etherSound - an interactive sound installation

Henrik Frisk
Malmö Academy of Music - Lund University
Box 8203, 20041 Malmö, Sweden
henrik.frisk@mhm.lu.se

September 18, 2007

Abstract

This article describes the interactive instrument/sound installation *etherSound* and discusses its artistic and ethical implications. *etherSound* is a work in progress and the main intention is to create a vehicle for audience participation through the use of SMS (Short Message Service). The two different contexts in which *etherSound* has been tried (in concert with performers and as a sound installation without performers) are discussed as well as the design of the system and the mapping between text and sound. A notion of a 'democracy of participation' is introduced. The relatively fast response of the system, the familiarity of the interface (the cellular phone) and the accessibility of the system suggests that the cellular phone can be successfully integrated in a sonic art work.

Contents

List of Figures

List of Tables

1 Introduction

etherSound was commissioned by the curator Miya Yoshida for her project The Invisible Landscapes and was realized for the first time in August 2003 at Malmö Art Museum in the city of Malmö, Sweden. The curatorial concept for The Invisible Landscapes project was the use of cellular phones in the context of experiencing and creating artistic expressions. The principle idea behind *etherSound* came to be an attempt at developing an instrument that can be played by anybody who has knowledge about how to send an sms (Short Messages Service) from their cellular phones. The focus of my research project, of which *etherSound* is a part, is interaction between computers and musicians as well as non-musicians. *etherSound* is an investigation of some of the aspects of interaction between the listener, the sounds created and the musicians playing, and also of the formal and temporal distribution of the music that this interaction results in.

While interaction is an important aspect of musical performance in many genres, active audience participation is not as evolved in the western music tradition, and when explored, the result is usually not labeled as music, but rather as a sound installation, soundscape, sonic art or some other term that indicates alienation from the traditional notion of music. Opening up a musical work for others than trained musicians is not a trivial task; careful attention has to be paid to the purpose of doing so and to the intentions of the work. It is relevant to pose the question whether it is possible to reach a satisfactory result with almost no limitations on participation and, if so, can the result not be called music? However, before these questions can be addressed we need delineate the purposes for wanting to allow for public participation.

Public participation has been explored in the visual arts for almost a century, for artistic as well as political reasons, and if we look at it from a performing arts perspective, the audience visiting a performance can be said to participate in it - if only in a limited sense. Concurrently, especially in spheres of distribution and consumption of music, there is a tendency to objectify the musical work. As the power and irrational control excercised by the institutions of distribution increases, the freedom of choice and influence of the listener decreases [Adorno,]. Furthermore, western art music is to a considerable extent looked upon as a hierarchic process; a process that begins in the mind of the composer and ends at the level of the listener or, even before that, at the level of interpretation. It is fair to assume that bringing in an uncontrollable agglomeration of participants influencing the distribution of musical events will disturb this order .

In their article on the multi-participant environment *The Interactive Dance Club*, Ulyate and Bianciardi defines one of the design goals as wanting to 'deliver the euphoria of the artistic experience to "unskilled" participants' [Ulyate and Bianciardi, 2002]. Instead of sharing merely the result with an audience, they attempt at unfolding the creative process leading to the result and invite the audience to take part in this process. This ambition points to another issue: how to design musical interfaces that have a 'low entry fee, with no ceiling on virtuosity' [Wessel and Wright, 2002, Jordà, 2002] (see also [Rowe, 1993, Freeman et al., 2004]). With the recent technological advances there are innumerable tools that can be used for collaborative efforts [Barbosa and Kaltenbrunner, 2002], affordable devices that easily can be used as interfaces to computer mediated art works. Not only has this the potential of changing our perception of the arts, it can also help us understand and this new technology and the impact it has on our lives.

Traditionally there is an intimate association between social class, level of education and cultural interests [DiMaggio and Useem, 1978, Bourdieu, 1979] that affects cultural consumption. Is it possible to make music that can counteract this 'closeness' of contemporary art and music; that can make conditions for classless and unprejudiced participation in the arts without compromising the content and the expression? I believe it is and I believe collaborative music is one way to achieve this. Roy Ascott, in addressing the issue of 'content' in art involving computers and telecommunications writes:

> In telematic art, meaning is not something created by the artist, distributed through the network, and *received* by the observer. Meaning is the product of interaction between the observer and the system, the content of which is in a state of flux, of endless change and transformation [Ascott, 1990].

A similar idea is put forward by Guy E. Garnett in his article on interactive computer music aesthetics. In the quote below he refers to "very unfixed works" such as "human improvisation with computer partner":

> Since the human performance is a variable one, by its nature, that variability can become the focus of aesthetic issues, even simple ontological issues. Because the performance changes from time to time and from performer to performer, the notion of 'the work' becomes more and more clouded. The work, even from an objective rather than an immanent point of view, becomes something open-ended. Each performance becomes an 'interpretation' of the possibilities inherent in whatever was 'composed.' However, each of these concepts is highly problematic. This 'interpretation' can have significant consequences for the meaning - and therefore value - of a work in a cultural context. Since the work is not fixed, it is open to new interpretations, and therefore the possibility at least exists for the growth of the work over time or across cultural boundaries. The work can thus maintain a longer life and have a broader impact culturally, because it is able to change to meet changing aesthetic values. [Garnett, 2001]

Following these lines of thought, it may be concluded that the need for a thorough insight in the history of art or electronic music is no longer a prerequisite for understanding a collaborative, interactive work. This limits the advantage of the educated listener and makes room for new interpretations of the term 'understanding' in the arts.

2 The Design

etherSound is an attempt to open a musical work to the uninitiated and provide for a notion of 'democracy of participation': all contributions are equally valuable. Accessibility without prior knowledge of music or musical training is an end in itself in this project. It should be noted that this obviously presupposes that the participant knows how to send a SMS and that the system makes it difficult for those who are not familiar with this technology[1]. It should also be made clear that, using SMS text messages for interaction as it is implemented here, does not allow for direct dynamic control. Every message generates one 'message-composition' and all control data is derived from the content of the message.

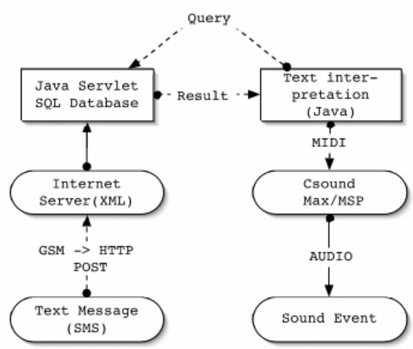

Figure 1: Communication in the first version.

2.1 Communication - first model

In the first version, realized in August 2003, the communication between the participant and the system was accomplished according to Figure 1. A sms sent to the specified number was transformed to a XML file and transferred to a URL by a HTTP POST request. This part was handled through an external service. At the called URL, a JSP (Java Server Pages) was directing the POST data to a Java Bean [J2EE 1.4.1, 2004] that handled the parsing of the data and the connection to a MySQL database in which it created a new entry with the relevant fields.

It was due to security reason at the museum where this version was realized that the HTTP request could not be handled locally. Instead, the local computer queried the server database for new entries on regular intervals. After some testing, sending a SQL query once every second seemed like a reasonable time interval. Shorter time intervals didn't accomplish a perceivably quicker response time and, since the synthesis program was running on the same machine, I didn't want to use more processing and network activity than necessary for this task (see section 3 for further discussion). After the text message had been processed, control signals where sent by MIDI to the synthesis engine.

2.2 Communication - current model

Although the first version worked well and was fairly stable, it was a solution that required an external SMS processing service, and a local, reliable

[1]Yet there is a great commercial interest in increasing the use of sms and, in Sweden, there has been a tremendous effort from the part of the GSM service providers to teach their customers how to use it.

network connection. In order to make the piece more 'portable' and independent, the message receiving part has been rebuilt. Using the gnokii API [gnokii, 1995] it is relatively easy and reliable to connect a GSM phone to a computer and gain access to the storage and status of the phone which enables reception of the SMS messages locally. To have the possibility to review the activity of transmission, the messages are, just as in the first model, written to a database. In other words, the client-server model is retained but on one and the same machine. Furthermore, the MIDI connection between the control application and the synthesis engine has been replaced with OpenSound Control (OSC) [Wright et al., 2003, OSC, 1997] for speed, reliability and flexibility, using the library JavaOSC (see `http://www.mat.ucsb.edu/~c.ramakr/illposed/javaosc.html`).

2.3 The text analysis

The program handling the text processing and the mapping of text to control signals for the sound synthesis is written in Java [J2SE 1.4.2, 2004] and features a simple but useful GUI for control and feedback about the status of the system. It is here, in the mapping between the text and the sound, that the compositional choices have been made. There are three groups of parameters that are being extracted for every message:

- The length of the whole event
- The rhythm and articulation of the individual sound events
- The pitch and character of individual sound events

For the timing there are two parameters; a local 'life' index shaping the rhythms and the length of the current message and a global index that influences the current and subsequent 'message-compositions'. The global index is a function of the current and previous messages local indexes. The purpose of the local index is to make a simple semantic analysis of the message and thus discriminate between a set of random letters and real words. The participant should be 'rewarded' for the effort of writing a message with substance, where 'substance' is defined here as a short message with a credible average word length and a reasonable distribution of vowels within these words. The local index is calculated by looking at the average length of words and the average number of syllables per word and comparing these with constants:

$$i_1 = \frac{1}{(w(\frac{c}{w_c}) - w_l)^{1/2} + 1} \qquad i_2 = \frac{1}{(w(\frac{s}{w_c}) - s_l)^{1/2} + 1} \tag{2.1}$$

where c and s are the total number of characters and syllables, w_c is the number of words in the current message, w_l and s_l are constants defining the 'optimal' mean number of words/syllables. w is a weight defined by

$$w = \frac{1}{w_c - s_c + 0.5} \tag{2.2}$$

where s_c is the total number of words that contains vowels. Through w, the index is decreased if the message contains words without vowels. The mean

message	*life index*
hello	0.18882
Hello, my name is Henrik	*0.81032*
hjdks la s duyfke jhsldf hasdfiw uehr jkdsl	0.14448
From fairest creatures we desire increase, That thereby beautys rose might never	*1.44618*

Table 1: Life index for four different messages

value of i_1 and i_2 is then multiplied by the arcus tangens of the number of words in relation to a third constant parameter, o_w, delimiting the optimal number of words per message[2] according to (2.3).

$$lifeIndex = \frac{i_1 + i_2}{2}\arctan(\frac{w_c}{o_w}) \tag{2.3}$$

If we set w_l to 4.5, s_l to 2.0 and o_w to 10 the result on four different messages can be seen from Table 1; the method distinguishes fairly well between nonsense and real words at a low computational cost. Similar or better results could conceivably be achieved in a number of different ways but this method appears to work well for the purpose. Since there is only audio feedback, it is important that all, even empty messages, will lead to a perceptible change in the sonic output. The total length of the music derived from the message is calculated by multiplying a constant preset time with the local index. Any new messages received adds its local index to the instantaneous global index which constantly decreases exponentially at a set rate[3]. If a message causes the global index to reach maximum, it stops the playback of the current message and begins playing back a precomposed pattern, sonically different from the output of a typical message, for about 30 seconds before resuming ordinary mode and starts playing back the message that caused the break. This feature is added to reward collaborative efforts. The global index controls mainly the density and the overall volume of the output, but also the distribution of random and stochastic processes in the synthesis.

2.4 The synthesis

The synthesis engine is written as a Csound orchestra [Boulanger, 2000] (see also `http://www.csounds.com/`) running inside a Max/MSP (`http://www.cycling74.com/products/maxmsp.html`) patch through the use of the `csound~` object (see `http://www.csounds.com/matt/`). The 'score' for the message to be played back is sent to Max/MSP using OSC. Max/MSP is responsible for timing the note events and preparing valid information for the `csound~` object and the orchestra file associated with it. Due to processing power limitations only one message can be played back simultaneously; if a message is received before the previously received message has finished playing back, the new message will interrupt the current message.

All sounds heard in *etherSound* are generated with FOF (Fonction d'Onde Formantique) synthesis as this technique is implemented in Csound [Clarke, 2000,

[2]Since a SMS is limited to 160 characters these constants are set according to what kind of message content should be rewarded.

[3]The rate is context dependent. In a performance with improvisation it would be shorter than in an installation.

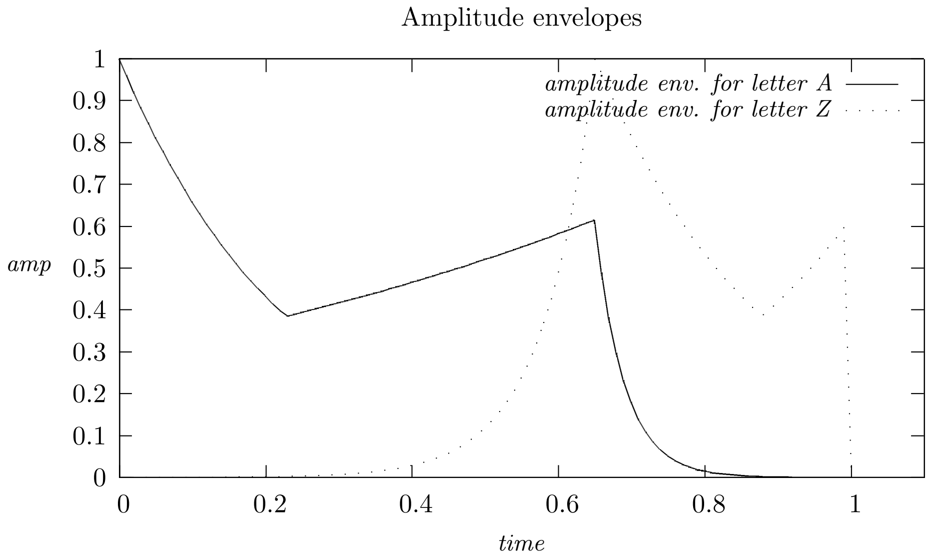

Figure 2: Amplitude envelopes for Instrument A. Linear interpolation between these two envelopes is performed for every character between A and Z.

Byrne Villez, 2000], using both samples and simple sine waves as sound sources. There are two distinct timbres played by two different instruments in each 'message-composition': (A) granulated samples of a male reading a text in english[4] and (B) a bell like sound whose timbre is governed by the series of vowels in the text. The timbre as well as the generative rules of the first voice are in contrast with the second voice.

2.4.1 Instrument A

Every word of the message is considered one phrase or *bar* of music in the resulting message composition. The number of beats per bar is aproximately equal to the number of syllables in the word, where a syllable is defined as a vowel or group of consecutive vowels or a punctuation mark. The rhythmic subdivision of each bar is equal to the number of characters, including punctuation and white space, in each syllable. Thus, a one syllable word such as 'my' followed by a white space results in a phrase consisting of one bar of one beat and two notes and one pause, i.e. three (eight-note) triplets of which the last is silent (see Table 2). If a word ends with a full stop, a comma, an exclamation mark or a question mark, more emphasis is put on the end of the bar containing the punctuation mark and the last note of the resulting phrase will be elongated. A note close to a vowel will more likely be accented than a note away from a vowel.

The amplitude envelope curve of each note is related to the letter the note corresponds to. Envelopes are mapped linearly to characters; letter 'A' has a short attack and a long decay and letter 'Z' has a long attack and a short decay (see Figure 2). The amount of overlapping between notes, i.e. the lengths of

[4]An excerpt of the recording of one of John Cage's lectures at Harvard College 1989.

the notes, is influenced by the current life index *and* the global index where higher values will result in longer notes and thus in smoother transitions between timbres. The notes of Instrument A does note have a perceivable pitch. Twenty-eight short sample buffers (typically 32.768 samples or aproximately 0.7 seconds), one for each letter, are mapped one to one to the characters in the message. The FOF synthesis is used to granulate these samples, creating an erratic, non-tonal texture however still, in most cases, reminiscent of speech.

Figure 3: Rhythmic distribution of notes in Instrument B.

2.4.2 Instrument B

The phrasing of the notes of the second instrument is somewhat more complex than that of Instrument A. This instrument has, at the most[5], as many voices as there are words in the message. If the polyphony of this instrument is limited to four voices the rhythmic mapping of the notes using the message in Table 2 is shown in Figure 3. For this instrument the number of beats per bar (i.e. per word) is equal to the *number of letters* per word, including trailing punctuation marks and white space. If there are less words than the maximum polyphony, the number of voices is equal to the number of words; the first voice correspond to the first word, the second voice to the second word and so forth. For every bar, each voice has as many potential excitations as there are letters in the corresponding word. After the initial excitation, which will always be played, the likelyhood that a given note will be played is related to the life index *and* the global index: If the normalized sum of the local index and the global index is 0.5, half of the excitations will be performed. The amplitude envelope curve for the notes played by this instrument is either of a bell like character or of its inversion, and notes close to the beginning of a bar has a greater likelyhood of being emphasized. The initial pitches are derived from the ocurrence of certain key letters in the originating text[6]. The first unique occurrence of one of the key letters, searched for from the first letter of the word corresponding to the current

[5]The maximum number of voices is set based on what the processing power of the system is.

[6]For the sake of experiment and variation, I am changing these 'key notes' for every performance of *etherSound*.

Figure 4: Harmony for Instrument B as a result of a received message "*Hello, my name is Henrik.*".

voice until the end of the message, becomes the initial pitch for each voice. If none is found the voice is deleted. The voicing of the initial chord is constructed so that the first voice will be the top note of the chord and consecutive voices will be laid out below this using octave transposition, aiming for the closest possible voicing.

The exact microtonal centre pitch between the highest and lowest note of the initial chord is then calculated (this would be the pitch 'D' if the initial chord is a major third up from 'C'). After the initial chord has been introduced, all voices begin a virtual glissando toward the centre between the outer limits of the chord, creating microtonal variations of an ever decreasing harmony, ending at a unison. For each excitation of each voice, the instantaneous value of the corresponding glissando sets the pitch for that excitation. Figure 4 shows the initial chord and the glissandi towards the centre the message from Table 2 would result in if the max polyphony value is set to five or higher and the 'key' characters were mapped by German note names (a to A, b to Bb, c to C, ... ,h to B and s to Eb).

The timbre of the voices played by this instrument is also shaped by the vowels contained in the message and the order in which they appear. For non real time processing this is achieved by synthesizing the first five formants of the first vowel found in the word corresponding to the current voice and then interpolating between the formant spectrum of the remaining vowels of the message (see Table 3). As this method is very expensive - it requires allocation of five times more voices - a cheaper variation has been implemented for real time usage. By modulating the formant frequency of one single FOF voice with Frequency Modulation whose carrier signal and index is derived from the vowel interpolation described above, the effect of molding the formants spectrum with data about the content of the message is retained. However, it should be made clear that the sonic output of these two models is very different.

2.5 Sound event generation and synthesis - conclusion

The two instruments offer two different interpretations of the message played back in paralell. As Instrument A performs a linear displacement within the message, Instrument B gives a snapshot image of the entire text at once, an image which is gradually dissolving over time. One instrument is modelling the discrete words and characters as they appear in time, the objective flow of the components of the message, and the other deals with the continuous meaning, or subjective understanding, of the message as it is understood in its entirety. Although the result can be rather complex and abstract it is my intention that certain conceptual elements of the input should be retained in the output.

3 Discussion

As has already been explained, the main issue for *etherSound* is to allow for unconditioned participation. It comes natural that unconditioned perception should similarly be allowed for. In the first versions of *etherSound* the message compositions where less dependent on input than what they are in the current version. I was more concerned with the collection of diverse input than I was of giving the contributor a sense of control, or participation. My grounds for letting the messages generate a musical event with a clear form stems from the wish to retain a perceptible connection - even though this connection may only be dismantled by the *change* in output - between input and output.

The process of designing the analysis and synthesis programs described above is to a considerable extent tantamount with the process of composing in the traditional meaning. In a sense, *etherSound* is an algorithmic or ruled based composition with stochastic elements, methods which have been explored by many composers for many years. However, there is a great difference between a traditional composition and a work such as *etherSound*and in the dimension of time, as the latter does not have a fixed beginning nor an end. An interactive, ongoing and indeterminate, musical creation will inevitably dismantle the traditional idea of musical form. There is nothing new with the "permanent event" [Barbosa, 2003] or the infinite musical form - it has been explored by many composers for many years - but it is the effect the indeterminate form has on the *understanding* and *interpretation*[7] of the work from the point of the participant, and whether the closed form of the message compositions enhance or degenerate this effect, that is my concern. Will a random collection of message compositions, each one with a sense of musical form, generate a large scale (closed) form or will they result in something else, conceptually different from musical form? I believe both is possible and, in this particular case, they are both part of the very core of the artistic intent. It is a question of perspectives. By opening up the form, the listening experience is likewise opened up and a multiplicity of perceptive perspectives becomes possible. This multiplicity is also reflected in the relation between the individual act of participating and the perceptual experience where the individual act can be difficult to discriminate from the totality.

3.1 The context of concert performance

etherSound has been used in two different contexts. As a stand alone sound installation that users can interact with but also in combination with one or several improvising musicians playing acoustical instruments. In this situation, which resembles a traditional concert, the audience is 'playing' the electronic instrument and are given an important role in the development of the form. As can be gathered from the description of the system above, the sonic outcome of a received SMS is fairly strictly controlled. On the individual level, only a limited amount of control over detail is offered, and it is debatable whether *etherSound*

[7] Perhaps *reading* would be a better word to avoid confusion with the musical term interpretation.

can be called an 'instrument' at all. This was however never the intention. It is the *desire* to contribute to the whole that was intended to be the ruling factor, not the individuality of expression or the virtuosity of performance.

An interesting aspect of the concert performance context appears if we compare it to an interactive performance for computer and instrument where the performer influences the output of the computer. In this model the performer and the computer constitute an ontological entity, a closed system that the audience can observe and listen to. However, in *etherSound*, the computer generated sounds becomes the common ground between the performers and the audience, a sonic field of communication and the audience can no longer be disunited from the content.

Whether or not the participants felt they had influence and whether this influence set creative energies in motion within the participant can only be proved, if at all, by performing empirical studies that are beyond my intentions and competence. I can however offer the lightweight, subjective analysis that improvising along with an audience in a way that can be done with this work, is an experience incomparable to traditional group improvisation.

4 Future improvements and further work

The aspect of 'democracy of participation' could be further expanded by also streaming the sound on the Internet, inviting participants anywhere to collaborate. It would also be desirable to allow for simultaneous playback of multiple messages, possibly through the use of several computers, and to add more depth to the interface and allow for 'expert' performers. One thought is to add the possibility to make a voice call to the phone connected to the system and interact in true real time, either by voice or by pressing digits. The text analysis responsible for calculating the life index could be further evolved, i.e. to allow for, and equally reward, typical SMS language such as 'c u 4 dinner 2nite'.

The latency of the system in the first model, measured from when the participant presses the 'send' button to when sound is heard, is in the range of less than a second to a little over two seconds. This may seem long but, in fact, many users commented the fact that they experienced the response time as being short. The second model remains to be tested, but it is fair to assume that the response will be slower and an effort should be made to shorten the response time, possibly by letting the cellular phone interface notify the Java application of new messages.

Since every performance of *etherSound* is 'recorded' in the database, the music can be recreated and altered. I am currently working on a fixed media piece using the data collected during one of the performances of *etherSound*.

5 Acknowledgments

I wish to thank Miya Yoshida who commisioned *etherSound*, Leif Lönnblad for his work, Lund University/Malmö Music Academy for financial support and Vodafone™for financial and technical support in the beginning of the project. Furthermore, I wish to acknowledge the people behind the *gnokii* project and the people behind the library *JavaOSC*.

References

[Adorno,] Adorno, T. W. chapter 12, page 211. In [Adorno, 1962].

[Adorno, 1962] Adorno, T. W. (1962). *Inledning till Musiksociologin (Original title: Musiksoziologie).* Bo Cavefors Bokförlag.

[Ascott, 1990] Ascott, R. (1990). Is there love in the telematic embrace? *Art Journal - Computers and Art: Issues of Content*, 49(3):241–247.

[Barbosa, 2003] Barbosa, A. (2003). Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal*, 13:53–59.

[Barbosa and Kaltenbrunner, 2002] Barbosa, A. and Kaltenbrunner, M. (2002). Public sound objects: a shared musical space on the web. In *Proceedings. Second International Conference on WEB delivering of Music (WEDELMUSIC'02*, pages 9–16. IEEE, IEEE.

[Boulanger, 2000] Boulanger, R., editor (2000). *The Csound Book, Perspectives in Software Synthesis, Sound Design, Signal Processing and Programming.* MIT Press, 2 edition.

[Bourdieu, 1979] Bourdieu, P. (1979). *Distinction: a social critique of the judgement of taste.* Harvard University Press. Translation by Richard Nice.

[Byrne Villez, 2000] Byrne Villez, P. (2000). *Processing Samples with Csound's FOF Opcode*, chapter 15, pages 307–320. In [Boulanger, 2000], 2 edition.

[Clarke, 2000] Clarke, M. (2000). *FOF and FOG synthesis in Csound*, chapter 14, pages 293–306. In [Boulanger, 2000], 2 edition.

[DiMaggio and Useem, 1978] DiMaggio, P. and Useem, M. (1978). Social class and arts consumption: The origins and consequences of class differences to the arts in america. *Theory and Society*, 5(2):141–161.

[Freeman et al., 2004] Freeman, J., Ramakrishnan, C., Varnik, K., Neuhaus, M., Burk, P., and Birchfield, D. (2004). Adaptive high-level classification of vocal gestures within a networked sound instrument. In *Proceedings of the International Computer Music Conference 2004.* ICMA.

[Garnett, 2001] Garnett, G. E. (2001). The aesthetics of interactive computer music. *Computer Music Journal*, 25(1). The MIT Press.

[gnokii, 1995] gnokii (1995). Web page. http://www.gnokii.org/index.shtml.

[J2EE 1.4.1, 2004] J2EE 1.4.1 (2004). *Java Enterprise Edition, API Specification 1.4.1.* Sun, `http://java.sun.com`, 1.4.1 edition.

[J2SE 1.4.2, 2004] J2SE 1.4.2 (2004). *Java Standard Edition, API Specification.* Sun, `http://java.sun.com`, 1.4.2 edition.

[Jordà, 2002] Jordà, S. (2002). Fmol: Towards user-friendly, sophisticated new musical instruments. *Computer Music Journal*, 26(3):23–39.

[OSC, 1997] OSC (1997). Web page. `http://www.cnmat.berkeley.edu/OpenSoundControl/`.

[Rowe, 1993] Rowe, R. (1993). *Interactiv Music Systems: Machine Listening and Composing*. MIT Press.

[Ulyate and Bianciardi, 2002] Ulyate, R. and Bianciardi, D. (2002). The interactive dance club: avoiding chaos in a multi-participant environment. *Computer Music Journal*, 26(3):40–49.

[Wessel and Wright, 2002] Wessel, D. L. and Wright, M. (2002). Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22.

[Wright et al., 2003] Wright, M., Freed, A., and Momeni, A. (2003). Opensound control: State of the art 2003. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*, pages 153–159, Montreal, Canada. NIME-03.

<table>
<tr><td></td><td>H</td><td>E</td><td>L</td><td>L</td><td>O</td><td>,</td><td></td><td>M</td><td>Y</td><td></td><td>N</td><td>A</td><td>M</td><td>E</td><td></td><td>I</td><td>S</td><td></td><td>H</td><td>E</td><td>N</td><td>R</td><td>I</td><td>K</td><td>.</td></tr>
<tr><td>bar</td><td colspan="7">1</td><td colspan="3">2</td><td colspan="5">3</td><td colspan="3">4</td><td colspan="7">5</td></tr>
<tr><td>beats per bar</td><td colspan="7">3</td><td colspan="3">1</td><td colspan="5">2</td><td colspan="3">1</td><td colspan="7">3</td></tr>
<tr><td>subdivision</td><td colspan="2">2</td><td colspan="3">3</td><td colspan="2">2</td><td colspan="3">3</td><td colspan="2">2</td><td colspan="3">3</td><td colspan="3">3</td><td colspan="2">2</td><td colspan="3">3</td><td colspan="2">2</td></tr>
<tr><td>accents</td><td></td><td>></td><td></td><td></td><td>></td><td></td><td></td><td></td><td>></td><td></td><td></td><td>></td><td></td><td>></td><td></td><td>></td><td></td><td></td><td></td><td>></td><td></td><td></td><td>></td><td>></td><td></td></tr>
</table>

Table 2: Rhythmic distribution of notes in Instrument A.

<table>
<tr><td></td><td>H</td><td>E</td><td>L</td><td>L</td><td>O</td><td>,</td><td></td><td>M</td><td>Y</td><td></td><td>N</td><td>A</td><td>M</td><td>E</td><td></td><td>I</td><td>S</td><td></td><td>H</td><td>E</td><td>N</td><td>R</td><td>I</td><td>K</td><td>.</td></tr>
<tr><td>voice 1</td><td colspan="3">E</td><td colspan="3">O</td><td colspan="3">Y</td><td colspan="3">A</td><td colspan="3">E</td><td colspan="3">I</td><td colspan="3">E</td><td colspan="4">I</td></tr>
<tr><td>voice 2</td><td colspan="4">Y</td><td colspan="4">A</td><td colspan="4">E</td><td colspan="4">I</td><td colspan="4">E</td><td colspan="5">I</td></tr>
<tr><td>voice 3</td><td colspan="5">A</td><td colspan="5">E</td><td colspan="5">I</td><td colspan="5">E</td><td colspan="5">I</td></tr>
<tr><td>voice 4</td><td colspan="8">I</td><td colspan="8">E</td><td colspan="9">I</td></tr>
</table>

Table 3: Influence of vowels on four consecutive voices of Instrument B.