ACOUSTICS2008/3233
Single sensor singer/music separation using a source/filter model of the singer voice

J.-L. Durrieu[a], B. David[b] and G. Richard[c]

[a]GET-ENST Télécom-Paristech, 46, rue Barrault, 75013 Paris, France

[b]Télécom ParisTech (ENST) / TSI - CNRS LTCI, 46 rue Barrault, 75634 PARIS Cedex 13, France

[c]ENST / TSI, 46, rue Barrault, 75634 Paris Cedex 13, France

durrieu@enst.fr

Separating the singer voice from polyphonic music signals has many useful applications, such as demixing/remixing, desoloing or audio indexing. In the works of Benaroya on single sensor blind source separation, the signals are modelled by Gaussian mixtures (GMM) such that each state is characterized by a spectral shape. The separation itself is then done by adaptive Wiener filtering. However, to better fit general signals, the number of states for the vocal model should be equal to the number of notes multiplied by the number of vowels (or canonical vocal tract shapes) that the singer uses. Therefore, in order to separate a singer voice from background music, we suggest a source/filter model for the singer signal, keeping the same models as used by Benaroya for the background music signal. Assuming the presence of only one singer, we separate the desired part from the rest by first estimating the sung melody thanks to the source part of our model and then re-evaluating the parameters of our model. This technique does not require a training step and can be generalized to the separation of signals involving a predominant instrument that also follows a source/filter model.