What Can Speech Researchers Bring to Music Processing?
Prof. Shigeki Sagayama
Graduate School of Information Science and Technology, The University of Tokyo

The speech research community has developed powerful approaches which are potentially applicable to other technological areas. As music is the counterpart of speech in the sense of them being the two most important information-rich categories of acoustic signals understood by humans, music processing can be a good application target of speech technologies. Recently, music technology research has been growing rapidly, fueled by a high demand in music entertainment and a general need for music information retrieval. Since the speaker started working on music processing research in 1998, he has been continuously seeking good models and algorithms for music processing inspired by speech technologies, as well as new solutions to music-specific problems which will hopefully help speech processing in the future.

The speaker will give some examples from his and his colleagues' recent research activities, where speech processing algorithms play an important role in music processing both for audio and symbolic (typically, MIDI) music inputs. Applications of HMMs (Hidden Markov Models), DP (Dynamic Programming) and their generalizations: Dynamic Bayesian Networks (DBNs) include chord and key modulation detection, music transcription, harmonization of given melodies, counterpoint, rhythm recognition, score following, song composition from given lyrics, piano fingering, etc. Applications of Gaussian mixtures and the EM algorithm include multiple F0 estimation, precise onset detection, sound separation in polyphonic music, deletion and modification of notes and reconstruction of missing parts in audio signals, etc. Language modeling approaches are applicable to musicological analysis and harmonization of melodies. Research in music also motivates us to develop music-specific acoustic signal processing methods such as Non-negative Matrix Factorization (NMF) for music transcription, harmonic/percussive sound separation and microphone array techniques for music signal separation.