

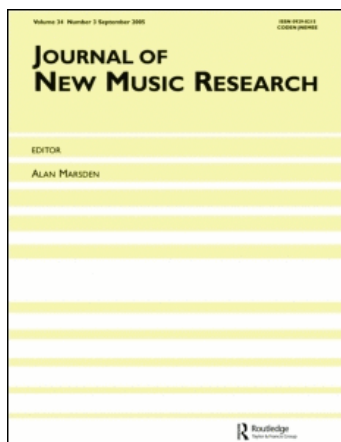
This article was downloaded by: [Ingenta Content Distribution - Routledge]

On: 10 March 2009

Access details: Access Details: [subscription number 791963552]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of New Music Research

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713817838>

### Introduction-From Genres to Tags: A Little Epistemology of Music Information Retrieval Research

Jean-Julien Aucouturier <sup>a</sup>; Elias Pampalk <sup>b</sup>

<sup>a</sup> Temple University Japan, Tokyo, Japan <sup>b</sup> Last.fm, London, UK

Online Publication Date: 01 June 2008

**To cite this Article** Aucouturier, Jean-Julien and Pampalk, Elias(2008)'Introduction-From Genres to Tags: A Little Epistemology of Music Information Retrieval Research',Journal of New Music Research,37:2,87 — 92

**To link to this Article:** DOI: 10.1080/09298210802479318

**URL:** <http://dx.doi.org/10.1080/09298210802479318>

## PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

---

## Introduction – From Genres to Tags: A Little Epistemology of Music Information Retrieval Research

---

**1996:** Possibly the first research publication trying to access musical content in a database, based on algorithms that analyse the audio content rather than on editorial information (Wold and Blum, 1996; see also Kassler, 1966; Ghias et al., 1995).

**2000:** First ISMIR conference – establishes music information retrieval (MIR) as a research agenda, and automatic genre classification as a flagship application.

**2008:** The vision of MIR seems to have become a reality: people routinely connect to online resources where they browse and search for music based on content information.

**Surprise:** This didn't quite happen as expected. Most of the information is annotated manually (no automated analysis), unstructured (no taxonomy), in a collaborative, dynamical and unmoderated process (unlike a centralized library). Millions of users routinely connect to web-sites such as last.fm, MusicStrands, MusicBrainz or Pandora, where they enter free descriptions (aka *tags*) of the music they like or dislike. Each user's tags are available for all to see and influence the way other users describe or look for music. The result is a collaborative repository of musical knowledge of a size and richness unheard of so far. "The Beatles" used to be "British pop". What they are now is something akin to Figure 1.

We have entered the era of collaborative tagging, folksonomies and social networks. This special issue proposes to look at the implications of this new context for Music Information Retrieval research. The issue is composed of technical contributions in the field of signal processing, pattern recognition and artificial intelligence, which report on experiments based on real-world tagging data. Thanks to the dedication of the authors invited here and the editorial staff at JNMR, we achieved an unusually short publication cycle: some of the research presented here is only a few months' old. All the papers in this issue analyse data and reflect trends that can still be found online as you read this.

Nearly all the papers, that is. For the one opening this special issue is ten years old, and it is an interesting story to tell. The paper is an extended version of "*Scanning the dial: An exploration of factors in the identification of musical style*", by David Perrott (currently Jury Consultant at Trialgraphix-Kroll Ontrack, New York) and Robert O. Gjerdingen (currently Professor of Music in Northwestern University). This paper never existed in print: it was presented in 1999 at the annual meeting of the Society for Music Perception and Cognition (SMPC), in Northwestern University (Evanston, IL) on 14–17 August, and there weren't any associated proceedings.

In their SMPC presentation, the authors reported on a psychological study of the human ability to categorize the genre of musical signals. Participants were found to exceed chance by far even on very short extracts (as short as 250 ms), as if they were very rapidly "scanning the [radio] dial" to identify a musical channel they like. This notably demonstrates the near immediacy of genre identification as a response to a musical stimulus, and questions the common assumption that we identify genre as a construct of several "lower-level" component features, such as rhythm or melody.

This study on "Scanning the dial" is well known in the MIR community. It has been referred to very frequently as one of the few experiments measuring human ability to classify musical genre, thus providing a ground-truth to compare automatic genre classification algorithms with. Out of the 24 papers that appeared in the ISMIR conferences from 2000–2007 and included the keyword "genre" in their title, 10 quote "Scanning the dial". A Google search reveals more than 100 conference and journal articles including it as a reference. The first author of these words did quote it too himself, in some of his papers.

Yet, until today, it wasn't possible to find a written account of this research, neither on the internet nor in the library. The only way for referencing authors to have had a direct encounter with these results is either by attending the 1999 SMPC meeting, or by personal communication with the authors. There is little value in tracking down the plausibility of both options. A number of authors did indeed contact Gjerdingen and Perrott directly, as revealed

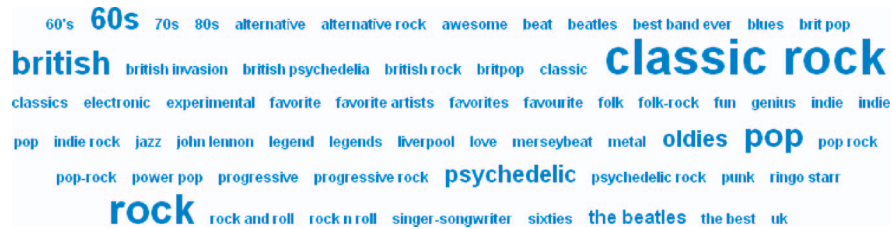


Fig. 1. Most popular tags used to describe the artist “The Beatles” in the tagging community of the website last.fm. Data retrieved from <http://www.last.fm> in August 2008. Larger fonts indicate more popular tags.

by personal communication, but it appears that a probable majority of the works referencing the “scanning” study did so without ever possibly accessing it directly. In amusing support of this, many of the references are plagued with the same recurring typographical errors in the authors’ names. As shown in Table 1, only 20% of the referencing ISMIR papers use the correct spelling. Googling “Gjerdigen (sic) scanning the dial” yields 69 hits against 37 with correct spelling. Readers please note, for good, that “Gjerdigen” takes a “n” in the middle and “Perrott”, a double “t”. Note also that, in the extended version published here, the order of the authors have been reversed and the original subtitle modified.

Referencing work that one has not explicitly seen is generally acknowledged as bad scholarly practice. As the interested reader can now ascertain, there are many subtleties of the study as published in this issue that were never taken into account by subsequent works referencing it, at the risk of using it for erroneous conclusions. However, the goal of these words is neither to pontificate nor to self-flagellate. We are also found culprit here. In fact, the typographical error on “Perrot” (no double “t”) can be traced back as far as we could see – and we did try! – to a review on genre classification published in this journal in 2003 . . . by the first author (Aucouturier and Pachet, 2003)! More appropriately, this informal study gives us data to reflect on the nature of our work, how we do research, how we build intuition, and how our work navigates between empirical data, computational modeling and societal use of our object of investigation: music.

With a little research, references can be traced back to a 1999 PhD dissertation by Eric D. Scheirer, advised by Barry Vercoe at MIT Media Lab (Scheirer, 2000). Scheirer’s work has been also very influential in the MIR community for various reasons: he designed one of the first real-time beat-tracking algorithms (a topic which is now the subject of an international contest as part of the Music Information Retrieval Evaluation eXchange (MIREX)) and arguably pioneered the use of low-level signal features to do audio classification, e.g. discrimination of speech and musical signals. We reproduce here the original citation from the dissertation:

“A preliminary report on “scanning the dial” behavior and its implications regarding the use of music to mediate social

Table 1. Number of papers quoting the “Scanning the dial” study with typographical errors on the authors’ names, in the ISMIR conferences 2000–2007 (among all 24 papers with keyword “genre” in their title, out of total 609 documents available online) and in the works indexed by Google (retrieved with the query: *[name as printed]* + “scanning the dial”).

Name as printed	ISMIR	Google
Perrott	2 (20%)	10 (9.7%)
Perrot (sic)	8 (80%)	93 (90.3%)
Total	10	103
Gjerdigen	3 (30%)	37(35%)
Gjerdigen (sic)	7 (70%)	69 (65%)
Total	10	106

relationships was recently presented by Perrott and Gjerdigen [sic] (1999). They found that college students were able to accurately judge the genre of a piece of music (about 50% correct in a ten-way forced choice paradigm) after listening to only 250-ms samples. This remarkable result forces us to confront the issue of musical surface directly. The kind of musical information that is available after only 250 ms is quite different than the kind of information that is treated in the traditional sort of music-psychology experiment (notes, chords, and melodies). 250 ms is often shorter than a single note in many genres of music; therefore, listeners must be making this decision with information that is essentially static with regard to the musical score (although certainly not stationary in the acoustic signal)” (Scheirer, 2000, pp. 61–62).

This citation is interesting because it illustrates one of the arguments often invoked when referencing Gjerdigen and Perrott’s (G&P) study: the fact that genre can be identified (at least significantly better than random) using a very short extract suggests that identification does not rely on long-term music-theoretical constructs. This argument is the main conclusion proposed by G&P themselves in their paper. Scheirer uses this finding to indicate the technical possibility of using short-time signal features (typically FFT-based) as a basis for audio pattern recognition. Since then, this has become a standard methodology, which makes this seminal argument a possible founding act of MIR. Scheirer’s choice of the

word “musical surface” is interesting, as it doesn’t appear in the original argumentation of G&P and it is not a standard musicological term. This makes it easy to trace down in subsequent referencing works. The word “surface” appears 20 times in the 106 works indexed by Google, in contexts illustrated here:

“[...] judge genre using only the musical *surface* without constructing any higher level theoretic descriptions [...]”  
 “[...] by using only an immediately accessible *surface* [...]”  
 “[...] using only music *surface* features [...]”

Besides the recurring argument of “surface” versus “higher level”, one can identify several ways authors have interpreted the “scanning the dial” study:

**To justify technical constructs:** For instance, the size of the audio segments used to train a classifier (“there must be a sufficient amount of information in very short segments”), or the choice of features to be incorporated in the information flow: “style identification does not require beat estimation and chord recognition”, “style is found in vertical structure as well (i.e. the harmonic relationship of notes)”

**To document human performance:** Unsurprisingly, the actual quantitative results of the original study suffer from statistical spread after multiple indirect references. Human performance with 250 ms extracts is reported in at least three variants: 40%, 50% and 53%, while the paper as published here reports on 54% performance for instrumental data, averaged over all tasks with stimulus size between 250 and 500 ms. Performance at 3 s length is reported as 70% (which is the figure given by G&P) and 72%. Some authors also report 72% precision at 300 ms, in probable confusion with 3000 ms.

**As a ground truth to validate technical work:** in this case, the experimental conditions at different stimulus lengths are often eluded, and the unique figure of 70% precision is used. Automatic performance is often compared favourably: “clear that automatic performance is not far away from humans”, “technique is better than the 70% accuracy reported for humans”, “one can clearly say that the precision achieved here is satisfying high” ... Many authors point out that there are problems with using this data as a ground-truth, notably the fact that humans and algorithms are not tested on the same database, that the human database was undocumented until now and that the human study was not intended as a test of inter-participant consensus but rather an investigation of temporal factors.

**As indicative of the nature of musical genre and listening patterns:** e.g. “genre is not decided by the type of music but rather by market pressure”, “the typical time horizon for a human to identify a genre is from 250 ms to a few seconds”.

Contrary to the “surface” argument, it is interesting that many of these interpretations are not suggested by G&P, and in many cases, seem to only bear distant relation with the original study. The results obtained for a specific set of genres are often generalized to broader aspects of music perception (“just a few short extracts are enough to summarize rhythm”). Moreover, the same quantitative results are often used towards contradictory conclusions depending on the author referencing them. 70% performance is taken to indicate that “ordinary listeners are remarkably adept at classifying music” and “humans are remarkably good”, just as well as showing “the fuzzy nature of musical genre boundaries” and that “the problem is challenging even for humans”.

This brings up several issues. First, it appears that our scientific activity is very much a social one, and that we take information from a social network, and not simply from our own scholarship. The significance of the “scanning the dial” study is indistinguishable from the way it was *used* by the seminal works of our community in the years 2000–2001. In many aspects, the original claims have vanished under their subsequent adaptations of greater appeal: for the reader of the paper as published in this issue, it will likely be a surprise that P&G didn’t use the word “surface” at all, nor conclude on the absolute level of competence of the participants.

Second, this questions the scientific validity of the way we use references in our scholarly communication. Classical epistemology since Descartes and Bacon has insisted that scientific practice relies on the formal knowledge of logics, and we like to consider ourselves as building logically-valid constructs based on evidence previously accumulated and tested. How our community has used this particular study bears little resemblance to this process, but rather to the metaphorical constructs pervasive in natural languages. In *Metaphors We Live By*, Lakoff and Johnson (1980) show how expressions like “the deficit is soaring” or “his income fell” are instances of the generic schema that “more is up”. It is easy to see why “more is up” is a better metaphor than “more is down”, but one still has to learn which of the many reasonable metaphors are actually used within a culture. Such metaphors are easily “motivated” but impossible to “predict”. In many ways, the various interpretations made by authors referencing the “scanning the dial” study are also impossible to *predict* but easily *motivated*. Each individual interpretation (that, say, 70% should indicate poor performance) finds its place in the author’s own local logics, revealing the subjective importance and significance that the “scanning the dial” study had on her.

That this should take place in a scientific process is a disturbing thought. However, many distinguished scientists, including Nobel prize winner Santiago Ramón y Cajal, have noted that the most brilliant discoveries often rely less than expected on formal logics than on this

“acute inner logic that generates ideas” (Ramón y Cajal, 1999) (also quoted in Thagard, 2005).

G&P have a recurring phrase in their paper: “**the customer is always right**”. If someone insists that what “The Beatles” do is “Baroque” music, there is no right to question this statement. It is a datum, subjectively correct regardless of what anyone else says. The target of MIR research is a moving target. Our goal is to design algorithms and systems that can simulate and assist human judgements that are inherently subjective, dynamic, contextualized in a society and a personal history, motivated rather than predicted. While it was possible to ignore this in the early attempts to categorize musical genre using simplifying assumptions, this problem appears in very crude light in the context of collaborative tagging.

**Is the tagger “always right”?** In the data retrieved from last.fm by Paul Lamere in his article appearing in this issue, “brutal death metal” is the most common tag applied to “Paris Hilton”. The music of Paris Hilton bears timbral similarity with such artists as “Madonna” and is generally classified as “pop music” by music vendors. Yet, describing it as “brutal death metal” carries a lot of meaning: it is derogatory of musical quality or sophistication, uses obvious irony to indicate certain musical properties (it is “light” and “cheesy” rather than “hard” and “metal”) and also indicates socio-cultural contempt at what is judged to be too commercial or opportunistic. In many ways, the tagger (all the more so since these are – as Lamere notes – 558 different taggers) is very much right, here. This tag conveys far more information than the arguably “correct” one (“pop”). Sadly, being able to mine this information and not simply discarding it as a statistical outlier is a problem probably as vast as the whole of Artificial Intelligence.

**Is the scholarly reference “always right”?** MIR is a field in the making: its problems, its methodologies, its goals are changing as the discipline adapts to new contexts, new types of data, new challenges. Our practice encompasses following contradictory intuitions and frequently questioning the assumptions that were made in the past. The contributions to this special issue are no exception, as they address the questions raised by collaborative tagging:

- What should MIR do with tags? Should MIR aim at tag classification, just as it considered genre classification, i.e. consider tags as a ground-truth for pattern recognition approaches? Or, on the contrary, should pattern recognition be used to validate and filter out tags that are “musically meaningful” for MIR systems? In his review paper “Social Tagging and Music Information Retrieval”, **Paul Lamere** (Sun Labs, Mass. USA) gives an overview of how tags are collected and used in current systems. This article defines many of the important concepts of collabora-

tive tagging and social networks, and is suitable for the non-technical reader.

- Are tags just more complicated genres, or different semantic concepts altogether? What is the relationship between tags and the traditional entities addressed by MIR, such as genres or moods? In their contribution to this issue, **Mark Levy** and **Mark Sandler** from the Centre for Digital Music in Queen Mary University of London, UK, apply latent semantic dimension reduction to show that collections of tags follow an underlying topology that is strongly organized by genres and artists.
- Do the current computational paradigms, designed for learning in closed-world classification tasks, scale up to the diversity and complexity of tags? Can they be adapted, or do they need to be completely revised? In their paper entitled “Autotagger”, **Thierry Bertin-Mahieux**, **Douglas Eck**, **François Maillet**, **Thierry Bertin-Mahieux**, **Douglas Eck**, **François Maillet** from the Université de Montreal, Canada and **Paul Lamere** from Sun Labs report on large scale experiments to predict tags from audio signals using an online ensemble learning algorithm FilterBoost based on the concept of “infinite training data”.
- Are certain approaches to gather tags approaches better suited for MIR than others? What are the respective advantages of tags produced by unconstrained online communities such as last.fm to those produced in the emergent context of human computation games? In their contribution to this issue, **Michael Mandel** and **Dan Ellis** from Columbia University describe a web-based game, MajorMiner, designed to collect objective descriptions of musical excerpts. Their experiments show that the tags collected by MajorMiner are useful for training automatic music description algorithms, more so than social tags from a popular website.

In many ways, the articles in this issue offer more questions than answers. When they do, they often take contradictory points of view. What this collection of articles does offer, in our opinion, is a very exciting picture of the challenges faced by the MIR community and the finest illustration of the formidable dedication and intellectual honesty with which these challenges are undertaken by its practitioners.

The papers in this special issue were invited based on remarked contributions at the International Conference on Music Information Retrieval held in September 2007 in Vienna, Austria. They went through two rounds of reviews and generated intense debate as they were modified to the point of sometimes doubling their length and scope and including many new empirical results. We wish to equally thank the authors and the reviewers for all their hard work – needless to say, each of them was instrumental to the success of this initiative. To the list of authors of the works published here, we wish to associate **Douglas**

**Turnbull** and his colleagues at the University of California San Diego, CA, USA, as well as **Ciro Cattuto** and his colleagues at University of Roma “La Sapienza” and the ISI Foundation, Italy. Both teams entertained the project of contributing to this issue and did not only by lack of time. We consider their respective excellent work, published elsewhere, as “virtual” articles to this issue. Naturally, our gratitude also goes to Editor-in-chief **Alan Marsden**, for his unconditional support and all the suggestions, as well as the Publishing Editor at Taylor&Francis, **Elizabeth Ryan**. Thanks to **Stephen Downie** for providing hosting space on music-ir.org for the supplemental data of “Scanning the dial”. Thanks finally to **Michael Fingerhut** for maintaining the repository of ISMIR papers and making possible our awkward datamining in the search for strange typographical errors.

Ten years separate the first presentation of “Scanning the dial” and the most recent research published in this special issue. In this lapse of time, the MIR community has made a transition from genres to tags. It’s been an exciting 10 years and we had an exciting time preparing this issue too. We hope you will share our feeling when discovering these pages.

*Jean-Julien Aucouturier<sup>a</sup> and Elias Pampalk<sup>b</sup>*  
(Guest editors)

<sup>a</sup>Temple University Japan, Tokyo, Japan;  
Riken Brain Science Institute, Tokyo, Japan  
Email: aucouturier@gmail.com

<sup>b</sup>Last.fm, London, UK  
Email: elias@last.fm

## References

- Aucouturier, J.-J. & Pachet, F. (2003). Representing musical genre: A state of art. *Journal of A New Music Research (JNMR)* 32(1), 83–93.
- Ghias, A., Logan, J., Chamberlin, D. & Smith, B.C. (1995). Query by humming – musical information retrieval in an audio database. In *Proceedings of ACM Multimedia*, San Francisco, California.
- Kassler, M. (1966). Toward musical information retrieval. *Perspectives of New Music*, 4, 59–67.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press (reprinted, 2003).
- Ramón y Cajal, S. (1999). *Advice for a young investigator*. Cambridge, MA: MIT Press.
- Scheirer, E.D. (2000). Music-listening systems. PhD thesis, Media Lab, Massachusetts Institute of Technology, Cambridge (MA).
- Thagard, P. (2005). How to be a successful scientist. In M.E. Gorman, R.D. Tweney, D.C. Gooding & A.P. Kinnon (Eds.), *Scientific and technological thinking* (pp. 159–171). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wold, E. & Blum, T. (1996). Content based classification, search and retrieval of audio. *IEEE Multimedia*, 3(3), 27–36.

## Appendix: Works used in this study which quote the original “Scanning the dial” study

Works appearing in ISMIR proceedings:

- George Tzanetakis, Georg Essl & Perry Cook, Automatic Musical Genre Classification Of Audio Signals, in Proceedings ISMIR 2001, Bloomington Indiana.
- George Tzanetakis, Andrey Ermolinskyi & Perry Cook, Pitch Histograms in Audio and Symbolic Music Information Retrieval, in Proceedings ISMIR 2002, Paris, France.
- Cory McKay & Ichiro Fujinaga, Automatic Genre Classification Using Large High-level Musical Feature Sets, in Proceedings ISMIR 2003, Barcelona, Spain.
- Ming Li & Ronan Sleep, Genre Classification Via an LZ278-based string kernel, in Proceedings ISMIR 2005, London, UK.
- Anders Meng & John Shawe-Taylor, An Investigation of Feature Models for Music Genre Classification Using the Support Vector Classifier, in Proceedings ISMIR 2005, London, UK.
- Nicolas Scaringella & Giorgio Zoia On the Modelling of Time Information for Automatic Genre Recognition Systems in Audio Signals, in Proceedings ISMIR 2005, London, UK.
- Cory McKay & Ichiro Fujinaga, Musical genre classification: Is it worth pursuing and how can it be improved?, in Proceedings ISMIR 2006, Victoria, BC, Canada.
- James Bergstra, Alexandre Lacoste & Douglas Eck, Predicting genre labels for artists using FreeDB, in Proceedings ISMIR 2006, Victoria, BC, Canada.
- Alastair J. D. Craft, Geraint A. Wiggins & Tim Crawford, How Many Beans Make Five? The Consensus Problem in Music-Genre Classification and a New Evaluation Method for Single-genre Categorization Systems, in Proceedings ISMIR 2007, Vienna, Austria.

Other works

- Karin Kosina, Music Genre Recognition, Master Thesis, Department of Media Technology and Design, Hagenber University, June 2002.
- Toni Heittola, Automatic Classification of Music Signals, Master Thesis, Department of Information Technology, Tampere University, Finland, Feb. 2003.
- William P. Hannah, Automated Music Genre Classification Based on Analyses of Web-based Documents and Listeners’ Organizational Schemes, Master thesis, School of Information and Library Science, University of North Carolina at Chapel Hill, May 2005.

- Stephen Downie & Joe Futrelle, Terascale Music Mining, Proceedings of 2005 ACM/IEEE SC-05 Conference.
- Ulaç Bağcı & Engin Erzin, Inter Genre Similarity Modelling for Automatic Music Genre Classification, in Proceedings DAFX 2006, Montreal, Canada.
- Tao Li & Mitsunori Ogihara, Toward Intelligent Music Information Retrieval, in IEEE Transactions on Multimedia, Vol. 8, No. 3, June 2006.
- Nicholas Scaringella, Giorgio Zoia & Daniel Mlynek, Automatic genre classification of music content: a survey, IEEE Signal Processing Magazine: Special Issue on Semantic Retrieval of Multimedia, 2006.
- Luigi Lancieri & Lucille Tanquerel, Mesure rapide de similarités musicales: Perception du rythme, in Proceedings Compression et Representation des Signaux Audiovisuels, Nov. 2006.
- André Holzapfel & Yannis Stylianou, Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features, in IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, No. 2, Feb. 2008.