

**ACOUSTICS2008/1853**  
**Music and speech signal processing using harmonic-temporal clustering**

Jonathan Le Roux<sup>a</sup>, Hirokazu Kameoka<sup>b</sup>, Nobutaka Ono<sup>a</sup>, Alain De Cheveigne<sup>c</sup> and Shigeki Sagayama<sup>a</sup>

<sup>a</sup>University of Tokyo, Sagayama/Ono Laboratory, 7-3-1 Hongo, Bunkyo-ku, 113-8656 Tokyo, Japan

<sup>b</sup>NTT Communication Science Laboratories, NTT Corporation, 3-1 Morinosato wakamiya, 243-0198 Atsugi, Kanagawa, Japan

<sup>c</sup>CNRS, Universite Paris 5, Ecole Normale Supérieure, 29 rue d'Ulm, 75230 Paris, France

We present here the principle of the recently introduced Harmonic-Temporal Clustering (HTC) framework and its applications in both music and speech signal processing. HTC relies on a precise parametric description of the harmonic parts of the power spectrum through constrained Gaussian Mixture Models. The model parameters of all the elements of the acoustical scene are estimated jointly by an unsupervised 2D time-frequency clustering of the observed power density. HTC is effective for multi-pitch analysis of music signals and F0 estimation of single and multiple speaker speech signals in various noisy environments. It also enables to perform extra processing of monaural music and speech signals, such as isolation or cancellation of a particular part, noise reduction and source separation.