ISSE: AN INTERACTIVE SOURCE SEPARATION EDITOR

NICHOLAS J. BRYAN, GAUTHAM J. MYSORE, AND GE WANG

For many music- and audio-related tasks, it is desirable to separate a single-channel recording of a mixture of sounds (e.g. pop song) into its respective sources (e.g. drums, guitar, vocal, etc.). This process is referred to as single-channel source separation and it is useful for audio denoising, music transcription, music remixing, audio-based forensics, and similar tasks. One of the most promising approaches found for this problem is based on non-negative matrix factorization and its probabilistic latent variable model counterparts (PLVM). These methods model audio signals as a linear combination of prototypical frequency components over time and can achieve high-quality separation results. Unfortunately, however, they also require isolated training data to perform separation and offer no mechanism to correct for unsatisfactory results.

To overcome these issues, we have developed an interactive source separation editor (ISSE). ISSE is a free, open-source (GPL v.3), cross-platform audio editing tool written in C++ that allows a user to separate a single audio recording into its respective sources. The core separation technology is based on our recently proposed interaction paradigm and algorithm [1, 2], which incorporates user-feedback into PLVM-based source separation techniques. The method works by allowing end-users to roughly paint on time-frequency visualizations of sound. We then use the painting annotations as continuous-valued feature labels that are, in turn, used to constrain our PLVM via posterior regularization, update the separation estimates, and iteratively improve the separation quality.

In contrast to many open-source machine learning projects, this methodology embodies a fully interactive machine learning approach. In particular, end-users are employed to both train and deploy a learning algorithm and complete a complicated task that both humans and machines cannot otherwise perform independently. In addition, it is interesting to note that the use of interaction is essential not only because the feature labeling drastically improves performance, but because it helps users to learn how to correctly label the data and understand how the algorithm reacts to the labeling.

For evaluation, we submitted the proposed software to the recent community-based signal separation evaluation campaign and achieved state-of-the-art separation quality for several tasks related to professionally produced music recordings [3]. In addition, a video demonstration and several audio examples are posted online demonstrating the software can be used in many real-world audio and music editing scenarios. To download the application, code, and audio/video examples, please see `http://mloss.org/software/view/507/` or `http://isse.sourceforge.net/`.

[1] N. J. Bryan and G. J. Mysore. Interactive refinement of supervised and semi-supervised sound source separation estimates. In IEEE International Conference on Acoustics, Speech, and Signal Processing, May 2013.

[2] N. J. Bryan and G. J. Mysore. An efficient posterior regularized latent variable model for interactive sound source separation. In International Conference on Machine Learn- ing, June 2013.

[3] SiSEC Audio Committee. Signal separation evaluation campaign. http://sisec.wiki.irisa.fr/tiki-index.php, 2013.