Dynamic time warping for segmentation of temporally expressive Irish traditional music recordings

Hannah Robertson
MUMT 621 — Winter 2012
Project proposal — 20 March

March 20, 2012

In this project, I will attempt to segment audio recordings of traditional Irish tunes with expressive tempo variations. To do this, I plan to incorporate dynamic time warping (DTW) into the self-similarity matrix method for tune segmentation developed by Kelly et al. in 2010.

Discussion

Folk music is by nature social, in that it is passed among 'folks' orally (and aurally!) rather than written down. This means that any symbolic representation of a folk tune is either a general snapshot of the tune's general contours at one instant in time or a complete transcription of one player's unique rendition in a single performance. In either case, the symbolic representation cannot be treated as any sort of golden standard. In order to study folk music effectively, then, it is critical to be able to work with and compare multiple audio recordings of the same tune over time and by different musicians, whether by studying that audio directly or by transcribing portions of it. To aid in human memorization and recall, material within the oral tradition often involves idioms and repetition. For this reason, the ability to recognize reoccurring patterns within and among folk tunes is also critical to their study. In order to study large amounts of folk music, computational methods for similarity recognition come into play.

Like the folk music in many cultures, Irish traditional music has a highly repetitive and highly patterned structure. While skilled players increasingly ornament the repetitions of each tune, a tune's initial melody maintains a strong presence throughout the entire duration of the tune. This is in contrast to music like jazz, for example, which includes improvisational sections that deviate from primary tune. The Audio Research Group at the Dublin Institute of Technology has used self-similarity matrices (also known as recurrence plots) to segment traditional Irish tunes into self-similar parts (Kelly, Gainza, Dorran, and Coyle 2010a), both for tune identification and tune fingerprinting: the ability to play a 'representative sample' of a tune that include each unique segment from the tune but none of the repetitions, thus preserving and presenting a tune's trademark elements in an efficient and compact way. Their methods make use of the fact that Irish traditional music tends to have a steady tempo; by looking at similarity matrices built on the chroma vectors found at beat locations, they are able to identify musical segments within tunes.

Not all traditional Irish music has an even and constant tempo, however. While such a constant tempo is critical for dance-accompanying tune types such as jigs and reels, airs and other slow tunes tend to be performance pieces in their own right, and as such involve significant tempo variation in order to convey emotion. This means that tune segmentation methods based in any way on the identification of a beat, such as the method utilized by Kelly et al., may not perform as well on slower tunes with a less obvious beat.

To counteract this issue, I plan to implement dynamic time warping (DTW) on the chroma vectors within a tune, in an attempt to align similar segments with themselves. DTW is well suited to this task since it requires a one-to-one note mapping but makes allowances for "missing" data, which is perfect for comparing two variously ornamented repetitions of a single tune with one another, since the ornaments present in one tune might be non-existant in the other and vice versa. I hypothesize that incorporating such a step into the workflow already used by Kelly et al. will improve segmentation of temporally expressive Irish traditional music recordings.

Sub-goals

Dates	**Tasks**
March 20 to 27	Study DTW. Write up DTW and similarity matrices for class presentation. Write paper introduction.
March 27 to April 3	Acquire and annotate data set. Use jMIR (or SALAMI?) to create feature vectors (chroma, signal strengths, ...others? TBD).
April 4 to 10	Code like crazy: implement similarity matrix identification, sans DTW.
April 10	Evaluate current project state: is the project moving quickly enough that there is still a possibility of meeting the ISMIR deadline? If so, *drop everything else* and write/submit an abstract by the 13th; bump up all additional steps and be ready to submit on the 18th. If not, carry on with this original schedule!
April 11 to 17	Implement DTW. Fine-tune code. Get results(!). Write draft of results.
April 18 to 25	Panic! Commence this extra week of late-night debugging to account for the inevitable fact that something *will* be broken. Drink too much tea.
April 26 to May 4	Proof-read and submit(!).
May 5	Frolick in the park.

Tentative Bibliography

Antonopoulos, I., A. Pikrakis, S. Theodoridis, O. Cornelis, D. Moelants, and M. Leman. 2007. Music retrieval by rhythmic similarity applied on greek and african traditional music. *Proceedings of the 8th International Society for Music Information Retrieval Conference*.

Cooper, M., and J. Foote. 2002. Automatic music summarization via similarity analysis. *Proc. IRCAM*: 81–5.

de Nooijer, J., F. Wiering, A. Volk, and H. J. Tabachneck-Schijf. 2008. Cognition-based segmentation for music information retrieval systems. *Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08)*.

Duggan, B. 2009. Machine annotation of traditional irish dance music. *PhD Thesis School of Computing Dublin Institute of Technology*.

Duggan, B., M. Gainza, B. O'Shea, and P. Cunningham. 2009. Compensating for expressiveness in queries to a content based music information retrieval system. *Proceedings of International Computer Music Conference (ICMC)*.

Duggan, B., B. O'Shea, M. Gainza, and P. Cunningham. 2009. The annotation of traditional irish dance music using matt2 and tansey. *8th Annual Information Technology & Telecommunication Conference, Galway Mayo Institute of Technology*.

Foote, J. 2000. Automatic audio segmentation using a measure of audio novelty. *IEEE*.

Foote, J., and M. Cooper. 2001. Visualizing musical structure and rhythm via self-similarity. *Proceedings International Conference on Computer Music*.

Izumitani, T., and K. Kashino. 2008. A robust musical audio search method based on diagonal dynamic programming matching of self-similarity matrices. *ISMIR 2008 – Session 5b – Feature Representation*.

Kelly, C., M. Gainza, D. Dorran, , and E. Coyle. 2009. Structural segmentation of irish traditional music using chroma at set accented tone locations. *127th Audio Engineering Society Convention*.

Kelly, C., M. Gainza, D. Dorran, and E. Coyle. 2010a. Audio thumbnail generation of irish traditional music. *ISSC*.

Kelly, C., M. Gainza, D. Dorran, and E. Coyle. 2010b. Locating tune changes and providing a semantic labelling of sets of irish traditional tunes. *International Society for Music Information Retrieval,*.

Martin, B., M. Robine, and P. Hanna. 2009. Musical structure retrieval by aligning self-similarity matrices. *10th International Society for Music Information Retrieval Conference*.

Müller, M., and M. Clausen. 2007. Transposition-invariant self-similarity matrices. *Proceedings of the 8th International Society for Music Information Retrieval Conference*.

Müller, M., and F. Kurth. 2007. Towards structural analysis of audio recordings in the presence of musical variations. *EURASIP Journal on Advances in Signal Processing 7*.

Ness, S. 2009. Content-aware visualizations of audio data in diverse contexts. *University of Victory Thesis*.

van Kranenburg, P. 2010. A computational approach to content-based retrieval of folk song melodies. *PhD thesis, Utrecht University*.

van Kranenburg, P., J. Garbers, A. Volk, F. Wiering, L. Grijp, and R. C. Veltkamp. 2007. Towards integration of music information retrieval and folk song research. *Proceedings of the 8th International Conference on Music Information Retrieval*: 505–8.