordings are pre-processed and transcribed into the MIDI format. The MIDI files are then encoded to text strings using N-grams. These text strings are then indexed and tested for retrieval using both database and Information Retrieval (IR) systems. The Longest common subsequence (LCS) matching algorithm was used for identifying similarities from the database. With IR, full text indexing of the recordings was used and retrieved using known item searches from a search engine. The feasibility of these text based retrieval approaches are shown from retrieval experiments with around 100 digital heart sound recordings and eleven categories of heart diseases. Each recording include eight heart beat cycles and duration of ten seconds.