myMoodplay: an interactive mood-based music discovery app

Alo Allik, György Fazekas, Mathieu Barthet, Mark Sandler
Centre for Digital Music, Queen Mary University of London
{m.barthet,g.fazekas,a.allik,mark.sandler}@qmul.ac.uk

ABSTRACT

myMoodplay is a web app that allows users to interactively discover music by selecting desired emotions. The application uses the Web Audio API, JavaScript animation for visualisation, linked data formats and affective computing technologies. We explore how artificial intelligence, the Semantic Web and audio synthesis can be combined to provide new personalised online musical experiences. Users can choose degrees of energy and pleasantness to shape the desired musical mood trajectory. Semantic Web technologies have been embedded in the system to query mood coordinates from a triple store using a SPARQL endpoint and to connect to external linked data sources for metadata.

1. INTRODUCTION

A number of technologies have recently been developed in response to the exponential growth of music collections, both for private users and content providers. This development is fundamentally changing music search strategies for both personal and commercial use as the traditional search methods are rendered ineffective by the sheer volume of available data. New search algorithms are typically based on content-based audio feature extraction and focus on similarity measures between tracks in a collection, however, low-level features in general provide insufficient meaning in terms of musical attributes with which most people are familiar. myMoodplay app allows users to explore music in a large collection through an intuitive touch interface in which moods are represented in a 2-dimensional mood space. The played music expresses user selected emotions. By interacting with the control surface, users create a personalised trajectory through the mood space that is visualised with individualised selection indicators. myMoodplay uses crowd-sourced tag statistics to determine the location of each track in the mood space, rather than content-based features to better approximate perceived emotions the music expresses. The client application connects to a dedicated API for data retrieval and music streaming. The system embeds Semantic Web technologies to query mood coordinate values for each audio track from a triple store and music metadata from external linked data sources.

Web Audio Conference WAC-2016, April 4–6, 2016, Atlanta, USA.

2. BACKGROUND

2.1 Related work

Mood driven music navigation systems have been explored in different contexts for various purposes. Content-based feature extraction systems like SmartDJ[1] are based on the concept of building feature similarity networks of songs. SmartDJ uses a set of low-level features and then applies dimension reduction techniques to place the songs into a 2-dimensional space. The users are able to visualise and transition between songs based on tempo, spectral similarity and tonality. A different way to select music is available in affective biofeedback players that use previous listening data to reinforce user experiences. For example, AffectiveDJ[4] and AMP[8] capture listeners' biosignals and derive mood information from skin conductance. Both these systems monitor and store skin conductance data during the listening experience to be able to match users' previous experiences and select music based on listening history on subsequent occasions. An alternative strategy to feature extraction is crowd-sourcing mood tags from social networks and creating track networks based on tag statistics. Mood Cloud 2.0 [9] employs this kind of approach by extracting information from Last.fm tags and aligns these to a 2-dimensional mood space. There are a number of online interfaces to music libraries that enable browsing or selection according to mood, for example Moodstream[1] or Musicovery[2]. However, in these kind of services, the mood annotation method as well as the details of the underlying audio dataset tends to be obscured from users making it difficult to evaluate and compare methodologies.

2.2 Mood Conductor

Mood Conductor [6] is an interactive audience-performer system that facilitates audience feedback during improvised live music performances [5]. In order to "conduct" such performances, Mood Conductor allows audience members to send emotional directions using their mobile devices. Audience indicated emotion coordinates in the arousal-valence (AV) space, a model proposed by Thayer [19] and Russell [13] to characterise core or basic emotions, are aggregated and clustered to create a video projection. This is used by the musicians as guidance as well as visual feedback to the audience. Mood Conductor addresses the problem of investigating and exploring the dynamic emotion variations due to the interaction between artists and audiences in live per-

[1] `http://moodstream.gettyimages.com`
[2] `http://musicovery.com`

formances. It opens a direct communication channel from audience members to performers, providing an experimental framework to study this phenomenon. The system proved highly successful in several public performances delivering both entertainment value and useful insight into the interplay of music and emotions [7, 11].

2.3 Moodplay

The development of the myMoodplay app follows from the interactive Moodplay system, a collaborative jukebox that enables users to collectively select music based on mood [2]. Moodplay uses an extended version of the distributed client-server architecture originally developed for Mood Conductor, with several new components created particularly for this installation, including a triple store, a large annotated music dataset, an audio player and a lighting client.

The Moodplay interactive installation was exhibited at two public events: the Digital Shoreditch (DS) Festival[3] and the BBC R&D "Sound: Now and Next" (BBCSNN) conference[4] both held in London in May 2015. We conducted a user survey during these events in order to assess the system in real world scenarios with a large number of users from different backgrounds. Visitors from these events who came to the Moodplay installation were invited to interact with the system. Consenting visitors were given brief instructions on how to use the system which they could then explore and experience without timing or task-based constraints. After their trial of the system visitors were invited to take part in our survey. In total 126 participants completed the survey (113 from DS and 13 from BBCSNN, 47.6% female, 52.4% male).

One of the main goals of the user survey was to encourage participants - who came from a wide range of professional backgrounds and age groups - to suggest ideas for future applications and improvements to the prototype system. Two of the questions were designed to address these concerns: *What would you use the system for?* and *How would you improve the system?* Qualitative analysis of recurrent themes and underlying patterns of the responses revealed personalisation to be one of the most prominent themes. There were a number of requests for a personal player ("be able to use outside exhibition", "have the app on a headset") or a music streaming service ("link to Spotify or other music service (Last.fm) playlists of the user and vote cast are based on their library") or to live streams ("allow for importing live streams"). 13 participants suggested to add functionalities for user preference and history monitoring ("upvoting", "ability to save playlist i created"), user customisation ("have an individual base level song to calibrate system","tailoring the music to suit the user?",'ability to configure?",'select genres"). Another theme the current development is striving to address is music identification. Some participants wished to be able to identify the song being played ("show what song is playing","indicate song names / artists"). The myMoodplay app is being developed based on these suggestions incorporating user responses into system requirements.

2.4 Music Mood Annotation

myMoodplay uses a database of audio recordings exceeding 10.000 tracks (see Section 3.1) automatically annotated with mood ratings in the two dimensional arousal-valence space. Manually annotating songs on this scale is not feasible, however, several automatic and semi-automatic approaches have been proposed to solve this problem. These solutions range from games-with-a-purpose [20] and crowd-sourcing, to the estimation of mood labels using machine learning models trained on small manually labelled datasets. Typically, these models learn the association between audio features and mood labels. A prime example of the former approach is MoodSwings [17], a collaborative online game that leverages crowd-sourcing to collect AV annotations directly. Audio-based approaches include auto tagging that rely on early categorical models of emotions. For instance, [10] used an audio database with manually labelled adjectives belonging to one of 13 categories and trained Support Vector Machines (SVM) on timbre, rhythmic and pitch features. However, categorical representations of emotions have been criticised for their numerous restrictions [12, 3]. For instance, the discretisation of moods into a set of "landmarks" prevents the use of emotions which differ from these landmarks. To solve this issue, music emotion labelling may be formulated as a regression problem [21] to map high-dimensional features extracted from audio directly to the AV space.

Audio-based mood annotation is still challenging however and far from being a fully solved problem [16]. This is especially true for the valence dimension, where typical performance is low ($R^2 < 0.5$) compared to human labelling. Crowd-sourcing AV annotation for a large dataset using online games or Mechanical Turk is also problematic, due to the long time it takes and the potential cost it may involve. For these reasons, in the present work we opt for a third approach which relies of folksonomy or crowd-sourced tag data collected from an online radio station Last.fm. We then map tagged tracks to the AV space using the approach proposed in [15] and validated in [14]. This process yields the mood annotations for the dataset described in Section 3.1.

3. MYMOODPLAY WEB APP

3.1 Audio dataset

In order to create a new musical experience, myMoodplay relies on large dataset called ILM10K consisting of samples from 10,199 commercial tracks from iLikeMusic[5] (ILM). To create this database, first social tag data was collected from Last.fm for over a million tracks as in [15]. We found 218,032 tracks in the ILM database that matched one of the Last.fm tracks using string matching between the artist names and song titles and a duration difference below 0.5s. We then applied a two stage sampling method, removing duplicates and fulfilling several potentially conflicting criteria, to arrive to the 10,199 tracks used in myMoodplay.

To provide a unique musical experience, the aim was to create a diverse dataset that includes many artists, albums and genres and provides a good coverage of the mood space. To this end, we sort songs into 16 genre buckets representing broad genre categories such as rock and jazz, obtained by applying hierarchical clustering to expert annotations available from ILM. We also assign each track to one from a list of unique artists available in the collection and track the number of songs sampled from each single artist. We then assign

[3] http://digitalshoreditch.com/

[4] http://www.bbc.co.uk/rd/events/sound2015

[5] www.ilikemusic.com

artists to multiple genre buckets and the genre categories are ordered by the number of unique artists associated with them. To provide a good coverage of moods, we apply Multi Dimensional Scaling (MDS) to the mood tags and obtain a coordinate in a 3-dimensional space for each track. A Gaussian Mixture Model (GMM) with 5 components is fitted on this data. The likelihood under the model is then used to sort tracks in the respective genre and artist buckets before sampling. During iterative sampling, we first choose a genre category using a pseudo random process that favours genres with fewer artists associated with them. Then we randomly pick an artist from this genre bucket and choose the least likely song from the selected artist under the GMM model fitted on the MDS space. This ensures entropy in the resulting sample is maximised without having to directly estimate it. Selected songs are removed from the pool and the process is repeated until a desired number of tracks is reached or no more songs can be selected given the criteria related to the maximum number of songs chosen from the same artist, album or genre. We experimented with different values for genre and artist thresholds, as well as the number of components for the GMM and evaluated different collections by looking at maximising the number of unique artists as well as the coverage of the mood space. The selected collection includes tracks from over 5600 unique artists. Ten of 16 genre categories are equally well represented in the database (about 7% each) including pop, rock and jazz, while classical and latin are somewhat underrepresented. The database was then automatically annotated with arousal and valence mood ratings using the Affective Circumplex Transformation (ACT) proposed in [15]. This method ensures that coordinates obtained by applying MDS to crowd-sourced tags conform to locations of mood adjectives validated in human psychological experiments.

3.2 Web-based user interface

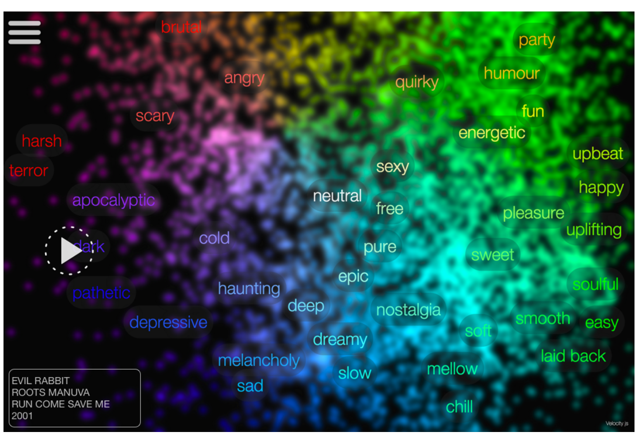

Figure 1: The myMoodplay app user interface with mood tags directing users in the 2-dimensional valence-arousal space.

The myMoodplay app is implemented as a JavaScript client application. It uses the Web Audio API[6] for streaming and cross-fading between tracks and enhances the user interface with audio analysis based animation. The interactive visualisation is implemented in the Velocity.js[7] animation framework.

The interface allows users to select points in the mood space. Once a location has been selected, the application sends a request to the Moodplay server for the nearest track information. The response contains an audio file URL and a MusicBrainz universal identifier. The system then simultaneously starts streaming the audio while querying the MusicBrainz server for metadata of the selected track, including name of the track, artist, album, release year, and duration. Tags collated from Last.FM tag statistics for the tracks in the ILM10K dataset are displayed in the mood space as semantic indicators for the valence-arousal coordinates. Another visual cue is provided by the distribution of tracks by mood coordinates in the background, that indicates the density of tracks in different areas. As the user explores the mood space, each time a new location is selected, a corresponding track is retrieved from the triple store and the audio is cross-faded with the track that was streamed previously. User selection trajectory is visualised by a circular multi-layered cursor that is stretched between selection points for the duration of the cross-fade. Web Audio API is also used to enhance audiovisual interactivity by animating the circular coordinate indicator. The audio context analyser waveform and frequency domain data is used to control animation parameters of the cursor. myMoodplay system keeps track of all the tracks played during a session. Users can also compile their personal playlist by adding tracks to it as they interact with the interface. There is functionality to download personal playlists in M3U or N3 format. The latter provides data in the terminology defined in the Music Ontology.

3.3 myMoodplay API Server

The myMoodplay server uses the Python Web application server module CherryPy[8]. It provides APIs for myMoodplay client applications for interface initialisation, nearest track search and audio streaming. It implements a query engine similar to that of the standard for relational databases where stored procedures and views are incorporated into the server file system. Each operation necessary from the client perspective is defined as a web service method to respond with a specific data structure. There are dedicated web service methods that access the Moodplay SPARQL endpoint to retrieve the nearest neighbour coordinates to user selections, MusicBrainz identifiers for querying track metadata, and audio file URLs for audio streaming. There is also a method to access mood tags for initialising the user interface. SPARQL queries are stored in the server file system and implemented so as to allow parameter definition and on-demand substitution on execution. An example SPARQL stored query is illustrated in Listing 1 where floating point values for valence and arousal parameters have placeholders denoted with an initial '@' that facilitates substitution each time a user selects a point in the 2-dimensional mood space.

3.4 Semantic data modelling and retrieval

The ACT coordinate data from different experimental configurations and tagging information from Last.fm is linked to internal ILM identifiers and audio file paths using a light-weight OWL ontology. OWL is a language for

[6] `https://webaudio.github.io/web-audio-api/`

[7] `http://julian.com/research/velocity/`

[8] `http://www.cherrypy.org/`

```
prefix mood: <http://isophonics.net/content/mood-play/>
prefix mo: <http://purl.org/ontology/mo/>
SELECT ?valence ?arousal ?mbid ?path
WHERE
{
        SELECT ?valence ?arousal ?mbid ?path
                ((ABS(@x-?valence)+ABS(@y-?arousal)) as ?diff)
        WHERE {
        ?coords mood:valence ?valence ;
                mood:arousal ?arousal ;
                mood:configuration mood:actfold4 .
        ?lfmid mood:coordinates ?coords ;
                mo:musicbrainz_guid ?mbid .
        ?trackid mood:filename ?path .
        } ORDER BY ?diff
} LIMIT 1
```

Listing 1: SPARQL query to retrieve ACT coordinates, audio file URL and MusicBrainz identifier for the nearest track to user selected coordinates.

knowledge representation and ontology authoring on the Semantic Web. The structure, as illustrated in Figure 2, relies on concepts defined in the Music Ontology and the Modular Unified Tagging Ontology (MUTO)[9]. The Music Ontology[10] uses the RDF/OWL framework to describe properties and concepts related to music, particularly metadata for published musical works. MUTO is a tag ontology which unifies core concepts of various ontologies on tagging and folksonomies in one consistent schema. The mood namespace extends the Track class in the Music Ontology by adding an object property for coordinates, which is necessary to associate a set of alternative mood reference configurations of the ACT coordinate data to each track. AudioFile class is extended as well with a property for storing an audio file URL for each track in the dataset. Last.fm tagging data is linked in the triple store by extending the MUTO ontology to facilitate expression of tag statistics for each track by individual tags as well as by tag types for mood and genre categories. An example of a track in the dataset expressed in Turtle syntax[11] can be seen in Listing 2. The user selected valence-arousal values are received as parameters by the API and substituted in the stored SPARQL query shown in Listing 1.

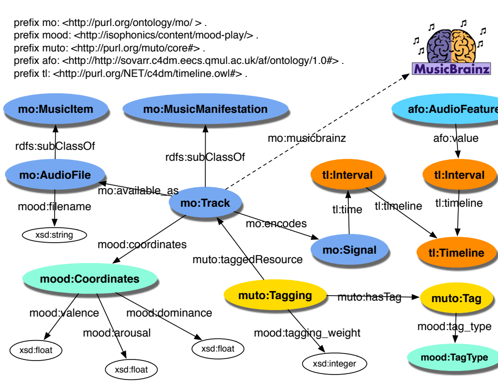

Figure 2: Moodplay uses structured data built on the Music Ontology and the MUTO tag ontology.

[9] http://muto.socialtagging.org/core/v1.html

[10] http://musicontology.com/

[11] http://www.w3.org/TR/turtle/

```
@prefix mood: <http://isophonics.net/content/mood-play/>
@prefix mo: <http://purl.org/ontology/mo/> .
@prefix muto: <http://purl.org/muto/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

mood:lfm13628 a mo:Track ;
        mo:available_as mood:af13628 ;
        mo:musicbrainz_guid "...f80ed9131450" ;
        mood:coordinates mood:actfold4-13628 ;
        mood:has_taggings mood:tagging_lfm13628_tag_darkgroove,
                mood:tagging_lfm13628_tag_spooky .

mood:af13628 a mo:AudioFile ;
        mood:filename "62400-14.01.wav" ;

mood:actfold4-13628 a mood:Coordinates ;
        mood:arousal -0.744659578346668 ;
        mood:configuration mood:actfold4 ;
        mood:valence -0.896364633037682 .

mood:actfold4 a mood:Configuration ;
        mood:reference "actfold4" .

mood:tagging_lfm13628_tag_darkgroove a mood:TrackTagging ;
        mood:tag mood:tag_alternative_dark_groove ;
        mood:tagging_weight 18 .

mood:tagging_lfm13628_tag_spooky a mood:TrackTagging ;
        mood:tag mood:tag_spooky ;
        mood:tagging_weigh 18 .

mood:tag_spooky a mood:LastFMTag ;
        rdfs:label "spooky" ;
        mood:total_count 775 .

mood:tag_darkgroove a mood:LastFMTag ;
        rdfs:label "dark groove" ;
        mood:total_count 56 .
```

Listing 2: Track and corresponding audio file representation in Turtle syntax.

The myMoodplay app connects to the MusicBrainz web services for track metadata. MusicBrainz[12] is an open and public encyclopedia for music metadata that allows anyone to access or contribute data in a structured, reliable format. The MusicBrainz search server exposes an Apache Lucene service that handles metadata queries and returns results in XML and JSON formats. The web client accesses track metadata from the Recording search interface at `http://musicbrainz.org/ws/2/recording/`specifying the *rid* parameter as the MusicBrainz identifier in the query URI: `?query=rid:55743cb9-46c0-4d45-96f4-f80ed9131450\&fmt=json`

4. CONCLUSIONS AND FUTURE WORK

The myMoodplay app was developed in response to suggestions from users who interacted with the prototype Moodplay system at public events and filled in a survey about their experience. The users were asked how the system could be improved and what kind of applications to the system they could imagine. The web-based application combines affective computing, the Web Audio API and Semantic Web technologies to create a new way to personalise music selection in large music datasets online. It is designed with available web technologies based on user suggestions taking into account the most recurrent themes from the survey responses: personalisation, visualisation, and iden-

[12] http://musicbrainz.org/

tification. Users can create playlists by selecting sequences of tracks in the mood space and adding the desired ones to their personal preferences, which can be downloaded to their devices. The system also keeps track of the entire mood browsing history which users can review during the session. The myMoodplay app provides a new way of browsing large music collections by selecting moods that the music represents. The survey results imply that most of the time the music played approximates the represented mood quite accurately, even though moods and musical preferences are rather individual and subjective. This relative success could perhaps be attributed to the crowd-sourced mood tags that enable more accurate reflection of the general consensus on the relationship between a musical work and the mood it represents. Building the system on Semantic Web technologies enables connections to external linked data sources that enables enriched metadata queries beyond the basic information about the particular audio track. Ongoing and future developments of myMoodplay are envisioned to include enhancing audio mixing when transitioning between tracks. This will be achieved by incorporating the semantics and functionality of dynamic music objects [18] into the system and extracting relevant content-based audio features for all tracks in the dataset. This will enable the selection process to search for the most suitable track for transitioning from a pool of nearest neighbor tracks according to content-based audio features like tempo, key and spectral similarity.

5. ACKNOWLEDGMENTS

This work is supported by the "Fusing Semantic and Audio Technologies for Intelligent Music Production and Consumption" (FAST-IMPACt) project (EP/L019981/1).

6. REFERENCES

[1] M. S. Aw, C. S. Lim, and A. W. Khong. Smartdj: An interactive music player for music discovery by similarity comparison. In *Signal and Information Processing Association (APSIPA'13)*. IEEE, 2013.

[2] M. Barthet, G. Fazekas, A. Allik, and M. Sandler. Moodplay: an interactive mood-based musical experience. In *Audio Mostly (AM'15) October 7-9, 2015, Thessaloniki, Greece.*, 2015.

[3] R. Cowie, G. McKeown, and E. Douglas-Cowie. Tracing emotion: an overview. *International Journal of Synthetic Emotions, Special Issue on Benefits and Limitations of Continuous Representations of Emotions.*

[4] F. Dabek, J. Healey, and R. Picard. A new affect-perceiving interface and its application to personalized music selection. In *Workshop on Perceptual User Interfaces*, 1998.

[5] G. Fazekas, M. Barthet, and M. Sandler. Mood conductor: Emotion-driven interactive music performance. In *Affective Computing and Intelligent Interaction (ACII'13), 2-5 September, Geneva, Switzerland*, 2013.

[6] G. Fazekas, M. Barthet, and M. Sandler. The Mood Conductor System: Audience and Performer Interaction using Mobile Technology and Emotion Cues. In *10th International Symposium on Computer Music Multidisciplinary Research (CMMR'13), 15-18 October, Marseille, France.*, 2013.

[7] G. Fazekas, M. Barthet, and M. Sandler. *Novel Methods in Facilitating Audience and Performer Interaction using the Mood Conductor Framework*, volume 8905 of *Lecture Notes In Computer Science (LNCS)*. Springer-Verlag, Heidelberg, Germany., Sound Music and Motion edition, 2014.

[8] J. H. Janssen, E. L. van den Broek, and J. H. Westerink. Tune in to your emotions: a robust personalized affective music player. *User Modeling and User-Adapted Interaction*, 22(3):255–279, 2012.

[9] C. Laurier, M. Sordo, and P. Herrera. Mood cloud 2.0: Music mood browsing based on social networks. In *10th International Society for Music Information Conference (ISMIR), Kobe, Japan*, 2009.

[10] T. Li and M. Ogihara. Detecting emotion in music. *4th International Society of Music Information Retrieval Conference*, pages 239–240, 2003.

[11] T. Lou, M. Barthet, G. Fazekas, and M. Sandler. Evaluation and Improvement of the Mood Conductor Interactive System. *53rd AES Internationa Conference on Semantic Audio, 26-29 January, London, UK*, 2014.

[12] M. Mortillaro, B. Meuleman, and R. Scherer. Advocating a componential appraisal model to guide emotion recognition. *Journal of Synthetic Emotions, Spec. Issue on Benefits and Limitations of Continuous Repr. of Emotions*, 2012.

[13] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[14] P. Saari, M. Barthet, G. Fazekas, T. Eerola, and M. Sandler. Semantic models of musical mood: Comparison between crowd-sourced and curated editorial tags. In *International Conference on Multimedia & Expo, 15-19 July 2013, San Jose, CA, USA*, 2013.

[15] P. Saari and T. Eerola. Semantic computing of moods based on tags in social media of music. *IEEE Transactions on Knowledge and Data Engineering*, 26(10):2548–2560, 2014.

[16] P. Saari, T. Eerola, G. Fazekas, and M. Sandler. Using Semantic Layer Projection for Enhancing Music Mood Prediction With Audio Features. In *Sound and Music Computing Conference, Stockholm, Sweden*, 2013.

[17] E. M. Schmidt and Y. E. Kim. Modeling musical emotion dynamics with conditional random fields. In *12th International Society for Music Information Retrieval Conference, Miami, Florida, USA*, 2011.

[18] F. Thalmann, A. Perez Carillo, G. Fazekas, G. A. Wiggins, and M. Sandler. The mobile audio ontology: Experiencing dynamic music objects on mobile devices. In *Tenth IEEE International Conference on Semantic Computing*, Laguna Hills, CA, 2016.

[19] R. E. Thayer. *The Biopsychology of Mood and Arousal.* Oxford University Press, New York, USA, 1989.

[20] L. von Ahn and L. Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(57), 2008.

[21] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H.-H. Chen. A regression approach to music emotion recognition. *IEEE Trans. on Audio, Speech, and Language Proc.*, 16(2):448–457, 2008.