Discourse analysis evaluation method for expressive musical interfaces

Dan Stowell, Mark D. Plumbley, Nick Bryan-Kinns
Centre for Digital Music
Queen Mary, University of London
London, UK
dan.stowell@elec.qmul.ac.uk

ABSTRACT

The expressive and creative affordances of an interface are difficult to evaluate, particularly with quantitative methods. However, rigorous qualitative methods do exist and can be used to investigate such topics. We present a methodology based around user studies involving Discourse Analysis of speech. We also present an example of the methodology in use: we evaluate a musical interface which utilises vocal timbre, with a user group of beatboxers.

Keywords

Evaluation, qualitative methods, discourse analysis, voice, timbre, beatboxing

1. INTRODUCTION

One of the motives for founding the NIME conference was to foster dialogue on the evaluation of musical interfaces [11]. Yet a scan of NIME conference proceedings finds only a few papers devoted to the development or application of rigorous evaluation methods. Many published papers do not include evaluation, or include only informal evaluation (e.g. quotes from, or general summaries of, user interviews). This may of course be fine, depending on the paper's purpose and context, and the stage of development of the research. But the further development of well-founded evaluation methods can only be of benefit to the field.

In a very useful discussion, Wanderley and Orio [18] look to the wider field of Human-Computer Interaction (HCI) for applicable methodologies, and suggest specific approaches for evaluating musical interfaces. Much of HCI focuses on interfaces which can be evaluated using goal-based tasks, where measurements can be made of (for example) how long a task takes, or how often users fail to achieve the goal. Wanderley and Orio's framework follows this route, recommending that experimenters evaluate users' precision in reproducing musical units such as glissandi or arpeggios. Later work uses Wanderley and Orio's framework [9, 10].

Precision is important for accurate reproduction. But for composers, sound designers, and performers of expressive or improvised music, it is not enough: interfaces should (among other things) be in some sense intuitive and offer sufficient freedom of expression [11, 8]. "Control $\neq$ expression" [4].

Using precision-of-reproduction as a basis for evaluation also becomes problematic for musical systems which are not purely deterministic. "Randomness" would seem to be the antithesis of precision, and therefore undesirable according to some perspectives, yet there are many musical systems in which stochastic or chaotic elements are deliberately introduced.

The question arises of how to evaluate interfaces more broadly than precision-of-reproduction. It is difficult to design an experiment that can reliably and validly measure qualities such as expressiveness and aesthetics.

Poepel [10] operationalises "expressivity" into a number of categories for stringed-instrument playing, and investigates these numerically using tasks followed by Likert-scale questionnaires. This limits users' responses to predefined categories, although a well-designed questionnaire can yield useful results. Unfortunately Poepel analyses the data using mean and ANOVA, which are inappropriate for Likert-scale (ordinal) data [6]. The questionnaire approach also largely reduces "expressivity" down to "precision" since in this case, the tasks presented concern the reproduction of musical units such as vibrato and dynamical changes.

Paine et al [9] use a qualitative analysis of semi-structured interviews with musicians, to derive "concept maps" of factors involved in expressive performance (for specific instruments). These are not used for evaluation, rather to guide design. In the evaluation of their instrument, the authors turn to a quantitative approach, analysing how closely users can match the control data used to generate audio examples.

We propose that qualitative methods approaches may prove to be useful tools for the evaluation of musical interfaces. This paper aims to be a contribution in that area, applying a rigorous qualitative method to study the use and affordances of a new musical interface.

1.1 Discourse Analysis

Interviews and free-text comments are sometimes reported in studies on musical interfaces. However, often they are conducted in a relatively informal context, and only quotes or summaries are reported rather than any structured analysis, therefore providing little analytic reliability. Good qualitative methods penetrate deeper than simple summaries, offering insight into text data [1]. *Discourse Analysis* (DA) is one such approach, developed and used in disciplines such as linguistics, psychology, and social sciences [14, chapter 6].

Essentially, DA's strength comes from using a *structured method* which can take apart the language used in discourses (e.g. interviews, written works) and elucidate the connections and implications contained within, while remaining faithful to the content of the original text [1]. DA is designed to go beyond the specific sequence of phrases

NIME08, Genova, Italy

used in a conversation, and produce a structured analysis of the conversational resources used, the relations between entities, and the "work" that the discourse is doing.

Uszkoreit [17] summarises the aim of DA very compactly:

> The problems addressed in discourse research aim to answer two general kinds of questions: (1) what information is contained in extended sequences of utterances that goes beyond the meaning of the individual utterances themselves? (2) how does the context in which an utterance is used affect the meaning of the individual utterances, or parts of them?

We should point out that DA is not usually regarded as one single method – rather, it's an approach to analysing texts. Someone looking for the single recipe to perform a DA of a text will be disappointed. However, specific DA methods do exist in the literature. Our DA method is elaborated in section 3.3.

In this paper we use DA to analyse interview data, in the context of a project to develop voice-based interfaces for controlling musical systems. First we give an overview of the interface we wish to evaluate.

2. VOICE TIMBRE REMAPPING

With recent improvements in timbre analysis and in computer power, the potential arises to analyse the timbre of a signal in real-time, and to use this analysis as a controller for synthesis or for other processes – in particular, the potential to "translate" the timbral variation of one source into the timbral variation of another source. This is the process which we refer to as *timbre remapping* [16]. De Poli and Prandoni [3] made an early attempt at such control, more recently investigated by Puckette [13].

One of the main issues is the construction of a useful *timbre space* for the purpose of timbre remapping. Timbre is often very loosely defined, and often taken to refer to all aspects of a sound beyond its pitch and loudness [7]. There are many options as to which acoustic features to derive, and how to transform them, so as to provide a continuous space that provides useable control to the performer. Some features exhibit interactions with pitch, and the variation of some features may depend strongly upon the type of source.

In the present work we derive a heterogeneous set of timbral features, mostly spectral but some time-domain. We then apply a Principal Components Analysis (PCA) to decorrelate the features and reduce dimensionality. Finally we apply a piecewise linear warping (using the range, mean, and standard deviation statistics) to shape the distribution of data points; we will come back to the reasons for this shortly. The construction of the timbre space is summarised in figure 1.

Thus far we have a procedure for creating a timbre space based on any input signal. We might want to analyse two different classes of signal in this way, and then map the timbral trajectory of one system onto another: for example, use the timbral trajectory of a voice to control the settings of a synthesiser, and produce the corresponding timbral trajectory. To do this, we take a point in the voice's timbre space, and find its nearest neighbour in the synthesiser's timbre space. If we can retrieve the synthesiser parameters which created this timbre, we can send those parameters to the synthesiser, thus "remapping" from the vocal timbre to the synthesiser timbre. This approach has the advantage of being independent of the exact relation of the target system's control space to its timbre space: it works even if the target system's controls have highly nonlinear and obscure relation to the timbres produced.

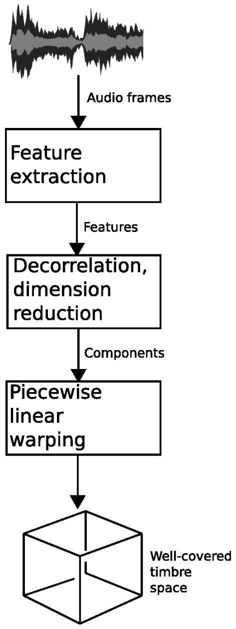

Figure 1: Constructing a timbre space for timbre remapping

Such a mapping from one timbre space to another depends on being able to find a suitable "nearest neighbour" in the target space. This is facilitated if the spaces are covered by a similar distribution of data points, ensuring that the resolution of a timbral trajectory can be adequately reflected in the target timbre space. This is why we perform a warping during the construction of the timbre space: it ensures that the timbre dimensions are covered in a certain way (guaranteeing various aspects of the distribution such as that it is centred and its standard deviation lies within a certain range).

One aspect of our timbre remapping system is that we typically wish to remove pitch-dependencies from the timbral data. Many acoustic measures such as MFCCs or spectral percentile statistics can exhibit interactions with pitch. Our current approach to mitigating this is to *include* a pitch analysis as one of the features passed to the PCA and therefore used in constructing the space. We then identify the PCA component with the largest contribution from pitch, and discard that, on the assumption that it is essentially composed of "pitch plus the pitch-dependent components of other features". This approach makes simplifying assumptions such as the linearity of pitch and timbre dimensions and their interaction, but it leads to usable results in our experience.

2.1 Real-time operation

We wish to develop a timbre remapping system that can operate efficiently in real-time, so the relative speed and efficiency of the processes used is paramount. In fact this is the strongest motivation behind using PCA for the decorrelation, dimension reduction, and pitch-removal. PCA is a straightforward process and computationally very simple to apply. More sophisticated methods, including non-linear methods, exist, and may be capable of improved results (such as better pitch-removal), but imply a significant cost in terms of the processing power required.

Efficiency is also important in the process which retrieves a nearest-neighbour data point from the target system's timbre space. We use a kd-tree data structure [12, chapter

2] for fast multidimensional search.

3. METHOD

In evaluating a musical interface such as the above, we wish to develop a qualitative method which can explore issues such as expressivity and affordances for users. Longitudinal studies may be useful, but imply a high cost in time and resources. Therefore our design aims to provide users with a brief but useful period of exploration of a new musical interface, including interviews and discussion which we can then analyse.

In any evaluation of a musical interface one must decide the context of the evaluation. Is the interface being evaluated as a successor or alternative to some other interface (e.g. an electric cello vs an acoustic cello)? Who is expected to use the interface (e.g. virtuosi, amateurs, children)? Such factors will affect not only the recruitment of participants but also some aspects of the experimental setup.

Our method is designed either to trial a single interface with no explicit comparison system, or to compare two similar systems (as is done below in our case study). The method consists of two types of user session (solo sessions followed by group session(s)), plus the Discourse Analysis of data collected.

3.1 Solo sessions

In order to explore individuals' personal responses to the interface(s), we first perform solo sessions in which a participant is invited to try out the interface(s) for the first time. If there is more than one interface to be used, the order of presentation is randomised in each session.

The solo session consists of three phases for each interface:

Free exploration The participant is encouraged to try out the interface for a while and explore it in their own way.

Guided exploration The participant is presented with audio examples of recordings created using the interface, and encouraged to create recordings inspired by those examples. This is not a precision-of-reproduction task; precision-of-reproduction is explicitly not evaluated, and participants are told that they need not replicate the examples.

Semi-structured interview The interview's main aim is to encourage the participant to discuss their experiences of using the interface in the free and guided exploration phases, both in relation to prior experience and to the other interfaces presented if applicable. Both the free and guided phases are video recorded, and the interviewer may play back segments of the recording and ask the participant about them, in order to stimulate discussion.

The raw data to be analysed is the interview transcript. Our aim is for the participant to construct their own descriptions and categories, which means it is very important that the interviewer is experienced in neutral interview technique, and can avoid (as far as possible) introducing labels and concepts that do not come from the participant's own language patterns.

3.2 Group session

To complement the solo sessions we also conduct a group session: peer group discussion can produce more and different discussion around a topic, and can demonstrate the group negotiation of categories, labels, comparisons, etc. The focus-group tradition provides a well-studied approach to such group discussion [15]. Our group session has a lot in common with a typical focus group in terms of the facilitation and semi-structured group discussion format. In addition we make available the interface(s) under consideration and encourage the participants to experiment with them during the session.

As in the solo sessions, the transcribed conversation is the data to be analysed, which means that a neutral facilitation technique is important – to encourage all participants to speak, to allow opposing points of view to emerge in a non-threatening environment, and to allow the group to negotiate the use of language with minimal interference.

3.3 Data analysis

Our DA approach to analysing the data is based on that of [2, p. 95–102], adapted to our study context. The DA of text is a relatively intensive and time-consuming method. It can be automated to some extent, but not completely, because of the close linguistic attention required. Our approach consists of the following five steps:

(a) Transcription

The speech data is transcribed, using a standard style of notation which includes all speech events (including repetitions, speech fragments, pauses). This is to ensure that the analysis can remain close to what is actually said, and avoid adding a gloss which can add some distortion to the data. For purposes of analytical transparency, the transcripts (suitably anonymised) should be published alongside the analysis results.

(b) Free association

Having transcribed the speech data, the analyst reads it through and notes down surface impressions and free associations. These can later be compared against the output from the later stages.

(c) Itemisation of transcribed data

The transcript is then broken down by itemising every single object in the discourse (i.e. all the entities referred to). Pronouns such as "it" or "he" are resolved, using the participant's own terminology as far as possible, and for every object an accompanying description is extracted, of the object as it is in that instance – again using the participant's own language, essentially by rewriting the sentence/phrase in which the instance is found.

The list of objects is scanned to determine if different ways of speaking can be identified at this point. Also, those objects which are also "actors" (or "subjects") are identified – i.e. those which act with agency in the speech instance; they need not be human.

It is helpful at this point to identify the most commonly-occurring objects and actors in the discourse.

(d) Reconstruction of the described world

Starting with the list of most commonly-occurring objects and actors in the discourse, the analyst reconstructs the depictions of the world that they produce. This could for example be achieved using concept maps to depict the interrelations between the actors and objects. If different ways of speaking have been identified, there will typically be one reconstructed "world" per way of speaking. Overlaps and contrasts between these worlds can be identified.

The "worlds" we produce are very strongly tied to the participant's own discourse. The actors, objects, descriptions, relationships, and relative importances, are all derived from a close reading of the text. These worlds are essentially just a methodically reorganised version of the participant's own language.

In our particular context, we may be interested in the user's conceptualisation of musical interfaces. It is particularly interesting to look at how these are situated in the described world, and particularly important to avoid preconceptions about how users may describe an interface: for example, a given interface could be: an instrument; an extension of a computer; two or more separate items (e.g. a box and a screen); an extension of the individual self; or it could be absent from the discourse.

(e) Examining context

The relevant context of the discourse typically depends on the field of study, for example whether it is political or psychological. Here we have created an explicit context of other participants. After running the previous steps of DA on each individual transcript, we compare and contrast the described worlds produced from each transcript, first comparing those in the same experimental condition (i.e. same order of presentation, if relevant), then across all participants. We also compare the DA of the focus group session(s) against that of the solo sessions.

4. THE METHOD IN ACTION: EVALUATING VOICE TIMBRE REMAPPING

In our study we wished to evaluate the timbre remapping system with beatboxers (vocal percussion musicians), for two reasons: they are one target audience for the technology in development; and they have a familiarity and level of comfort with manipulation of vocal timbre that should facilitate the study sessions.

We recruited by advertising online (a beatboxing website) and around London for amateur or professional beatboxers. Participants were paid £10 per session plus travel expenses to attend sessions in our (acoustically-isolated) studio. We recruited five participants from the small community, all male and aged 18–21. One took part in a solo session; one in the group session; and three took part in both. Their beatboxing experience ranged from a few months to four years. Their use of technology for music ranged from minimal to a keen use of recording and effects technology (e.g. Cubase).

In our study we wished to investigate any effect of providing the timbre remapping feature. To this end we presented two similar interfaces: both tracked the pitch and volume of the microphone input, and used these to control a synthesiser, but one also used the timbre remapping procedure to control the synthesiser's timbral settings. The synthesiser used was an emulated General Instruments AY-3-8910 [5], which was selected because of its wide timbral range (from pure tone to pure noise) with a well-defined control space of a few integer-valued variables. We used the method as described in section 3. Analysis of the interview transcripts took approximately 10 hours per participant (around 2000 words each).

We do not report a detailed analysis of the group session transcript here: the group session generated information which is useful in the development of our system, but little which bears directly upon the presence or absence of timbral control. We discuss this outcome further in section 5.

In the following, we describe the main findings from analysis of the solo sessions, taking each user one by one before drawing comparisons and contrasts. We emphasise that although the discussion here is a narrative supported by quotes, it reflects the structures elucidated by the DA process – the full transcripts and Discourse Analysis tables are available online[1]. In the study, condition "Q" was used to refer to the system with timbre remapping active, "X" for the system with timbre remapping inactive.

4.1 Reconstruction of the described world

User 1

User 1 expressed positive sentiments about both Q and X, but preferred Q in terms of sound quality, ease of use and being "more controllable". In both cases the system was construed as a reactive system, making noises in response to noises made into the microphone; there was no conceptual difference between Q and X – for example in terms of affordances or relation to other objects.

The "guided exploration" tasks were treated as reproduction tasks. User 1 described the task as difficult for X, and easier for Q, and situated this as being due to a difference in "randomness" (of X) vs. "controllable" (of Q).

User 2

User 2 found the the system (in both modes) "didn't sound very pleasing to the ear". His discussion conveyed a pervasive structured approach to the guided exploration tasks, in trying to infer what "the original person" had done to create the examples and to reproduce that. In both Q and X the approach and experience was the same.

Again, User 2 expressed preference for Q over X, both in terms of sound quality and in terms of control. Q was described as more fun and "slightly more funky". Interestingly, the issues that might bear upon such preferences are arranged differently: issues of unpredictability were raised for Q (but not X), and the guided exploration task for Q was felt to be more difficult, in part because it was harder to infer what "the original person" had done to create the examples.

User 3

User 3's discourse placed the system in a different context compared to others. It was construed as an "effect plugin" rather than a reactive system, which implies different affordances: for example, as with audio effects it could be applied to a recorded sound, not just used in real-time; and the description of what produced the audio examples is cast in terms of an original sound recording rather than some other person. This user had the most computer music experience of the group, using recording software and effects plugins more than the others, which may explain this difference in contextualisation.

User 3 found no difference in sound or sound quality between Q and X, but found the guided exploration of X more difficult, which he attributed to the input sounds being more varied.

User 4

User 4 situated the interface as a reactive system, similar to Users 1 and 2. However, the sounds produced seemed to be segregated into two streams rather than a single sound – a "synth machine" which follows the user's humming, plus "voice-activated sound effects". No other users used such separation in their discourse.

"Randomness" was an issue for User 4 as it was for some others. Both Q and X exhibited randomness, although X was much more random. This randomness meant that User 4 found Q easier to control. The pitch-following sound was

[1] http://www.elec.qmul.ac.uk/digitalmusic/papers/2008/Stowell08nime-data/

felt to be accurate in both cases; the other (sound effects / percussive) stream was the source of the randomness.

In terms of the output sound, User 4 suggested some small differences but found it difficult to pin down any particular difference, but felt that Q sounded better.

4.2 Examining context

Effect of order-of-presentation

Users 1 and 2 were presented with the conditions in the order XQ; Users 3 and 4 in the order QX. Order-of-presentation may have some small influence on the outcomes: Users 3 and 4 identified little or no difference in the output sound between the conditions (User 4 preferred Q but found the difference relatively subtle), while Users 1 and 2 felt more strongly that they were different and preferred the sound of Q. It would require a larger study to be confident that this difference really was being affected by order-of-presentation.

In our study we are not directly concerned with which condition sounds better (both use the same synthesiser in the same basic configuration), but this is an interesting aspect to come from the study. We might speculate that differences in perceived sound quality are caused by the different way the timbral changes of the synthesiser are used. However, participants made no conscious connection between sound quality and issues such as controllability or randomness.

Considerations across all participants

Taking the four participant interviews together, no strong systematic differences between Q and X are seen. All participants situate Q and X similarly, albeit with some nuanced differences between the two. Activating/deactivating the timbre remapping facet of the system does not make a strong enough difference to force a reinterpretation of the system.

A notable aspect of the four participants' analyses is the differing ways the system is situated (both Q and X). As designers of the system we may have one view of what the system "is", perhaps strongly connected with technical aspects of its implementation, but the analyses presented here illustrate the interesting way that users situate a new technology alongside existing technologies and processes. The four participants situated the interface in differing ways: either as an audio effects plugin, or a reactive system; as a single output stream or as two. We emphasise that none of these is the "correct" way to conceptualise the interface. These different approaches highlight different facets of the interface and its affordances.

During the analyses we noted that all participants maintained a conceptual distance between themselves and the system, and analogously between their voice and the output sound. There was very little use of the "cyborg" discourse in which the user and system are treated as a single unit, a discourse which hints at mastery or "unconscious competence". This fact is certainly understandable given that the participants each had less than an hour's experience with the interface. It demonstrates that even for beatboxers with strong experience in manipulation of vocal timbre, controlling the vocal interface requires learning – an observation confirmed by the participant interviews.

The issue of "randomness" arose quite commonly among the participants. However, randomness emerges as a nuanced phenomenon: although two of the participants described X as being more random than Q, and placed randomness in opposition to controllability (as well as preference), User 2 was happy to describe Q as being more random and also more controllable (and preferable).

A uniform outcome from all participants was the conscious interpretation of the guided exploration tasks as precision-of-reproduction tasks. This was evident during the study sessions as well as from the discourse around the tasks. As one participant put it, "If you're not going to replicate the examples, what are you gonna do?"

A notable absence from the discourses, given our research context, was discussion which might bear on expressivity, for example the expressive range of the interfaces. Towards the end of each interview we asked explicitly whether either of the interfaces was more expressive, and responses were generally non-commital. We propose that this was because our tasks had failed to engage the participants in creative or expressive activities: the (understandable) reduction of the guided exploration task to a precision-of-reproduction task must have contributed to this. We also noticed that our study design failed to encourage much iterative use of record-and-playback to develop ideas. In section 5 we suggest some possible future directions to address these issues.

5. DISCUSSION

The analysis of the solo sessions provides useful information on the user experience of a voice-controlled music system and the integration of timbre remapping into such a system. Here we wish to focus on methodological issues arising from the study.

Above we raised the issue that our "guided exploration" task, in which participants were asked to record a sound sample on the basis of an audio example, was interpreted as a precision-of-reproduction task. Possibilities to avoid this in future may include: using audio examples which are clearly not originally produced using the interface (e.g. string sections, pop songs), or even non-audio prompts such as pictures; or forcing a creative element by providing two examples and asking participants to create a new recording which combines elements of both.

Other approaches which encourage creative work with an interface could involve tasks in which participants are asked to create compositions, or iteratively develop live performance. We would expect that the use of more creative tasks should produce more participant discussion of creative/expressive aspects of an interface.

Such tasks could also be used to provide more structure during the group sessions: one reason the group session produced less relevant data than the solo sessions is (we believe) the lack of activities, which could have provided a more structured exploration of the interfaces.

6. CONCLUSIONS

We have applied a detailed qualitative analysis to user studies involving a voice-driven musical interface with and without the use of timbre remapping. It has raised some interesting issues in the development of the interface, including the unproblematic integration of the timbral aspect, and the nuanced interaction of issues such as control and randomness.

However, the primary aim of this paper has been to investigate the use of Discourse Analysis to provide a robust qualitative approach to evaluating the affordances and user experience of a musical interface. Results from our DA-based user study indicate that with some modification of the user tasks, the method can derive detailed information about how musicians interact with a new musical interface and accommodate it in their existing conceptual repertoire.

We have presented one specific method for evaluating a musical interface, but of course there may be other appropriate methods. As discussed in the introduction, the state of the art in evaluating musical interfaces is relatively underdeveloped, and we would hope to encourage others to explore reliable methods for evaluating new musical interfaces in authentic contexts.

7. REFERENCES

[1] C. Antaki, M. Billig, D. Edwards, and J. Potter. Discourse analysis means doing analysis: A critique of six analytic shortcomings. *Discourse Analysis Online*, 2004.

[2] P. Banister, E. Burman, I. Parker, M. Taylor, and C. Tindall. *Qualitative Methods in Psychology: A Research Guide*. Open University Press, Buckingham, 1994.

[3] G. De Poli and P. Prandoni. Sonological models for timbre characterization. *Journal of New Music Research*, 26(2):170–197, 1997.

[4] C. Dobrian and D. Koppelman. The 'E' in NIME: musical expression with new computer interfaces. In *Proceedings of New Interfaces for Musical Expression (NIME)*, pages 277–282. IRCAM, Centre Pompidou Paris, France, 2006.

[5] General Instrument. *GI AY-3-8910 Programmable Sound Generator datasheet*, early 1980s.

[6] J. Harvey. *Evaluation Cookbook*, chapter So You Want to Use a Likert Scale? Learning Technology Dissemination Initiative, 1998.

[7] J. Kreiman, D. Vanlancker-Sidtis, and B. R. Gerratt. Defining and measuring voice quality. In *Proceedings of From Sound To Sense: 50+ Years of Discoveries in Speech Communication*, pages 115–120. MIT, June 2004.

[8] T. Magnusson and E. H. Mendieta. The acoustic, the digital and the body: A survey on musical instruments. In *Proceedings of New Interfaces for Musical Expression (NIME)*, June 2007.

[9] G. Paine, I. Stevenson, and A. Pearce. The thummer mapping project (ThuMP). In *Proceedings of New Interfaces for Musical Expression (NIME)*, pages 70–77, 2007.

[10] C. Poepel. On interface expressivity: A player-based study. In *Proceedings of New Interfaces for Musical Expression (NIME)*, pages 228–231, 2005.

[11] I. Poupyrev, M. J. Lyons, S. Fels, and T. Blaine. New Interfaces for Musical Expression. Workshop proposal, 2001.

[12] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, 1985.

[13] M. Puckette. Low-dimensional parameter mapping using spectral envelopes. In *Proceedings of the International Computer Music Conference (ICMC'04)*, pages 406–408, 2004.

[14] D. Silverman. *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*. Sage Publications Inc, 2nd edition, 2006.

[15] D. W. Stewart. *Focus groups: Theory and practice*. SAGE Publications, 2007.

[16] D. Stowell and M. D. Plumbley. Pitch-aware real-time timbral remapping. In *Proceedings of the Digital Music Research Network (DMRN) Summer Conference*, July 2007.

[17] H. Uszkoreit. *Survey of the State of the Art in Human Language Technology*, chapter 6 (Discourse and Dialogue). Center for Spoken Language Understanding, Oregon Health and Science University, 1996.

[18] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, 26(3):62–76, 2002.