Audio Source Separation via the Sum–of–Sinusoids Model

The sparse nature of musical sounds allow for their succinct representation via a sum-of-sinusoids model [1]. This model is convenient for its relative parsimony while keeping common signal transformations pertinent to computer music easy to perform. Operations like time-stretching, transposition, noise-removal and filtering are simple and efficient to perform on the sinusoid model. In theory, the act of separating multiple superimposed sounds—audio source separation—should also be simple: the desired sinusoidal components are chosen and the rest discarded. The difficult problem lies in choosing this group of sinusoidal components that come from the same source.

Theoretical Background

In recent years techniques estimating the parameters of more descriptive sinusoidal models have been developed that allow for instantaneous frequency– and amplitude–modulations (FM and AM) [2]. In addition, the problem of simultaneously tracking multiple sources has been made tractable using linear programming optimizations [3]. Previously, it had been shown that common FM [4] and AM [5] were plausible data with which to group sinusoids common to one source. The approach in the former paper is perceptual but the analytical perspective of the latter will be adopted here: In general, sources will appear superimposed in a traditional time–frequency representation (such as the short-time Fourier transform) while a plot of their FM and AM parameters will clearly group the sources without superimposition (provided the model is close enough to the true model and the parameters are different enough) [6].

Objectives

For non–stationary signals, a common approach to estimating their signal parameters is to calculate a set of parameters for the signal over a short period of time. Using the set of AM and

FM parameters estimated at one time, we will use the gaussian mixture clustering algorithm to group parameter elements [7]. It remains to group these elements across time. While various methods have been proposed for this, we will try a new method based on a technique for tracking multiple objects across video frames [8]. The result will be groups of parameters, each corresponding to a unique sound source.

Methodology

The thesis will be organized as follows: the proposed signal model will be described and a survey of techniques to estimate its parameters will be presented. The expectation maximization algorithm [9], its relationship to clustering, as well as its application to other audio source separation models will be described [10]. A brief overview of convex optimization relaxation solutions to combinatorial problems will be presented and related to the problem of tracking multiple sources. Finally, two experiments will be carried out. The first will demonstrate the plausibility of source grouping via common modulation by evaluating the clustering algorithm on synthetic sources with known parameters. The second will investigate grouping common sources across time using the modulation parameters in the cost function of a linear program designed for multiple target tracking. Possible improvements will be proposed, as well as a discussion of this strategy's possible integration with another popular source separation algorithm: non-negative matrix factorization [11].

Contribution

This thesis demonstrates the mastering of a variety of musical signal processing techniques and their integration into a larger framework at an attempt to solve the difficult problem of audio source separation with a single sensor. While documenting the knowledge

gained throughout the master's program, the contribution is furthermore applicable to the domains of audio editing, repair and analysis, and the development of new audio effects.

Bibliography

[1] Serra, Xavier, and Julius Smith. "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition." *Computer Music Journal* 14, no. 4 (1990): 12-24.

[2] Hamilton, Brian, Philippe Depalle, and Sylvain Marchand. "Theoretical and practical comparisons of the reassignment method and the derivative method for the estimation of the frequency slope." In *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*, pp. 345-348. IEEE, 2009.

[3] Boyd, Stephen, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[4] McAdams, Stephen. "Segregation of concurrent sounds. I: Effects of frequency modulation coherence." *The Journal of the Acoustical Society of America* 86, no. 6 (1989): 2148-2159.

[5] Li, Yipeng, John Woodruff, and DeLiang Wang. "Monaural musical sound separation based on pitch and common amplitude modulation." *Audio, Speech, and Language Processing, IEEE Transactions on* 17, no. 7 (2009): 1361-1371.

[6] Creager, Elliot. "Musical source separation by coherent frequency modulation cues." 2016. Masters Thesis, McGill University. Accessed April 18, 2016. http://digitool.library.mcgill.ca/thesisfile139201.pdf.

[7] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*. Vol. 1. Springer, Berlin: Springer series in statistics, 2001.

[8] Jiang, Hao, Sidney Fels, and James J. Little. "A linear programming approach for multiple object tracking." In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1-8. IEEE, 2007.

[9] Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the royal statistical society. Series B (methodological)* (1977): 1-38.

[10] Smaragdis, Paris, Bhiksha Raj, and Madhusudana Shashanka. "A probabilistic latent variable model for acoustic modeling." *Advances in models for acoustic processing, NIPS* 148 (2006): 7-1.

[11] Smaragdis, Paris, and Judith C. Brown. "Non-negative matrix factorization for polyphonic music transcription." In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pp. 177-180. IEEE, 2003.