6-3-2004

The Spell of Speech

Bruno Ruviaro
Santa Clara University, bruviaro@scu.edu

Follow this and additional works at: http://scholarcommons.scu.edu/music

Part of the Composition Commons

Recommended Citation
http://scholarcommons.scu.edu/music/20

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

This Thesis is brought to you for free and open access by the College of Arts & Sciences at Scholar Commons. It has been accepted for inclusion in Music by an authorized administrator of Scholar Commons. For more information, please contact rscroggin@scu.edu.
THE SPELL OF SPEECH

A Thesis

submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Master of Arts

in

Electro-Acoustic Music

by

Bruno Tucunduva Ruviar

DARTMOUTH COLLEGE

Hanover, New Hampshire

June 3rd 2004

Examining Committee:

Jon Appleton (chair)

Eric Lyon

Ioana Chitoran

Carol Folt (Dean of Graduate Studies)

http://eamusic.dartmouth.edu/~bruno

brunoruviaro@gmx.net
Abstract

This text is a complement to the musical composition *The Spell of Speech*, a piece for actress, electro-acoustic sounds and live-electronics composed in 2004 as part of the requirements to conclude the Master’s degree in Electro-Acoustic Music at Dartmouth College. The core of this research is the relationship between speech and music composition. The first chapter presents a theoretical framework comprising evolutionary musicology and music cognition. The second chapter contains a historical survey of electro-acoustic music in which human voice plays a fundamental role. Finally, the third chapter analyzes the composition *The Spell of Speech*, presenting the main techniques used and the musical ideas involved.
Preface

This text is about music and language. More specifically, it is about the relationship between spoken language and music from the standpoint of a composer.

Even though my interest in this subject dates back at least five years, it was during these last two years (2002-2004)—while enrolled at the Master’s program in Electro-Acoustic Music at Dartmouth College—that I could systematize and further develop the ideas presented here. This text was written during the compositional process of the piece *The Spell of Speech*, which is the center of my current research. This piece is scored for one actress, electro-acoustic sounds and live-electronics, including a variety of compositional approaches based on speech.

This written thesis should be seen as an accompaniment to *The Spell of Speech*. In the following pages, I discuss in detail some fundamental theoretical ideas that are involved in the interdisciplinary research on the music-language relationship, as well as presenting an historical survey of contemporary music in which the interplay between speech and music structure is a dominant element. Specific speech-related composition techniques are also described and discussed.

The first chapter presents an outline of current research in music cognition and evolutionary musicology concerning similarities and divergences between music and language. Special focus is given to the discussion of possible common origins for music and language, a hypothesis that is based on the analysis of their basic shared features. The problem of the identification of fundamental units in music is analyzed in the light of the analogue question in the language domain. General concepts and their definitions are borrowed from linguistics when needed, not only to clarify certain theoretical issues but
also to demonstrate how compositional thinking can benefit from a better understanding of the structures behind human speech.

After introducing some major questions that naturally arise in compositional projects involving human voice in general (and speech in particular), the second chapter presents a historical overview of some fifty years of electro-acoustic composition focusing on pieces that brought to light such questions. Pieces, composers and compositional trends are presented in chronological order. The chosen compositions are not thoroughly analyzed, for this would fall out of the scope of the present work. Instead, I tried to point out what is special in each piece in the relationship between composition and speech. The end of the chapter provides a table summarizing composers, pieces and significant dates.

Finally, the third chapter goes into details of my own compositional process. A number of techniques utilized in the creation of *The Spell of Speech* are described. Influences from historical pieces and ideas are pointed out, as well as the correlation between speech composition techniques and phonologic and prosodic issues. Audio and video recordings of the first performance are enclosed for reference.

*The Spell of Speech* was premiered on May 16th 2004 at Rollins Chapel (Dartmouth College, New Hampshire, USA). It was performed by actress Hannah Chodos. Live-electronics and electro-acoustic sounds were controlled and diffused by myself.

Bruno Ruviaro
Acknowledgements

I am very much indebted to all those people who helped me in some way to have this work done. Jon Appleton, my advisor, for his invaluable help and support in all aspects, academic and non-academic, throughout these two years I have lived at Dartmouth. Eric Lyon, Larry Polansky, Ioana Chitoran, Charles Dodge, Daniel Levitin and Jean-Claude Risset for their teaching and insightful conversations. Theresa Ciambra, for all the priceless help in every fundamental detail and organizational planning. Stefan Tomic, for having written the Perl script I am using in the voice analysis/resynthesis process. Hannah Chodos, for her dedication and artistic commitment. Percussionist Fabio Oliveira, for his wonderful participation in my final concert. Yuri Spitsyn, for helping me in the DVD creation. My friends Masaki Kubo, Steve Pierce, Yuri Spitsyn, Bill Haslett, Irina Escalante and Owen Osborn, for their extraordinary help in putting together the final concert. My parents and Tânia Lanfer Marquez, for more than obvious reasons.

Bruno Ruviaro

Hanover, May 26th 2004
# Table of Contents

Abstract ........................................................................................................... ii
Preface ........................................................................................................... iii
Acknowledgements ....................................................................................... v

Chapter 1 ........................................................................................................ 1  
  1.1 Introduction ........................................................................................... 2  
  1.2 The Musilanguage model ...................................................................... 3  
  1.3 Evolutionary musicology ...................................................................... 6  
  1.4 Music cognition experiments ............................................................... 8  
  1.5 The problem of structural unit definition ........................................... 13  
  1.6 An overview of speech prosody ............................................................ 24  

Chapter 2 ....................................................................................................... 30  
  2.1 General considerations about the use of voice in electro-acoustic music  31  
  2.2 Historical Overview ............................................................................ 39  

Chapter 3 ....................................................................................................... 56  
  3.1 The Spell of Speech ............................................................................... 57  
  3.2 Techniques I: Analysis and resynthesis ............................................... 58  
  3.3 Techniques II: composition on speech pitches ..................................... 69  
  3.4 Techniques III: the live-electronics ..................................................... 70  
  3.5 The Spell of Speech: formal organization .......................................... 80  
  3.6 The concert .......................................................................................... 83  
  3.7 Future work .......................................................................................... 85  

APPENDIX A (score) ...................................................................................... 87  
APPENDIX B (Perl scripts) ........................................................................... 94  
APPENDIX C (Csound files) ......................................................................... 103  
APPENDIX D (l’Attente, by Roland Barthes) ............................................. 107  
APPENDIX E (posters and program notes) ................................................ 112  
APPENDIX F (pictures) ................................................................................ 116  
APPENDIX G (CD and DVD) ....................................................................... 123  
REFERENCES ............................................................................................... 124
List of Tables and Figures

Table 1: Historical compositions.................................................................52
Figure 1: Sonogram in Audiosculpt; non-averaged partial tracking..................59
Figure 2: Sonogram in Audiosculpt; averaged partial tracking .......................60
Figure 3: Screenshot of Open Music. ..................................................................65
Figure 4: Raw conversion of MIDI data into a score and an elaboration on it .......68
Figure 5: First of two Max/MSP patches used in The Spell of Speech ..............72
Figure 6: Second of two Max/MSP patches used in The Spell of Speech ...........73
Figure 7: Patcher dopplerdelay .......................................................................74
Table 2: Schematized formal structure of The Spell of Speech .........................81
Chapter 1
1.1 Introduction

There is a significant amount of recent research in the quest for understanding the origins of music and language and the nature of their relationship. Studies range from general theories and models describing their possible common origins to specific experimental approaches trying to elucidate the details of such relationship. Although it is hard for someone to object to the striking similarities of these two highly specialized human-specific manifestations (music and language), it is still an open question as to whether these activities share common cognitive structures. In a review on the comparison of music and language, Besson and Schön (2001) list different theories that claim — in contrast with the Generative Grammar Theory — that language may actually be dependent, at least in part, on other general cognitive functions.

In this chapter we will describe the musilanguage model (a theory that suggests a common ancestor for music and language — see Brown, 2001), as well as summarize basic considerations of evolutionary theories of music (Huron, 2001; Cross, 2003). Further comparisons between the two domains are examined from a cognitive perspective, with a special focus on the problem of an adequate definition of structural units in music. Recent experimental data on the inter-relation between language (more specifically, speech) and music are also reviewed. A list of definitions of a few linguistic terms related to the study of prosody is also provided at the end of this chapter.

The purpose of this section is to present a broad range of ideas in which the music-language relationship is shown from different perspectives. We often go into detail in the presentation of each theory, and naturally we have tried to be as rigorous as possible in this study. However, it should be noted that this is not a scientific work. Drawn from the
existing literature, the body of research presented here represents an interdisciplinary theoretical framework that surrounds our compositional work.

1.2 The Musilanguage model

Guided by the belief that music and language possess strong biological similarities as well as important divergences, Steven Brown (2001) developed the Musilanguage model (ML) as an attempt to go beyond the limits of metaphor in comparing the two fields. Their most important points of convergence, according to him, are combinatorial syntax and intonational phrasing (Brown, p. 273).

Combinatorial syntax is the capacity of manipulation of discrete units to generate higher-order units that are in some way “meaningful”. These units, in both speech and music, are made up of sonic events. This ability to create permutations out of a basic set of smaller units is not exclusive to humans, though. Marler (2001) points out that many birds do have this kind of learned ability to reuse and recombine sonic elements in different sequences, in the same way as we unconsciously combine phonemes to speak words. This “phonocoding” or “phonological syntax” capability (Marler, 2001, p. 36) in birds does not mean, however, that they are actually “speaking”: “Song sequences are not meaningfully distinct, in the referential sense; they are rich in affective content, but lacking in symbolic content” (Marler, 2001, p. 40). Interestingly enough, phonocoding is present in birds but not in non-human primates, what leads that author to consider bird sonic behavior closer to human music than that of our direct relatives apes and monkeys.

Besides the ability of using sound elements as building blocks for complex combinatorial structures, Brown refers to intonational phrasing as the second major point
of convergence between language and music. This is defined as the “modulation of the basic acoustic properties of combinatorially organized phrases for the purposes of conveying emphasis, emotional state, and emotive meaning” (Brown, 2001, p. 273). It can take place at a local and at a global level, a differentiation mostly based on the time span of modulations. Small modulations in short intervals of time are seen as local and affecting individual elements within the phrase structure, whereas broader modulations affecting larger time intervals are analyzed as global expressive phrasing. It should be noted, however, that Brown seems to associate intonational phrasing with purely emotional meaning function, whereas specific linguistic studies on prosody clearly state that intonation is more than that. In her book on discourse analysis, Wennerstrom (2001) writes about intonation in the following way: “(…) speakers can manipulate their pitch on particular words, phrases, or even topic-sized constituents to convey meaning about relationships in discourse” (p. 17), and also that “prosodic features associate with particular constituents in discourse in a rule-governed and meaningful way” (p. 7).

Furthermore, as Brown notes, those shared aspects of music and language (combinatorial syntax and expressive or intonational phrasing) would develop into two higher levels, namely, the phonological and the meaning level. The former is basically the phonocoding capability explained above. In this phonological level, Brown (p. 275) draws a parallel between language and music by relating “phonemes” to “pitches” and “morphemes” to “motifs”. The problems created by this “note-based” hierarchy (which treats the “note” as the basic unit of music in the same way that linguists consider the phoneme the basic unit of language) will be discussed in more detail later in this chapter. On the other hand, the meaning level is, broadly speaking, the “symbolic content” that
Marler referred to, but Brown elaborates this concept in a more detailed manner. The idea of sound meaning as a continuum oscillating between two poles (referential versus emotive meaning) is central to the ML model. Music and language would differ more in emphasis than in kind (Brown, p. 275), and mixed manifestations like poetry or verbal songs would appear in the middle of this continuum. Referentiality would be characteristic of language, while emotivity would be characteristic of music. However, the concept of “emotive meaning” is never clearly defined in Brown’s text. One could draw a parallel of it with Umberto Eco’s semiotic approach in his book *The Open Work* (1989), in which he posits that everyday language tends to function best by avoiding ambiguity, while in arts ambiguity is desired and even expected (poetry being an interesting case where language itself has to be “forced” towards ambiguity). Thus we are assuming that emotive meaning — as one end of the continuum proposed by Brown — is related to the notion of ambiguity and non-referentiality (or non-denotative meaning).

Finally, the ML model is based on the assumption that “common features of these two systems are neither musical nor linguistic but musilinguistic, and that these properties evolved first” (Brown, p. 277). This musilinguistic stage is divided into two phases: in the first one, lexical tone plays a fundamental role, being the initial step for the discreteness of pitch units within broad semantic meaningful vocalizations. In a second step, combinatorial phrase formation and intonational (expressive) phrasing are developed. After this stage, music and language would eventually split into specialized human activities, without losing these basic shared ancestor features while giving rise to other exclusively musical or linguistic features.
1.3 Evolutionary musicology

The ML model is linked to a body of research identified as the evolutionary theories of music. There are many other evolutionary approaches to understanding the origins of music. Huron (2001) identifies up to eight broad lines of such theories, ranging from sexual selection to motor skill development and social cohesion, among others. Some of these approaches are not necessarily mutually exclusive, but rather they differ in focusing more on one or another possible explanation as to why and how music would have evolved. Other approaches are more controversial. For example, Steven Pinker believes that music has no intrinsic evolutionary value, being merely some sort of “by-product” of other important adaptive traits (Cross, 2001). Another example in a different direction is Geoffrey Miller’s theory of evolution of music through sexual selection (Miller, 2000; 2001a; 2001b): he departs from Darwin to develop the hypothesis that, ultimately, “music evolved and continues to function as a courtship display, mostly broadcast by young males to attract females” (Miller, 2001b, p. 354).

However, one central issue for further development of theories of this kind is what types of evidence one can present to support them. Huron (2001) lists four general types, namely: genetic, archaeological, ethological and neurological evidences. The first one is perhaps the more controversial, since to date there is no proof for the link between music and genes (Huron, 2001, p. 48).

The other kinds of evidence that furnish data on a number of different aspects of musical activity are:

- Archaeological findings that prove the existence of music dating back to the most ancients human settlements (Huron, 2001, p. 48);
• The ubiquity of music across all human cultures may be an anthropological evidence that supports an evolutionary argument (Huron, 2001, p. 50);
• Empirical data describing brain structures responsible for music cognition can provide important insights about how music works in humans’ minds, and what is innate and what is the result of learning. Experimental comparisons can help in further elucidating what is common and what is specific in music and language processing.

Hauser et al. (2003) proposed an interesting theoretical framework for comparative studies with animals. One of the major advantages of comparative studies is that, with animals, one can control in detail the degree of lifetime musical exposure to an extent not possible with humans. Also, Hauser argues that if any perceptual feature supposedly specific to music is found in nonhuman animals, it is unlikely that it was part of an adaptation for music, “and must instead represent a capacity that evolved for more general auditory scene analysis” (Hauser, 2003, p. 665). For example, an experiment is described in which rhesus monkeys showed octave generalization (Hauser, 2003, pp. 665-666). This would mean that octave generalization is an innate perceptual constraint not exclusive to humans. This kind of experiment represents a valuable path for future research not only to achieve a better understanding of the specificity of musical activities but also in the disentangling of the “nature versus nurture” problem.

Another evolutionary approach is the social bonding hypothesis. According to this view, one of the main adaptive values of music would be its capacity of creating or strengthening social cohesion among groups. Huron (2001) enumerates several arguments and evidence in support of this view. One example is his comparison between
two different mental disorders: Williams Syndrome and Asperger Autism. He points out the link of symptoms such as low sociability and low musicality in the latter, as opposed to high sociability and musicality in the former (Huron, 2001, pp. 54-55). Another example is the idea of music as a large-scale bonding mechanism more efficacious (for groups) than the rather interpersonal “grooming, gossiping, courting and conspiring”. Music, more than language, would have “allowed humans to live in larger groups with their attendant complex social relations” (Huron, 2001, p. 53).

To sum up, we have discussed a) the musilanguage model, which tries to explain how music and language would be related in terms of a common origin, and b) other evolutionary accounts that try to understand why music would have evolved and what could be its adaptive value. After reviewing some of these different accounts, we believe that the most promising theories are those pointing to an integrative view of the multiple functions that music may have. In other words, it seems to us that future research should try to integrate the multiplicity of functional levels often displayed by musical activities in a unifying theoretical framework, leading to what Brown has called “multilevel selection models” (Brown, 2001, p. 297).

1.4 Music cognition experiments

In their text “Is music autonomous from language? A neuropsychological appraisal”, Patel and Peretz (1997) conclude, “The evidence reviewed (...) suggests that music and language are not independent mental faculties, but labels for complex sets of processes, some of which are shared and some different. Neuropsychology allows the empirical delineation of the boundaries between these domains, as well as an exploration of their
overlap” (p. 208). It is this kind of research that can be helpful for further elaboration of theories that account for common origins of language and music, such as the musilanguage model described earlier. Let us now examine a couple of specific experiments that have been conducted in this area.

Words, music and brain imaging. Besson and Schön (2001) conducted experiments in order to detect event-related brain potentials (ERPs) elicited by different degrees of musical and linguistic events semantically unexpected or incongruent within a given context. It was known from previous linguistic experiments that unexpected words ending a sentence elicit in the brain a negative voltage variation component called N400. By testing musicians and non-musicians with familiar and unfamiliar melodies ending with or without “unexpected notes”, they found that musical incongruity within a tonal context causes a P600 component.\(^6\) This is seen as a very similar behavior in relation to the linguistic N400 component. Further experiments combining melody and words (by using opera themes) showed that incongruous words, even when they are being sung, do elicit a N400 component, while incongruous notes (no matter the “congruity” of the word accompanying it) always elicited a P600 component.\(^7\)

According to the authors, “this finding provides a strong argument in favor of the independence (i.e., the additivity) of the computations involved in processing the semantic aspects of language and the harmonic aspects of music” (Besson and Schön, 2001, p. 246). However, it may be possible that other levels of language and music processing share general cognitive processes. This would be the case for detecting syntactic and harmonic violations.\(^8\) The results of studies comparing ERPs associated to grammatically unexpected violations within sequences (language) and presentation of
chords from distant tonalities within a simple sequence (music) showed that “the effects of violation of syntactic and harmonic expectancies were not significantly different” (Besson and Schön, 2001, p. 250).⁹

Another interesting experiment was made by Besson and Schön to study the effect of temporal ruptures in musical and spoken sequences. Familiar and unfamiliar phrases were presented. Some of them had the final note or word delayed by 600 ms, thus breaking the expected flow of events in listeners. Different and complementary brain imaging methods were used to get a good combination of temporal resolution and good localization of brain activations. Briefly, results have shown that “similar brain areas were activated by temporal violations in both language and music”, leading the authors to conclude that “taken together [the] results suggest that processing temporal information in both language and music relies on general cognitive mechanisms” (Besson and Schön, 2001, p. 254).

Some cognitive processes may actually be of a much broader nature, indeed. An experiment conducted by Ramus and colleagues (2000) suggests that there are similarities in the perceptual mechanism of newborns humans and tamarins: under experimental conditions, both species were able to discriminate two rhythmically distinct languages (Dutch and Japanese). This implies that some aspect of speech processing might not be unique to humans. Speech rhythm is likely to be the key for explaining such results, and “although human infants may be equipped with a capacity to discriminate languages on the basis of rhythmic cues, the presence of this capacity in nonhuman primates that lack language suggests that it evolved for more general auditory purposes” (Hauser, 2003, p. 667).
**Rhythm in language and music.** Patel and Daniele (2003) have designed an experiment in which they apply a quantitative measure to compare musical and speech rhythmic patterns. They have compared English and French languages with English and French music themes from a specific era of music history. The hypothesis was that global patterns of rhythmic organization in music could be related to the composer’s native language rhythmic structure.

French and English are respectively recognized as a “syllable-timed” language and a “stress-timed language”. This basically means that there is quantitative rhythmic difference between these types of languages. These differences would exist partly because of variability of vowel durations, and indeed there is recent evidence for this claim: a special kind of measurement has been used by linguists to compare rhythm in English and French utterances, and the data show that significant differences were found (see Patel and Daniele, 2003, for details).

Based on that, Patel and Daniele (2003) designed their experiment using English and French instrumental music. They applied the same quantitative measure of speech rhythm taken from linguistics experiments\(^{10}\) to the music of 16 composers from those two countries. Their results showed that “English and French classical music have significantly different nPVI values, and that this difference is in the same direction as that observed for language” (p. B42). Thus the claim is that the particular rhythmic organization of a composer’s native language can influence the rhythmic structure of one’s compositions.\(^{11}\)

While appealing, however, this idea should be viewed with caution. The assumption that notes correspond to syllables was the basis of their experiment (“Vowels form the
core of syllables, which can in turn be compared to musical tones”, p. B37). This association is doubtful—the use of “note-units” in music is far more elastic and flexible than the use of syllables in speech. So such association could lead to mistaken comparisons between supposedly similar experiments. Besides, the musical material chosen came out of a quite specific historical period (all composers studied were born between 1835 and 1899). The authors consider the period limitation an advantage for their research. Such limitation comes together with stylistic issues, since all composers were in some way part of an age of national styles development, “in which speech prosody has been thought to play a role” (p. B39). The hypothesis is that composers can either “consciously or unconsciously draw on these [linguistic rhythmic] patterns in weaving the sonic fabric of their music” (p. B39) or be “influenced by music [they] hear[d] as children, such as popular and folk songs whose rhythms bear the imprint of their associated linguistic texts” (p. B39). On the other hand, considering the several layers of abstraction involved in the learning and use of musical techniques and development of individual compositional styles, the differences detected in Patel’s research could well be a consequence of specific, inherently musical problems faced by groups of composers. In short, the study is basically correlational: assuming its methodology is valid, it shows that two different comparisons appear to point to the same direction, but there is no direct evidence that they result from the same phenomenon.
1.5 The problem of structural unit definition

Linguists generally agree upon the concept of the phoneme as the most basic sound unit responsible for language structure. In an analogous fashion, a considerably large number of researchers take for granted that the “note” would be the musical unit, which would be the correlate of the phoneme, the syllable or the word, according to different research goals. As examples, one can see how such assumptions appear explicitly or not to different degrees in Aiello, 1994; Krumhansl, 2000; Besson & Schön, 2001; Brown, 2001; Patel, 2003; Patel and Daniele, 2003; Schwartz et al., 2003. Besides the work by Patel and Daniele (2003), from which we have already quoted in the previous section, let us extract a few more quotations from two other studies to make our point clearer.

- “(…) In both language and music, the phrase is the basic unit of structure and function.” (Brown, 2001, p. 273)
- “(…) The use of discrete building blocks and the generation of higher-order structures through combinatorial rules is a major point of similarity between music and language.” (Brown, 2001, p. 273)
- “(…) [The phonological level] is governed by a type of phonological syntax in which discrete acoustic units (phonemes, pitches) are combined to form functional units (morpheme, motifs).” (Brown, 2001, p. 275, our emphasis)

From the first two more open and general statements, Brown follows up implying that the acoustic properties of the fundamental units are pitch sets, intensity values and duration values, in music, whereas in language they are phonemes and phonological feet. Even though it appears somewhat mixed with the idea of “pitch”, it seems obvious that he is thinking of the concept of “note” to define a musical unit.
Besson & Schön (2001) leave the question a bit more open:

- “The segmentation of the sound continuum in discrete units (pitches or phonemes) is found in all music and languages. However, if we can eventually make an analogy between phonemes and intervals of a musical scale, we must also be aware of their differences.” (pp. 235-6)

- “Furthermore, some of the perceptual properties of the basic elements in music have no equivalent in language, as, for instance, the fact that octaves are perceived as equivalent in almost all cultures.” (p. 236)

We believe that even though this can be partly valid for some specific music styles, a more accurate look at the diversity of musical “facts” around the world can question this assumption. One can think of the minor importance of the traditional “note” concept in such different musics as many of Gyorgy Ligeti’s pieces, on the one hand, and the Tuvan throat-singers from Asia, on the other hand. In the same way, the definition of other higher-level units such as the musical “phrase” may not be so self-evident and therefore would benefit from a broader understanding of diverse musical manifestations that are not primarily note-structured. What would be a phrase, for example, in the opening of De Natura Sonorum, by French composer Bernard Parmegiani?

In order to continue this discussion, let us take a look at a few aspects of what is considered to be the basic structural unit in the language domain, namely, the phoneme.

“A phoneme, being a bundle of distinctive features, proves to be an ever important but derivative unit, a complex, simultaneous construct of a set of elementary concurrent units. It can be compared in this respect to the syllable, which is a constructive complex unit within the verbal sequence.” (Jakobson & Waugh, 2002, p. 29)
“Such components [that] enable the perceiver to distinguish zeal from deal, shows from showed, and sailor from tailor, or mad from bad, mobility from nobility, and fashion from passion are termed ‘distinctive features’.” (Jakobson & Waugh, 2002, p. 8)

Distinctive features are aspects of the sound related to its spectrum and timbral qualities, as we can see by the very words used to describe them in terms of oppositions: grave-acute, nasal-nonnasal, compact-diffuse, voiced-unvoiced, strident-mellow and so forth. Formants have a special role in this aspect: “A formant is a peak of energy in a spectrum, which can include both harmonic and inharmonic partials as well as noise. Formant peaks are characteristic of the vowels spoken by the human voice” (Roads, 1995, p. 296. See also Cook, 1999, especially chapters 9, 11 and 12).

Being basically defined according to their spectral qualities, phonemes are in some sense “atemporal”. In principle they can be defined without concern for their duration. As an example, the sound of the phoneme /a/ is defined according to its spectrum contents, and not according to its actual duration (whether it is 50 or 5000 milliseconds long).

Time relationships do matter if we are studying prosody: consonant and vowel length can be said to be a property of the prosodic structure. Thus, for the strict, structural definition of phoneme, time does not play a major role. The only time-component inherently present in such definition is the one imposed by the nature of the phonemes themselves: roughly, there are phonemes that can be extended in time and others that simply cannot. For example, sounds such as /sl/, /ʃl/, /a/ and /i/ can be sustained in time, whereas others like /p/, /t/, /k/ cannot. This level of temporal definition is part of a phoneme’s own nature. Duration is not able to fundamentally change the phoneme itself.
As the basic building block of a system, meaning discrimination comes from phoneme discrimination. Two words differing by one single phoneme are heard apart because of the aural cues provided by the binary opposition between the two different phonemes.

Now, is there such a concept in music? What basic component, when replaced by another distinct component of the same hierarchical level, would produce a significant change in a musical sequence, enough to cause a change in “meaning”?

As we have seen, in the “note paradigm”, used explicitly in many music-related scientific studies, and implicitly in many other ones, the note has been commonly taken as the fundamental unit. We believe that this approach is misleading or, at least, too limiting. It refers to an understanding of music that is limited to one type of musical organization that, even though culturally widespread today, is not the only one. We can use the term “tonality”, in a broader sense, to refer to this widespread musical system and its derivations. By derivations we mean other kinds of musical practice that are not strictly tonal in the historical sense but have a strong relationship with that historical system and share significant assumptions with it (for example, some hybrids between tonality and modality).

In any case, tonality is representative of a fraction of the total musical scenery in the world: the Western tradition. The growth of a cultural industry in the twentieth century has been responsible for the diffusion of highly standardized musical patterns, habits and attitudes throughout the world, often market-oriented and profit-driven. Tonal principles, sometimes colored by non-western musical influence, have provided the structural basis for the majority of musical “styles” that are available today in virtual or real music stores.
Even in the Western world, tonality represents a snapshot of a few centuries of musical production. The note-based assumption, however, comes at least in part from this tradition of looking at tonality or its derivates as the “natural” way of thinking about music. The note is identified as the almost-indivisible unit of music. Establishing a parallel with the phoneme, a “note” is defined by its placement in the frequency domain, or, more specifically, in an abstract grid called “scale”. Note names are the labels used to differentiate one note from the other. Thus, D is different from G, but two D’s in different registers are considered to belong to the same pitch class.

“We say that pitch has two separate aspects, a chroma, or color, telling where it stands in relation to others within an octave, and a tone-height telling which octave it belongs to. The term chroma allows for continuous variation of pitch, while pitch-class refers to the same property when we use only a finite number of discrete pitches within an octave.” (Hall, 1991, p. 399).

The note-based approach may disregard the time of individual notes, just as the phonemes do not depend on their actual duration in order to be defined. There is no explicit agreement as to whether note duration is part of the unit “note” or not. Thinking about tonal music, the foundation of the whole rhythmic building is the organization of the duration of individual events in a very close relationship with the pitch structure. One might say that, besides pitch, rhythm is the other fundamental organizational principle of music. Tonal music developed a very strict hierarchy for rhythmic and duration organization, mainly based on regular, notated proportions and subdivisions of individual events related to a constant pulse. So if we disregard duration and rhythm as part of the musical “unit”, we are disregarding a strong part of tonal music structure. On the other hand, if we do include rhythm in the unit definition, how should it be included? Concretely, the note C lasting 2 seconds would be a different unit if compared to the
“same” note C lasting 10 seconds? Or would rhythm and pitch consist of separate, but inter-connected structural domains, in a rough comparison with the way phonemes and syllables relate to metrical stress rules in linguistics?

Different from language, event durations are a fundamental building block of musical structure in a precise and quantified way. Furthermore, most of the time musicians are aware of such time manipulations: they are able to control it consciously. Of course, sounds have their natural time propensity in the same way that phonemes do; compare, for that matter, a string pizzicato to a sustained piano note. In music, however, the temporal behavior of every sound is not merely a function of its physical constraints. Rather it tends to be composed with a high degree of detail. Sounds are not only put together, but temporal relationships are built between sounds in many simultaneous levels.

We could make similar points on how much musical dynamics (correlate of loudness) are used as structural units by some composers and not by others, and whether it should be part of the “musical unit” or not. In fact, notes are not indivisible: frequency, time and amplitude are the three physical parameters that have direct influence on the perception of “note units” and their perceived pitch, duration and intensity. Besides these three acoustically measurable parameters (which, it is worth saying, not always correspond directly to perceptual phenomena), we have timbre as something else situated in a higher level of perception and with no single physical correlate. Timbre would be the sum of all those low-level properties, belonging to a qualitative perceptual sphere rather than a quantitative one (see Hall, 1991, p. 392 for a more detailed discussion).
Timbre is another characteristic that apparently does not count in the definition of a “note”. The “same” note C, played first by a violin and then by a trumpet, will still be called “the same”. This widely accepted “fact” hides an important assumption that must be relativized: this “sameness” is only possible within a musical system in which timbre is a secondary feature, in which it doesn’t play a role in the internal definition of the system itself.

If we think about basic assumptions of tonal music, what is the role of timbre? Is it part of the system or not?

The answer is: it depends. If one considers a musical piece with no specification as to instrumentation, such as The Art of Fugue by Johann Sebastian Bach, it can be argued that timbre is not at all an integral part of the those musical systems. One can also argue that some Mozart keyboard piece remains basically “the same” when played on the harpsichord or on the piano, or even on a synthesizer (a traditionally less accepted interpretation, but still the piece would be “the same”).

On the other hand, one could as easily find examples of the opposite: works in which timbre seems to have a much more intimate connection with the piece itself—if not as part of the structure, at least in the form of a clear “contamination” of the musical ideas by the constraints or capabilities of a given instrument. How much of a cello suite by Bach would be lost if someone attempted to play it on the flute? In the same direction, a solo trumpet cannot play a Chopin prelude without significantly changing the composition itself. Here we are in the domain of instruments’ idiosyncrasies, and it is hard to deny that some compositions are heavily dependent on them. In most of Webern’s music, timbre is a crucial component of composition, not to mention the famous example
from Schoenberg’s Op. 16, *Farben*, which is based on the composition of orchestral timbre (or “harmony-timbres”, so to speak). Electro-acoustic music in general is simply “not transposable” to any other media, precisely because timbre composition is as much a part of its structure as any other musical aspect according to a composer’s own goals.\(^\text{15}\)

This quick survey is enough to demonstrate that the role of timbre is a complex and delicate matter in music in general. Within the limits of tonal system and its derivates, it is at the same time hard and easy to separate timbre of a composition. The suspension of timbre considerations in musical analysis (while analyzing a Bach fugue, for example) may lead to an apparent “core” of the musical system, but actual musical practice makes it clear that timbre is never simply disregarded in any level (composition, interpretation, listening), even when it is not a structural component.

In conclusion, we hope to have demonstrated that even in the case of tonal music (and we are already referring to a great diversity of styles under this single term) the “note” concept as the fundamental unit of music is not only unclear, but it is fundamentally inappropriate. What to say, then, of different musical traditions from other cultures, and of other musical systems used by Western composers?

The question remains open. If we try to keep the music-language comparison, we can put the question this way: within a given musical system, what element, or elements, when replaced, will cause a meaningful perceptual change? And what elements will be perceived as equivalent (“the same”), as if they belonged to the same category but having localized changes or “accents”, and what elements would be opposed (“different”)? The note alone doesn’t seem to be able to account for the many different structural layers of many musical systems.\(^\text{16}\)
We think that the search for a fundamental unit of music must take into consideration the diversity of existing musical systems. There might not be one single unit. Two points of view come into play at this point: that of the composer and analyst and that of the listener.

A composer and a music theorist can develop coherent systems and explanations about how a given piece was composed, and what were the building blocks of that specific compositional project. Here is but one example of other ideas that may be present in the search for basic units of sound and music. Composer Barry Truax talks about the concept of grain as follows:

“What is most remarkable about the technique is the relation between the triviality of the grain (heard alone it is the merest click or 'point' of sound) and the richness of the layered granular texture that results from their superimposition. The grain is an example of British physicist Dennis Gabor's idea (proposed in 1947) of the quantum of sound, an indivisible unit of information from the psychoacoustic point of view, on the basis of which all macro-level phenomena are based. In another analogy to quantum physics, time is reversible at the quantum level in that the quantum grain of sound is reversible with no change in perceptual quality. That is, if a granular synthesis texture is played backwards it will sound the same, just as if the direction of the individual grain is reversed (even if it is derived from natural sound), it sounds the same. This time invariance also permits a time shifting of sampled environmental sound, allowing it to be slowed down with no change in pitch. This technique is usually called granulation.” (Truax, online resource)

A listener, on the other hand, may not know anything about the composer or about music theory at all, and this listener will understand music in his own way. How much of the composed structure is perceived by the listener? Does this perception need to be conscious? Language “users” don’t need to know linguistics to be able to communicate, to speak and to understand speech. In language, as well as in deeply rooted musical traditions, this seems to be possible due to a high degree of cultural agreement on the nature of the code itself and its uses (be it language or music). What happens when we
have a situation of rapidly changing musical systems? And when listeners are exposed to more than one single, widely accepted system? How do composers deal with the musical memory of their listeners, both locally (within a piece) and globally (historically)?

The musical mind seems to adapt itself to whatever musical system it has been exposed to for a long time (Tillmann & Bharucha, 2000). The alleged units for one system may not be relevant to another system. In terms of learning and interacting with new musical situations, the musical mind would have to learn as fast as the changes in musical systems occur. In the global cultural environment today, it is virtually possible to be exposed to radically different music styles within minutes.

Language seems to be a much more stable system than music. Even though languages do change through the centuries, their basic constituents remain unaltered, or at least they change very slowly. In music, the speed of structural changes throughout the centuries is much faster. Musical systems, at least in Western music, are continually in flux. Actually, a single “individual” in a musical collectivity has, to some extent, the “power” to make significant alteration in the definition of music (the system), influencing other music makers and listeners. In language, on the other hand, a single individual is much more tied to social constraints and norms, not having the power to introduce, alone, significant changes in the language structure of its social group (see Menezes, 1993a). Even though music making is also bound by social constraints and cultural tradition, it seems to have the capability of changing at a much faster rate than languages do.

As a further analogy, we can speculate that in music, today, creation is possible at all “levels”: from the phonemic and lexical levels to syntax and discourse organization. Language, for its purposes, keeps the first levels “locked up”: we are not able, or allowed,
to create our own phonemes and syllables; we are marginally able to create our own words and, apart a few possible violations of grammatical rules, we are free to combine words and sentences at our will to develop a discourse (Jakobson, 1963). In music, on the other hand, this hierarchy is a lot more unstable (Menezes, 1993a). Unless a musician “locks” himself (unconsciously or by conscious choice) in a very specific and well-defined musical tradition, we can say that one is relatively free today to compose music “from scratch”. This does not mean that one can compose without any influence from past music. On the contrary, it is exactly through the understanding of the present state of musical facts around the globe (especially including the transformations brought by the twentieth century) that we are today in position of rethinking music from its basics, from its inner structural constituents to its social roles. In this manner, any possible sound is a candidate for musical use, and definitions such as “noise” opposed to “musical sounds” have become obsolete. Sound objects can be chosen from the “natural” environment or can be carefully composed and synthesized, according to virtually any organizational principle that a musician might imagine. In other words, it is as if composers had the ability to invent or define their own phonemes, their own syllables and words, and then their own discourse.\textsuperscript{18} Of course the extent to which this hypothesis can be accepted is not unlimited. Humans are bound to their cultures and historical periods; we may not see a unifying agenda linking all musical styles today, but this may be identified in the future. At this point we also get to the perennial question of musical universals. This apparent creative freedom that we identify today may well be bound by human-specific biological constraints. If there are musical universals, what are they? Can they reconcile the incredible diversity of musics being created today?
1.6 An overview of speech prosody

To conclude this chapter, we will make a brief incursion to the domain of prosody by means of elaborating a basic prosodic glossary or list of definitions and concepts. The concepts to be exposed below are in more or less direct relation with the kind of vocal composition that we have been developing. Borrowed from linguistic studies, these ideas are useful in the creation of compositional tools and strategies to deal with human voice, especially speech. Presented here in their more theoretical form, it is in our actual composition (refer to the enclosed recording) that practical applications of such concepts may be seen (or rather, “heard”) in what we could call a “prosodic composition”.

The definitions presented below were extracted from Wennerstrom (2001), except when noted.

**PROSODY:** “General term encompassing intonation, rhythm, tempo, loudness, and pauses, as these interact with syntax, lexical meaning and segmental phonology in spoken texts.” (p. 4)

**INTONATIONAL MEANING:** “(…) speakers can manipulate their pitch on particular words, phrases, or even topic-sized constituents to convey meaning about relationships in discourse.” (p. 17); “Intonation is not derived automatically from the stress patterns or syntax of an utterance.” (p. 18)

**INTONATION UNIT:** There is divergence as to what should be considered a unit in intonational analysis. “Should the intonation contour of an entire phrase be interpreted as a single, meaning-bearing unit? Is it possible to identify smaller units as meaningful? Where does exactly a unit start and stop?” (p. 17); Making an adaptation of
complementary previous work by other researchers, the model proposed by Wennerstrom (2001) considers that “the intonation of an utterance can be analyzed as series of high and low tones, each conveying a particular meaning in discourse.” (p. 18) Her model includes four main categories:

- pitch accents
- pitch boundaries
- keys
- paratones

**PITCH ACCENTS:** “Pitch accents are the various tones associated with lexical items that a speaker decides are especially salient in the information structure of the discourse.” (p. 18-9); In Bolinger (1989) terms,

> “Intonation is the inclusive term, referring to all uses of fundamental pitch that reflect inner states. (It might be clearer to put this negatively: all uses that are not associated with the arbitrary tone of a tone language like Chinese or Mazatec.) **Accent** is intonation at the service of emphasis. In the shapes of the profiles it makes certain syllables stand out in varying degrees above others, revealing to our hearer how important the words containing them are to us, and revealing also, by the buildup of accents, how important the whole message is.” (p. 3)

**PITCH BOUNDARIES:** “Pitch boundaries are the pitch configurations at the ends of phrases, accompanied by a lengthening of the final syllables.” (p. 20)

**KEY:** “Key is the choice of pitch a speaker makes at the onset of an utterance to indicate the attitude or stance toward the prior one.” (p. 23)

**PARATONES:** “Paratones, the expansion or compression of pitch range, function as a kind of intonational ‘paragraphing’ to mark topical junctures.” (p. 24); Wennerstrom leaves open the hypothesis that key and paratones might be part of the same
phenomenon: “(...) they are both constituent-initial pitch range shifts that indicate the degree of continuity, or lack of it, between two units.” (p. 278)

**INTONATIONAL PHRASE:** This concept is proposed as the “unit” of intonation. “(...) a more or less continuous pitch contour with, at minimum, an initial key, a number of pitch accents, and a pitch boundary.” (p. 28); Correspondence between intonation and syntax is by no means absolute. (...) [A]ny subcomponent of a grammatical sentence can be uttered in a single intonational phrase.” (p. 29)

Common properties of intonational phrases (p. 29):

- “Speakers often pause between intonational phrases (physical requirement of breathing)”;
- “Tempo is faster at the beginning and slows toward the end” (final lengthening); “final syllables tend to be elongated”;
- “For some speakers, the intonational phrase ends with a change in voice quality as well, such as a creakier or breathier voice”;
- “Gradual lowering or “declination” of pitch throughout the duration of an intonational phrase as the speaker’s air pressure diminishes”; “pitch is reset back to a higher level at the onset of a new intonational phrase”;
- Important consequence: “although a speaker’s pitch may rise and fall to convey discourse meaning, the gradual trend of the phrase is usually from higher to lower. Thus, pitch accents within the same intonational phrase may perform a similar discourse function even though they do not have the same absolute pitch.”
• “It may coincide with syntactic constituents, however syntactic structure may be obscured by ellipsis, hesitations, repairs, and other fast-speech phenomena” (p. 31).

**PITCH ACCENTS AND STRESS PATTERNS:** Even though both stress and intonation rely upon pitch, loudness and duration changes in speech, they are not the same thing, for “stress is a phonological characteristic of lexical items and is largely fixed and predictable, whereas intonation (...) can be altered depending on the discourse role played by the constituents with which tones are associated.” (p. 47); “Pitch accents link to stressed syllables. Stress patterns of words are fundamental to the association between intonation and text.” (p. 46); “Rhythm provides an underlying hierarchical structure upon which stress is built, while intonation components are associated with the high points of these rhythmic hierarchies.” (p. 46)

**Notes — Chapter 1**

1 There is an ongoing debate in the literature about the existence of some degree of referentiality in food and alarm calls of some birds and mammals. The vervet monkeys from Africa, for example, use specific calls to refer to specific dangers such as snakes, leopards and eagles (Marler, 2001, pp. 32-36). To date, however, the majority of animal calls is not seen as symbolically meaningful, in spite of the significant findings above mentioned.

2 Phonocoding here refers solely to the capacity of making permutations using discrete sound units, which is commonly found in birds but not in primates. The problem of referentiality of sounds does not depend directly on this combinatorial ability.
Although the two authors (Pinker and Miller) present many valid arguments in defense of their theories, we agree with Ian Cross when he states that “[such] theories suffer from an attribute that disqualifies their conclusions from serious consideration, that of ethnocentricity. (…) [They focus] on only one dimension of music (…) : that of music’s inefficacy in any domain other than the individually hedonistic” (Cross, 2001). This hedonistic-oriented” view would be a reflection of the “subsumption of music into the capitalist economy as a tradable and consumable commodity”, after the development of “recording technology together with the reification of intellectual property and the globalization” (Cross, 2001). Underpinning this criticism is the study of the diversity of music around the world, the variability of its uses and the different social values it can have within and beyond Western culture. The question of how to reconcile nature and culture is still the center of the debate: what in music can be explained in general and scientific terms and what should be studied under the lights of particular cultures?

However, it is not clear whether or not the experiment controlled the monkeys’ previous exposure to western music. This could have affected the results, supposing that octave generalization can be learned somehow (Hauser, 2003, p. 666).

P and N mean, respectively, “positive” and “negative”, referring to voltage variations detected in the brain. The numbers 600 and 400 refers to how many milliseconds it took to the variation peak after the target onset (the incongruous note or word). The meaning of this positive-negative qualitative difference still remains unclear (Besson and Schön, 2001).

All possible permutations were equally tested during the experiment: (1) congruous final word sung “in tune”; (2) congruous final word “out of tune”; (3) incongruous final word sung “in tune” and (4) incongruous final word sung “out of tune”. See Besson and Schön (2001), p. 246, for details.

Tonal harmony is taken as “musical syntax” here. One can suppose that similar results would be achieved for any other kind of non-tonal musical structure (“syntax”), provided that the listeners have learned or internalized it in some way. This hypothesis, however, remains to be verified.

This experiment was made by Aniruddh Patel et. al (1998b) and summarized in Besson and Schön’s article (2001, p. 250).

The measurement is called “normalized Pairwise Variability Index” (nPVI). Refer to Patel and Daniele, 2001, p. B37 for more detailed information.

Béla Bartók addresses this question in a direct manner in his “ parlando” rhythms, in which he makes a conscious attempt to generate musical rhythmic material based on linguistic patterns from Hungarian (the slow movement from his fourth string quartet is an example). Thanks to Eric Lyon for this remark.

The importance given to the “note” still today is also linked to the very nature of the Western music notation and its historical development. Conservative music education is also responsible for the maintenance of this “status-quo”: many basic music theory methods in the market today still diffuse anachronistic concepts to new generations of musicians. Unfortunately, a deeper approach of this topic goes beyond the scope of the present work.
At least in part, this is probably a “musical” point of view of language: in actual daily-life speech, nothing is sustained more than certain limits defined by common sense. For a musician, however, manipulation of event durations is naturally part of the compositional process. It is also interesting to relate this opposition “continuous-discontinuous” to the two most basic and ancient instrumental paradigms: the voice and the percussion. Human voice offers the capability to control and vary the sound at any given moment of its production (the same is true for violin, clarinet etc). On the other hand, the “attack-resonance” model characterizes percussion instruments: after the sound is generated, there are little means of altering or varying its resonance, if any (the piano belongs to this category, for example). So phonemes such as /s/ and /t/ are closer to the voice paradigm whereas /p/ and /t/ are closer to the percussion model.

An extreme example outside the tonal music world is Aus den Sieben Tagen (by Karlheinz Stockhausen), which consists of a collection of poem-like texts with instruction for players to improvise (with varying degrees of vagueness). The growth of Electro-Acoustic music in the fifties, or, more specifically, its elektronische Musik branch developed in Germany, can be seen as the culmination of a “hyper-determination” tendency in contemporary music notation. The need for more precise scores had been increasing since the beginning of the twentieth century, as shared assumptions among composers and interpreters progressively disappeared (due to the vanishing of the common ground once provided by tonality). By making use of integral serialism in the electronic studio, composers attempted to overcome all natural “flaws” of a live performance by carefully composing every detail of the work to be fixed on tape. Paradoxically, the elimination of traditional musical notation (the “score”) came together with the apogee of determinacy. Even though the correlation is not so straightforward, we can suppose that the degree of “media transposability” for a given piece of music is related to the degree of determinacy involved in its “notation”, or “writing” (in the more specific sense of the French word “écriture”, which doesn’t requires the existence of a concrete, traditional “paper score”).

The well-established tradition of musical notation probably contributed to the spreading of this approach. In sheet music it seems very easy to identify written notes as “units”, and some composers might even have spoken of them as “units”—but an accurate examination of the actual musical practice shows that this is just part of the reality.

Menezes (1993a), from whom we borrowed this analogy, elaborates on the idea in the following way: “Enfin, avec le surpassement du système sériel intégral et la crise insurmontable—au moins jusqu’à nos jours—des systèmes de référence communs, la liberté s’intègre de nouveau dans la sphère des articulations stylistiques, s’interposant ainsi à tous les niveaux de l’articulation musicale, aux dépens, en revanche, d’un pouvoir de communication plus immédiat de ses énoncés face à sa réception sociale, dont le manque est précisément dû à cette absence de référentialité commune au niveau des méthodes, devenues de plus en plus personnelles” (p. 111).
Chapter 2
2.1 General considerations about the use of voice in electro-acoustic music

In acoustic vocal music, the presence of human voice is taken for granted: the singer is there, physically present on the stage. Moreover, there is a strict code of expected vocal behaviors that a musician must follow within each different traditional musical style. In a common sense, a musical use of voice is associated with singing most of the times. By force of cultural constraints, “pure” speech is not likely to be associated to music.¹

All that does not happen in electro-acoustic music. As it separates the sound from its physical origin, acousmatic music introduces a potential doubt about the cause of a sound. In an acousmatic situation, it is common for listeners to look for an explanation, an origin, a natural correlate or an existing acoustic model to match with the sounds they hear through the loudspeakers. This space, the space of creatively playing with the listener’s expectations and assumptions about the causes of a given sonic event, is unique to electro-acoustic music. The definition of causality fields in a piece can be seen as a new parameter for music composition, a parameter ingrained in the very nature of the acousmatic medium. Thus, in the context of our work, it is fundamental to think that the presence of voice cannot be taken for granted—actually, it is up to the composer to define how much of it will be explicit or not, and to redefine the entire role of it within a specific musical project. There can exist electro-acoustic pieces composed entirely from recorded voices in which no single clue of its presence can be found. In short, the perceived vocality of a piece and the precise role of human voice in it are issues to be defined from scratch by the electro-acoustic composer.

In his text on Spectro-Morphology, Denis Smalley (1986) proposes the idea of surrogacy orders to express the continuum between recognizability of causes:
If the instrumental source is electroacoustically transformed but retains enough of its original identity it remains a first order surrogate. Through sound synthesis and more drastic signal processing, electroacoustic music has created the possibility of a new, second order surrogacy where gesture is surmised from the energetic profile but an actual instrumental cause cannot be known and does not exist. It cannot be verified by seeing the cause. Second order surrogacy therefore maintains its human links by sound alone, rather than by the use of an instrument which is an extension of the body. Beyond this second order we approach remote surrogacy where links with surmised causality are progressively loosened so that physical cause cannot be deduced and we thus enter the realms of psychological interpretation alone (Smalley, 1986, p. 82-83).

The decision of using voice in a given piece in a “first order” manner, to use Smalley’s expression, implies that the composer will be dealing with a complex set of questions related to human voice perception. Any imaginable combination of one or more voices and vocal manifestations brings up compositional choices that are directly linked to the question: what layers of meaning are being added to the piece? These layers are often related to a variety of explicit and implicit cultural associations that can hardly be “wiped off” our perception. Let us give a few concrete examples of immediate impressions that the presence of a voice in a composition can elicit in the listener:

1) Physicality: the presence of a voice implies the presence of a person, real or fictional. Is it male or female? What is the age group of that person? Is it a baby, a child, an adult, an elderly person?
2) Personality: Who is this person? Do I know this person? Is he or she famous? Does the voice bring a recognizable character or type of person?
3) Space recognition: is it only one person there? Are there other voices? Is it a crowd? Is it indoors or outdoors?
4) Language vs. Non-Language: is there a specific language being used? Do I understand this language? Is it intelligible? What do I associate to this or that
language? Or, if there is no language, are those cries? Screams? Laughs? Sneezes? Sung melodies?

5) Meaning attribution: culturally coded and/or personal interpretations are attached to every vocal event presented vocal. What are the possible meanings of these language and non-language vocal sounds?

6) High-level (musical) structures: that is, to build a musical discourse with the vocal sounds. Emphasis can be given to the sounds by themselves, to the external meaning they represent, or to both. Within speech, specifically, there is an entire universe of linguistic parameters such as intonation, stress and durations that are highly meaningful and that can be subjected to compositional “interference”.

The list above is not supposed to be complete, nor should it imply that those are individual, isolated “stages” of perception well separated in time. Rather, listening—not to mention musical listening—is much more of a synchronic and recurrent experience of them all. We do expect to have pointed out how broad and complex is the range of questions that arise upon the decision of composing with the sound of human voice in an evident manner (in the sense of it being distinguishable as voice within a musical work). This being said, it is up to the composer to choose between (a) to worry about these issues and try to incorporate them in the compositional process to some extent, or (b) to discard them and attempt to overcome the vocality of the source material to get to a musical result where extra-musical implications of vocal sounds will be relatively irrelevant. The continuum between the extremes is a fertile territory. We will be able to survey later on how a few historical pieces dealt with these different possibilities.
Before that, however, we still need to discuss in more detail one of the above-mentioned aspects of the use of human voice in electro-acoustic music: the intelligibility of a text. Since speech and music are the central ideas of this work, let us take a closer look at the relationship between the composer and the text.

Luciano Berio described three fundamental tendencies of the text-music correlation within Western music history. The first one, according to him, would be the grand vocal polyphony of the past scope from which the European musical thinking has developed (Berio, 1995). In that period of vocal polyphony, composers “set to music” texts that were universally known (for example, texts of the Mass like Kyrie or Gloria). Complex vocal architectures often hindered the intelligibility of the words, but since texts were widely known, this was not a problem for listeners. This “acoustic freedom” in dealing with the text allowed composer to explore “new musical territories of great expressive intensity and complexity” (p. 102), and Berio included Ockeghem, Palestrina, Bach and Beethoven in this field.

The second tendency, on the other hand, would be characterized by “a metrical, prosodic, rhetoric, formal and expressive coincidence” (Berio, 1995, p. 102) between musical form and literature form. With the progressive abandonment of well-known texts, composers began to set to music increasingly complex literary texts, looking for some kind of isomorphism or “fusion of codes” (Menezes, 1994, p. 35). From baroque to romanticism, different aspects of such structural connections were experimented. Berio cites Mozart, Debussy and the romantic lied tradition as examples.³

Finally, the third tendency would be in the present, when composers should decide what part of the text should be intelligible or not, creating several perceptual layers of
text-music organization. This idea, in which Berio situates himself, “does not necessarily exclude the others. I think of the possibility of using musical criteria for the analysis and expressive rediscovery of a text, providing to the listener the possibility of a conscious, repeated travel through the wonderful itinerary that exists between sound and meaning” (Berio 1995, p. 102-103). The constitution of multiple layers of meaning is a concept very much close to what Umberto Eco expressed in his book The Open Work (Eco, 1989). Besides the idea of multiplicity of listening paths available to the audience, Menezes (1994, p. 36) points out two other significant ideas related to this tendency: text redundancy and relative separation between text and musical structure. Firstly, the use of repetition would allow the existence of a number of different renderings of one same piece of text. Secondly, to keep a conscious distance between text organization and musical structure would be an attempt to preserve relative autonomy and independence between the two codes, looking for a complementary text-music relationship.

With this brief historical survey we can see that there is a variety of compositional approaches to work with text, the intelligibility of which is one of the fundamental questions to be answered by the composer. We can draw an analogy of Smalley’s surrogacy orders to clarify a few ideas on this topic. As explained above, the three proposed levels form a perceptual continuum between the clear identification of a sound’s physical origin towards the progressive annulment of such cause. So our question is: is it possible to have a continuum between intelligibility and unintelligibility of a given text?"
To make our point more precise, let us first situate our question within the framework proposed by Smalley. Let us consider the three levels of surrogacy as I, II, and III (this last one being the “remote”):

I  II  III
Clearly identifiable cause ------------ Hardly identifiable cause ------------- Causes NOT
Gestural traces remain               identifiable anymore

We can agree that the space for text intelligibility is necessarily bound by the intelligibility of some kind of “voice” (or at least an imitation of its). The causality of a human voice must be present in order to a text to be delivered in any way. Therefore, the space for the existence of a text is the space where a voice can be recognized. This space is between the first (I) and second (II) surrogacy orders only, since moving from II to III means the progressive loss of any link with a surmised causality.

This being said, we can now refine our definition of this region, in other words, we can “zoom in” our lens to the space between levels I and II under a vocal perspective. A spoken text digitally recorded in the cleanest possible manner can be considered our ideal model for the sake of this argument: an optimally intelligible recorded voice (from an acoustic point of view). An optimal intelligibility is likely to correspond to a sonic event as close as possible to the original recorded voice (or a synthetic model of a natural voice). Thus the closest the vocal gesture is to the first order surrogacy, the more intelligible a text will possibly be. We may call $T_1$ this category of spoken sounds. In the same way, the other extreme would be the cases where human voice is not recognizable anymore, but the original gesture can still be perceived; in other words, the link to a
physical cause is not totally lost. This is the Smalley’s second order, and we can say that this is the end point for a text. This is the point where the text gets unintelligible, even though some clue of the original physical gesture may be present. Let us call this category T₂. All possible sonic transformations between a clearly intelligible text T₁ and the other extreme T₂ are situated in parallel movement in relation to levels I and II; in other words, the text-intelligibility line is a sub-case of the higher level of causality-recognition line:

\[
\begin{align*}
\text{I} & \quad \text{II} & \quad \text{III} \\
T₁ \quad \text{-------------------------} & \quad T₂
\end{align*}
\]

We can go even further and speculate that there is a substantial difference between the sound level and the meaning level. Explaining better: it is possible—not only with voice, of course—to transform one sound in a progressive manner in a way that an audible continuum emerges between the starting and ending sounds. Wishart (2000) calls this process metamorphosis: “(...) the sonic manipulation of a sound to produce [perceptually] related sounds” (p. 22). Text intelligibility, however, seems to be situated in a different plane, namely, that of a “step function”. Our hypothesis is that it is either comprehensible or not. Instead of a continuum, there would exist a break somewhere in the middle of the line from T₁ towards T₂ where intelligibility, in an absolute sense, would just no longer exist.
By “absolute” we mean that a given piece of transformed voice sound would not be heard as intelligible text by listeners with no previous knowledge of the original text or original recording, in any of its more comprehensible versions.

The question is rather complex, however, if we consider the mutual influence in our listening of different processed instances of the same original sound. In other words, after we hear for a few times a more or less clear recording of a sentence, our capacity of identifying it in even after heavy signal processing increases tremendously. Once a word or phrase is “understood” and fixed in our minds by repetition, we are able to understand it, to retrieve its meaning from transformed versions of it that would not be understandable by listeners not exposed to it before. This phenomenon is easily verifiable by any person working with voice transformation in electro-acoustic studios. Also, there is no single straight line between what we have named T₁ and T₂, since the multitude of existing signal processing techniques allows the creation of infinite ways to sonically move from one point to another.

Anyhow, for the sake of clarity, we can summarize the last considerations with the following graphic:

```
  I    II    III
T₁  ------------/-------------- T₂
```

I: Clearly recognizable source (human voice), most intelligible text
II: Gestural energy remains. Direct causal links with human voice do not exist anymore; consequently, text is not intelligible at all.
III: Most abstract level: hardly any physical cause can be linked to the sounds. No recognizable voice, no recognizable gesture at all.

As we said, a variety of sonic transitions can be created in between T₁ and T₂. Just to cite two simple examples:
a) The text may be subject of permutations in different levels (word, syllable or phoneme, for example). Even scrambled bits of an original text may still bear part of its meaning. If this process occurs long enough, and according to the intention of the composer, the overall meaning of a text can actually be non-linearly delivered almost in its entirety without ever presenting the original sequence. If the bits are extremely small and extremely shuffled, in a way that very few words “pop out”, we then get closer to the $T_2$ pole.

b) By processing the original spoken voice or parts of it in more dramatic ways (stretching, transposing, filtering etc), one can create other paths for the gradual obliteration of the speech flow (and consequently the gradual occultation of verbal meaning).

### 2.2 Historical Overview

Composers have addressed most of the questions discussed above since the very beginning of electro-acoustic music. Schaeffer and Henry’s *Symphonie pour un homme seul* (1950), one of the first major acousmatic compositions, makes extensive use of human voice. Words, sung notes, moans and giggles, among other vocal manifestations, are combined with different kinds of sonic objects in this 12-movement work that set the basis for the consolidation of concrete music in the fifties. Most sound transformations that today might be considered *clichés*, like reversed speech and simple transposition of voice to high registers, can be heard in *Symphonie* with the freshness of discovery and of true musical invention.
At the Cologne electronic music studio in the fifties, the need for a richer palette of sonic materials and the drawbacks created by the strict application of serial techniques to electronic music eventually led composers to abandon the exclusive use of electronic sources. *Pfingsoratorium: Spiritus Intelligentiae Sanctus* (1955-1956), by Ernst Krenek, is a work for soprano, tenor and electronic sounds; *Doppelrohr II* (1955), by Swedish composer Bengt Hambraeus, uses additive synthesis based on a concrete sound model (an organ). But it is after *Gesang der Jünglinge* (1955-1956), by Karlheinz Stockhausen, that the term “electro-acoustic music” became more common—understood, in this context, as the definitive suppression of the historical opposition between the “concrete” and “electronic” schools.

Ironically, one could say that this was made through the “concretization” of electronic music (Menezes, 1996, p. 39), that means, the intromission of the heretofore-undesirable concrete sounds in the music of the sinusoidal school. In *Gesang*, Stockhausen utilized the recorded voice of an adolescent as one of the main sonic materials. Since then, the use of recorded sounds together with synthesized sounds became more and more common, to the point where this apparent “opposition” was not an issue anymore.7

According to Pousseur, Stockhausen achieved an “admirably unified” piece that was at the same time much more diversified than previous electronic compositions. By using a human voice together with the usual electronic material (sine waves, square waves, white noise, impulses, filters and so forth), Stockhausen accepted “a certain degree of indeterminacy in controlling the material”, limiting some macroscopic structural factors to a purely statistic degree of perceptual control (Pousseur, 1996, p. 164).
This turning point inside the Cologne studio can be explained by two main reasons. Firstly, the limitations experienced in the generation of different sound qualities through simple addition of sine waves became evident by the mid-fifties, especially when confronted to new techniques and possibilities of sound generation and transformation. Secondly, the introduction of human voice came across the need for a semantic enrichment of their works. Verbal language, linked to the immense communicative power of non-linguistic vocal sounds in general, would serve as the basis of a new way of structural organization (Menezes, 1996, p. 40).

Besides Stockhausen’s Gesang, we also have Thema (Omaggio a Joyce) and Visage, composed respectively in 1958 and 1961 by Luciano Berio, and also Epitaph für Aikichi Kuboyama, composed by Herbert Eimert in 1960-1962. This kind of electro-acoustic composition intimately tied to the possibilities of human voice is sometimes called “Sprachkomposition” in Germany—“verbal” or “speech” compositions, or “text-sound” pieces, just to cite some other expressions in English.8

We shall open a brief terminological parenthesis related to this last point. The expression “verbal sound” (“Sprachklang”) is historically tied to this idea of “Sprachkomposition”. The core concept is that any electro-acoustic sound derived from human voice is a verbal sound (Humpert, p. 166), even those in which no vocal-related gestural or spectral clue can be found anymore. This leads to some interesting terminological problems. (1) What does the word “verbal” exactly intend to mean? According to a dictionary9, VERBAL means, “of, relating to, or consisting of words”; also “spoken rather than written” and “consisting of or using words only and not involving action”. We see that the overall definition of “verbal” connotes “spoken words”
more than anything else. In this sense, this term should not be used to non-spoken vocal phenomena such as screams, moans, coughs and the like, not to mention more ambiguous cases (onomatopoeia or sung words, for example). (2) Moreover, if a sound object derived from human voice is so heavily processed that it keeps no resemblance with voice at all, should it still be called a “verbal sound”? If we take a sample of a piano note and transform it until it bears no more perceptual hint of its original “pianoness”, there is no more reason to still call it a piano sound.

It is symptomatic that the expression “verbal sound” coined by members of the Cologne school seems to give more importance to the structural origin of a sound than to the actual perception of it. For all these reasons, in this work we do not use the expression “verbal sound” and its correlate “verbal composition” outside their historical context.

(End of terminological parenthesis)

The position of Luciano Berio in this context is worth considering. Although he was situated close to the post-Webernian musical thinking, Berio avoided attaching himself to any of these “schools” and their relative theoretical and practical rigidity in accepting some possibilities and rejecting others. The Studio di Fonologia Musicale (founded by Berio and Maderna in 1955 under the Italian Radio and Television in Milan) did not refuse the “musique concrète” experience: on the contrary, it was very much influenced by it, as well as by serial electronic music and by American tape music. Berio and Maderna were looking for a synthesis of the most recent and diverse musical experiences existing by then. If, on the one hand, we can see that Berio began his electro-acoustic research looking for a stylistic syncretism, on the other hand we observe how Eimert curiously insisted on the worn-out dispute between schools: “Whether we are dealing...
with concrete or electronic means—the pertinent question is never related to the means, but solely to what the artist does with such means. And this is not a technical question, but rather a moral-artistic matter” (Eimert, 1972, p.42-44).

Human voice occupies a very special place in the work of Berio. *Thema (Ommagio a Joyce)* (1958) and *Visage* (1961) are among the most relevant pieces in the history of electro-acoustic music, particularly in the field of voice-based compositions.

*Thema* makes use of an excerpt of James Joyce’s *Ulysses* (chapter XI) in three different languages. A careful organization of the sonic material is established in an attempt to create an “interpretation of Joyce’s text within a restrict field of possibilities provided by the text itself” (Berio, 1994, p. 127). Compositional procedures included the derivation of a series of “vocalic colors” from the words contained in the original text as well as the creation of artificial shocks of consonants to attain richer possibilities of musical articulation (See Berio, 1994).

*Visage*, originally conceived as a radiophonic program, is a tremendous sonic portrait of a human voice — that of American singer Cathy Berberian (1925-1983). The piece touches the borders between sound and meaning, and makes artistic use of issues borrowed from linguistics such as language acquisition, language simulation, prosody and aphasia. *Visage* intends to present language inflexions in a very neutral manner, in such a way that actual associations with real languages could hardly take place. At the same time, the non-sense speech has an undeniable potential of evoking a number of possible meanings according to the listener’s background. As Menezes (1993b) says, “Inasmuch as the concrete association is not *real* and the decoding is *virtual*, the operation of linguistic simulation [becomes] convincing” (p. 46). There is abundant
literature discussing this work and its multi-layered structure combining voice and electronic sounds (see Menezes, 1993b for a complete reference and analysis).

As a co-founder, with Berio, of the Studio di Fonologia Musicale, Bruno Maderna also experimented with human voice and electro-acoustic music. His *Invenzione su una voce* (1960), also known as *Dimensione II*, makes use of isolated phonemes as building blocks for “rhetoric gestures” deprived of any verbal meaning (see Boehmer, 1992). *Le Rire* (1962) has the laughter as its main musical material in a way that the listener may constantly move back and forth between a referential and an abstracted listening. The piece progresses later on to fragments of speaking and singing voices. The laughter becomes less present and less evident towards the end of the piece, as if it had turned into an abstracted gesture to accompany the speech fragments that became prominent.

Herbert Eimert’s *Epitaph für Aikichi Kuboyama* (1960-1962) is a very important early example of a composition specifically based on spoken word. *Epitaph* is perhaps one of the first examples of a politically engaged electro-acoustic piece (the text recitation at its very beginning is a criticism of the atomic bomb).¹⁰ Let us take a closer look at this piece now to see how Eimert dealt with some of the basic questions involved in the use of speech as musical material.

Based on the idea of departing solely from recorded spoken voice, Eimert undertook a “sonic decomposition” of speech in order to extract a varied range of sounds from the original material. Posteriorly, this new universe of verbal sounds became available to the creative “recomposition” by the composer. In *Epitaph*, this typically electro-acoustic process of “decomposition” and “recomposition” of sound shows us the large number of ways that can move from speech towards non-speech sounds. Eimert was completely
conscious of this idea, as we can read:

> With Epitaph I sought the realization the new principle of verbal sounds without compromises, in such a way that what is spoken would appear not only transformed, but also in an enlarged conception in which all possible layers between word and [pure] sound become compositionally available. (Eimert, 1963)

By filling the gaps between “Wort” (word) and “Klang” (sound), using human speech as his point of departure, Eimert was somehow seeking for the suppression of such a dualism commonly found in vocal music in general. In this case, “language is not ‘set to music’ anymore, but rather it is used in the very sonic creation [process] of music” (Humpert, p. 166).

In the instrumental domain, György Ligeti also developed an extremely interesting take on the issue of music and language that is worth mentioning. His two pieces *Aventures* and *Nouvelles Aventures*, composed between 1962 and 1965, go deep in the idea of musically composing an artificial language. According to the composer,

> All the ritualized human emotions that are expressed colloquially, such as understanding and dissension, dominion and subjection, honesty and deceit, arrogance, disobedience, indeed, even the subtlest nuances of irony hidden behind apparent agreement (…)—all this and very much more can be expressed exactly in the a-semantic emotional artificial language. (Ligeti, 1985, p. 8)

The vocal parts are all written using extremely detailed phonetic symbols that were “not drafted before the composition, but developed at the same time as the music” (Ligeti, p. 8). Compositional processes dealt directly with speech sounds. This way, most of the music is consistently drawn from the phonemic and prosodic levels of language. The choice of phonemes and their sonic composition was basically guided by “their ability to evoke the emotional content within a speech-like texture”. Thus, similarly to Berio’s *Visage*, one of the remarkable things about these pieces is this musical “leap” from the phonemic to the prosodic level, completely “skipping”, so to speak, the use of
actual words of any existing language. In this context, the instrumental parts are not merely accompanying the text. Instead, they “complete or highlight the human sounds: the phonetic composition penetrates the sphere of instrumental composition”. (Ligeti, p. 8)

This kind of “awareness of sound” influencing instrumental composition is without doubt related to the discoveries brought by electro-acoustic music after the fifties. Many other composers of this generation were influenced by electro-acoustic music in their instrumental and vocal writing—Ligeti’s micropolyphonic technique is one example of it.13

These first compositions from the fifties and sixties brought up most of the issues related to the expanded musical use of human voice, issues that composers deal with until today. However, even more new questions (or new approaches to previous ideas) arose in the last three decades of the twentieth century.

The piece *I am sitting in a room* (1970) by American composer Alvin Lucier is a 45 minutes long composition in which two central ideas come into play: the establishment of a continuum between “speech” and “music” and the use of an acoustical space as a musical instrument. *I am sitting in a room* is made of a recording of a single voice reading a text that makes the piece self-explanatory:

> I am sitting in a room different from the one you are in now. I am recording the sound of my speaking voice and I am going to play it back into the room again and again until the resonant frequencies of the room reinforce themselves so that any semblance of my speech, with perhaps the exception of rhythm, is destroyed. What you will hear, then, are the natural resonant frequencies of the room articulated by speech. I regard this activity not so much as a demonstration of a physical fact, but more as a way to smooth out any irregularities my speech might have. (Lucier, 1990, p. 1)
As the piece progresses, the words become more and more blurred; consequently, their meaning slowly fades away. By the end of the piece one can only hear harmonic textures made up of frequencies that were not filtered out by the room’s natural resonant properties. According to Nicolas Collins (in Lucier, 1990, p. 5), the ‘demonstration of a physical fact’ alluded by Lucier actually takes advantage of the very nature of speech and its irregularities to highlight a strong link between music and language.

*Speech Songs* (1972) and *Cascando* (1978) by Charles Dodge were some of the first uses of synthetic speech in music. *Cascando* is based on the homonymous Samuel Beckett’s play. A synthetic voice plays one of the parts with an artificial, almost sung intonation made possible through synthesis-from-analysis paradigm. “The part was read into the computer in the musical rhythm and, after computer analysis, resynthesized with an artificial (“composed”) pitch line in place of the natural pitch contour of the voice” (Dodge & Jerse, 1997, p. 238).

*Six Fantasies on a Poem by Thomas Campion* (1978-1979) by Paul Lansky is another famous example of a work dedicated to the inner musical qualities of speech. Different aspects of a reading of the poem are developed in each fantasy through speech resynthesis from analysis and other computer music techniques: expansion of pitch range and contour, harmony-timbre construction based on superimposition of transposed layers, speech reverberation and comb filtering are some examples.

Meanwhile, a strong movement of Swedish composers interested in the relationship between music and language was established in the transition from the sixties to the seventies. In 1967 Lars-Gunnar Bodin and Bengt Emil Johnson created the term “text-sound”. With the support of the Swedish Radio, many composers were able to explore the
“borderland between poetry and music” (Johnson). Lars-Gunnar Bodin has been one of the driving forces of the electro-acoustic music scene as a composer and administrator (he was chairman of the Fylkingen society, a Swedish group devoted to new music, and helped to establish the first Swedish Electro-Acoustic Studio). Besides musical theater, radio plays and “pure” electro-acoustic music, text-sound composition has been his main focus of interest. We can mention his piece On Speaking Terms II (1986) as one example of it. In this piece, several layers of the same recorded voice are superimposed with slightly different speeds, generating “cascades” of language sounds in which semantics is dissolved.

Sten Hanson was one of the composers that joined the Fylkingen society to pursue his artistic goals: an intermedium between literature and music, using oral elements of language in a polyphonic approach that would be essentially different from written poetry; finally, he also argued that the appropriate media for this kind of work would be the tape or the radio diffusion (Hanson). Hanson, who also had a fundamental role in the organization of new music in Sweden, considers his trilogy frarp(e)—Oips—OUHM (1969-1973) as the final result of his work with text-sound composition, where he makes “full use” of his ideas (Hanson). In the overall, the trilogy also deals with the idea of progression from clear verbal meaning to several degrees of abstraction based on the work with vocal sounds.

It is worth noting that the radio was considered a privileged medium for electro-acoustic and text-sound music since the beginning. It is not fortuitous that the first European studios were created with strong support of public radios. Visage was initially planned to be a radio program, and Berio himself brought this idea back with A-Ronne
(1974), which has the subtitle “a radiophonic documentary”. A-Ronne (which is not an electro-acoustic piece) is a remarkable composition that develops the concept of teatro per l’orecchi (“theater for the ears”). Five actors go through an extensive “catalogue” of vocal situations based on the continuous repetition of a poem by Edoardo Sanguinetti. Screaming, choking, breathing, laughing, crying, pain, gossip, relief, word fragmentation and singing are just a few examples of techniques, types of sound and emotions used or evoked. The “strangeness” of so many conflicting well-known situations and “moods” put together (in both synchronic and diachronic ways) is one of the strategies to allow their intrinsic musicality to come out and to be heard as such.

The American composer Robert Ashley developed a personal technique to compose with spoken word. A careful, detailed structuring of everyday speech is actually the basis of his piece eL/Aficionado (1987), one of four operas in which he used this same technique. “I work over the syllable count and the accent count in the lines, until they satisfy me, and until they satisfy some general requirement that I’ve discovered in the piece itself” (Ashley, 2001, p. 211). Another crucial component characteristic of Ashley’s approach to “speech composition” is pitch organization. Singers organize their pitch-inflections according to a given center pitch. After working many years with his own ensemble, Ashley actually composed many pieces thinking specifically in the typical speech contours and pitch range of each one of his singers.

“In every solo or ensemble part, the singer is given a ‘character defining’ pitch (that is, a pitch somewhere in the singer’s range that, understandably, forces a certain ‘character’ to emerge). Around this pitch the singer is asked to invent vocal inflections (pitch
changes, vocal techniques etc.) that express the intent or meaning of the text” (Ashley, 1993).

Long-sustained chords are usually played by an electro-acoustic part (the “electronic orchestra”), while occasional sung melodies break the dominant spoken narrative. The explicit harmony is responsible for providing the singers with their central tones. Each singer’s own creativity has a significant role in the creative process of the speech shapes. However, the interest and originality of this technique seem to us to be imprisoned in a permanent grid of a weak neo-tonal background and a rather limited development of the electro-acoustic sounds (mainly based on commercial synthesizer’s preset sounds in a conscious aesthetical option).

In his first electro-acoustic compositions—Phantom-Wortquelle: Words in Transgress (1987) and PAN: Laceramento della Parola (Omaggio a Trotskij) (1988), the Brazilian composer Flo Menezes develops what he calls “forma-pronúncia” (pronunciation-form), which is an attempt of transposing the form of phonetic phenomena to the musical form itself—that is, to the overall duration of a composition. The author describes it as a

… syntactic liberation [inside] the very word, [penetrating] in its internal phonologic structure, perceiving each detail of its expressive sonic dynamism, expanding its form beyond its presumed perceptual limits (what implies to exceed in an exaggerate manner the natural extension of words through its pronunciation, thus intensely amplifying the quantity—or duration—of each phoneme). (Menezes, 1996, p. 221)

Tongues of Fire (1994), by British composer Trevor Wishart, goes into a different direction. As in many of the pieces discussed above, the whole work is derived from human voice. In the composer’s definition,

The slightly angry, disgruntled, comic muttering with which the piece begins is the source of all sounds subsequently heard in the piece, sounds suggesting drums, water, metallic resonances, fireworks, or entirely imaginary materials or events, all generated through processes of sound transformation on the computer. (Wishart, 2000b)
Even though verbal meaning is never present in the piece, the vocal gesture keeps much of its strength throughout the music (whose focus is actually the metamorphosis and transformation of sounds—see earlier in this chapter).

*Machinations* (2000), by French composer of Greek origin George Aperghis, represents some of the latest developments of musical theater in which sound, action, visuals and language are all integrated in an indivisible unit. Called as a “musical spectacle for 4 women and computer”, the piece uses a text by François Regnault as a “pretext” for a laborious work on phonemic composition. Regnault says: “George has the ‘speakers’ produce associations of consonants and phonemes that are absolutely dizzying; he makes them do things that don’t exist in any language” (in Aperghis, p. 6). His work goes from this low-level organization of basic phonemes to higher levels of language phenomena like pitch organization of speech (intonation) and complex vocal gestures evoking different moods and emotions. The singers are also required to do other physical gestures as part of the score. These gestures, like the rising of a hand in a very specific point in time, are thought of as musical events, and “no longer have a sense. The gesture is freed. Similarly, when one removes its meaning from a spoken phrase, the phonemes start floating. Moreover, the images often seem to rise up from the interior of the body—they echo my idea on the production of phonemes” (Aperghis, p. 10). Vocal sounds in general and language sounds in particular have long been one of the main interests of Aperghis. His earlier work *Récitations* (1978), for one solo female voice, is an incredible travel through the possibilities of a musical and theatrical organization of spoken language.
<table>
<thead>
<tr>
<th>Name</th>
<th>Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCHOENBERG, Arnold (1874-1951)</td>
<td><em>Pierrot Lunaire</em> (1912)</td>
</tr>
<tr>
<td>SCHWITTERS, Kurt (1887-1948)</td>
<td><em>Ursonate</em> (1922-1932)</td>
</tr>
<tr>
<td>SCHAEFFER, Pierre (1910-1995); HENRY, Pierre (1927)</td>
<td><em>Symphonie pour homme seul</em> (1950)</td>
</tr>
<tr>
<td>LIGETI, György (1923)</td>
<td><em>Aventures and Nouvelles Aventures</em> (1962-1966)</td>
</tr>
<tr>
<td>MACHE, François-Bernard (1935)</td>
<td><em>Le son d’une voix</em> (1964)</td>
</tr>
<tr>
<td>LUCIER, Alvin (1931)</td>
<td><em>(I am sitting in a room)</em> (1970)</td>
</tr>
<tr>
<td>LANSKY, Paul (1944)</td>
<td><em>(Six Fantasies on a Poem by Thomas Campion)</em> (1978-1979)</td>
</tr>
<tr>
<td>BODIN, Lars-Gunnar (1935)</td>
<td><em>(On Speaking Terms II)</em> (1986)</td>
</tr>
<tr>
<td>ASHLEY, Robert (1930)</td>
<td><em>(eL/Aficionado)</em> (1987)</td>
</tr>
<tr>
<td>WISHART, Trevor (1946)</td>
<td><em>(Tongues of Fire)</em> (1994)</td>
</tr>
<tr>
<td>APERGHIS, George (1945)</td>
<td><em>(Machinations)</em> (2000)</td>
</tr>
</tbody>
</table>

Table 1: Compositions cited in chapter 1 plus a few other important related pieces that could not be discussed in the present text.

Notes — Chapter 2

1 Naturally, exceptions can be found in many places and times. The recitative in opera and the spoken word in rap are just two examples of the presence of speech features in a musical style.

2 Regarding the issue of source attribution in our perception of sounds, Wishart (2000) wrote that “no matter how abstracted from reality our sounds might be, we will still tend to attribute material and dynamic causation to them that is based on our experience of sound in the real world” (p. 24). In this sense, there would be no pure “remote surrogacy” as defined by Smalley, since even the more “abstract” sounds would be associated to an approximate plausible cause by listeners.

3 The fruitful collaboration between concrete poetry and the Brazilian New Music group in the sixties is a more recent example of a similar interaction.
Stockausen attempted to systematize the idea of a continuum from intelligible to unintelligible by means of serialism in *Gesang der Jünglinge* (1955-1956). He created seven levels of intelligibility for the text:

1) nicht verständlich = not intelligible
2) kaum verständlich = practically not intelligible
3) sehr wenig verständlich = only a little intelligible
4) fast verständlich = almost intelligible
5) nicht genau verständlich = not exactly intelligible
6) mehr verständlich = a little more intelligible
7) verständlich = intelligible

Stockausen used a non-orthodox serialism in *Gesang*. However, it seems to us that the idea of seven degrees of intelligibility is not more than an abstraction without link to the reality of human perception (a typical problem of the early years of serialism), especially within a complex musical composition in which several other parameters influence what is actually heard by listeners. On the other hand, one may argue that no listener is expected to fully hear or understand these degrees of intelligibility, which would be simply a working methodology for the composer—and as such they wouldn’t need to be transparent in the final audible result. On this respect, Humpert states that “(...) composing with ‘degrees of intelligibility’ (ordered in a scale) is fictitious, since after repeated listening our perception condenses everything down to two extremes: intelligible and unintelligible” (Humpert, 1987, p. 171; our adapted translation).

Wishart (2000) also defines the correlate concept of *transformation*, which for him “refers to a process of sonic development through time; that is, the use of sonic relationships between events to build musical structures in time” (p. 22).

Stockhausen structured the presence of text in his *Gesang der Jünglinge* (1955-56) according to seven degrees of intelligibility that he himself defined from “not intelligible” to “intelligible”. The composer Hans Humpert, making a veiled criticism of this approach, wrote that “the composition of ‘degrees of intelligibility’ (ordered in a scale from not intelligible to completely intelligible, passing by practically unintelligible and almost intelligible) is fictitious, since after several listenings our perception is reduced to the two extremes: intelligible and not intelligible”. See HUMPERT, 1987, p. 171.

It is curious to see that members of the Cologne School such as Mauricio Kagel and Herbert Eimert of the Cologne School later relativized the importance of such opposition or even the existence of a School at all. Kagel, for example, wrote that many of them were just visiting composers, and nobody felt as if they were members of such a unified “school” (See Kagel apud Menezes, 1996, p. 46). Eimert eventually considered that, even though there were differences in each approach, this hypothesis of the Paris-Cologne antagonism became a “ridiculous fight and competition” (Eimert, 1972, p. 42-44).

The term “text-sound” was created in 1967 by Lars-Gunnar Bodin and Bengt Emil Johnson, as we will see later on this chapter.


Aikichi Kuboyama was a Japanese fisherman that died in 1954 due to radioactive exposition. Less than a hundred miles from his boat, near the Bikini atoll in the South
Pacific, the US Navy was undertaking a series of experiments with the H-bomb. Many other fishermen were fatally injured in the incident.

11 The quarrel as to whether Epitaph is “concrète” or “elektronische” music is, in our understanding, completely irrelevant. The piece was composed between 1960 and 1962. Years before that, the “sinusoidal school” had already reached the deadlock of serialism, and had begun using “concrete” sources to enrich their sonic palette (Gesang der Jünglinge by Stockhausen is the pioneer and most well-known example of such move). Therefore, by the end of the fifties there is no more practical sense to speak of a bellicose opposition concrete versus electronic, even though peculiar aesthetical traits and influences may still be found among inheritors of one or the other orientation. Eimert, himself part of the serial tradition, stated that the formal structure of Epitaph was worked out only during the actual elaboration of the piece (Eimert, 196-).

12 Two technical procedures used by Eimert are interesting to mention. They were very typical of the production of verbal sounds in the electronic analog studio of the fifties. 1) The “Adjustable Frequency Band Amplifier” (Abstimmarer Anzeigeverstärker): in this method, spectral chunks from speech sounds were amplified in a series of selected bands. New spectra could be obtained by assigning new weights to different regions of partials. A variety of sound objects could be generated without losing the rhythmic aspect of speech. The occlusive consonants had an “agglutinating” role in the sonic blocks created this way. Since occlusives have a very dense spectrum and thus sound in most frequency regions, at least part of their constituents would be retained throughout this process. 2) Stationary sounds: a device called Tempophon was used to create stationary sounds departing from speech. A special type of rotative head (with four or more points for magnetic reading) could repeatedly play a fixed point on a magnetic tape. This is the equivalent of what nowadays we usually call “freezing”, or a special type of microscopic looping. The use of vowels or consonants yielded a variety of sound qualities as a result. According to Humpert, a regular sound processed this way would sound a ninth below the original frequency. The iterative quality of most sounds generated this way could be converted into steadiness by speeding up the tape (that is, making higher transpositions of it). Denser textures were obtained by overlapping these results, in a related process called “iterative copies” (Dauerkopierverfahren; see Humpert, p. 164-165).

13 It is also interesting to note that this innovative use of spoken voice in Aventures and Nouvelles Aventures opened up a breach amid his well-known style of that period: that is, the “Ligeti” sound, characterized by dense textures of sound with microscopic internal movements. The extensive use of a very idiosyncratic technique may sometimes “trap” a composer in his or her own unique style, where many pieces sound like “the same”. This is not necessarily criticisable or inherently bad, but if on the one hand this can be the expression of true artistic honesty and commitment to one idea, on the other hand it can well sound limited or even “fabricated”. Although very idiosyncratic, Ligeti’s techniques brought some of the most important contributions to the practice and theory of contemporary music until today. In this context we think that Ligeti’s adventures with human voice represent a breach in his typical sound of that period, predating the diversity of Le Grand Macabre (1974-1977), a piece that brought together many of his discoveries and that marked a turning point in his compositional work.
“I love factory sounds. I actually change them a little bit, but I love them. It’s like hiring a wonderful clarinet player in the old days (…)” (Ashley, in Chadabe, 1997, p. 259)

This piece is an extremely interesting early example of use of analysis of acoustic sounds as models for instrumental composition, and in this special case, the model was human voice. Unfortunately, we were unable to find much more information on it at this moment. The site


contains this short comment on Mâche’s work: “Parmi les multiples explorations esthétiques auxquelles il s’est intéressé figurent les modèles linguistiques. Dès 1959, il a transposé instrumentalement la structure phonétique d’un poème grec dans Safous Mêle. Il a aussi fait œuvre de pionnier dès Le son d’une voix en 1964, avec son usage de l’analyse par sonogrammes pour élaborer une écriture instrumentale, préfigurant l’école dite « spectrale ». Depuis, il a fréquemment exploité le potentiel musical de diverses langues rares ou éteintes, et proposé des méthodes d’analyse structurale inspirées par ses études linguistiques.”
Chapter 3
3.1 The Spell of Speech

This last chapter is about our composition The Spell of Speech, a piece for actress and electro-acoustic sounds with live-electronics that was composed between January and May 2004. The piece reflects many of the ideas discussed in the previous two chapters. The name of the piece was inspired by the last chapter of the book The sound shape of language (Roman Jakobson and Linda Waugh, 2002), which is called “The spell of speech sounds”. The Spell of Speech is a 25 minutes-long musical composition with a significant theatrical side. Composition using spoken word (and also other vocal manifestations) is the central idea. The text used is l’Attente (“Waiting”), by Roland Barthes. This text is a chapter of his book Fragments d’un Discours Amoureux, first published in 1977 (see full text in Appendix IV).

We shall describe the main techniques used in the composition as well as the overall form and its inner constituent parts. Enclosed with this thesis are a CD and a DVD with the first performance (May 16th 2004 in Rollins Chapel, Dartmouth College). There is also an additional CD containing the other pieces that were played in that same performance, most of them related in some way to our research on speech and music (See Appendix VII for CDs and DVD contents). A score for The Spell of Speech is also provided in Appendix I. It should be seen as a guide for the listener or analyst; it is not a score to be read by a performer in order to recreate the piece. We suggest to the readers to listen to the CD (or watch the DVD) following the score before reading this chapter.
3.2 Techniques I: Analysis and resynthesis

One of the sound processes used in The Spell of Speech is based on analysis and resynthesis of speech samples.\(^1\)

The speech analysis currently undertaken for our research is based on the *partial tracking* function in the software Audiosculpt (IRCAM). After obtaining a sonogram of the sample under study, the partial tracking function gives us a graphic representation of the partials by means of line segments; these lines are as straight as the partials are steady. In the case of human speech, the prosodic level is often characterized by continuous changes in pitch within words, thus resulting in curved lines for each partial (Figure 1).

We can export this data in text format for subsequent use in our resynthesis. Audiosculpt provides two ways of exporting partials. The first is to keep all minute details of each curved line segment, which yields a huge amount of data for each second of speech. The second is to average all partials in order to decrease the amount of data. In this case all partials are converted to simple straight lines (Figure 2). In this second method we lose the fine-grained details of the analyzed sound, but experience has proved that this is not necessarily inferior: the two methods just yield different musical results.
Figure 1: Sonogram in Audiosculpt. The lines show a typical partial tracking for a spoken sentence ("Check it out" spoke by a low male voice), with all their internal fluctuations (small glissandi).
Figure 2: Same sonogram as in Figure 1, but here the partials are averaged.

In both cases, the saved text file is simply a list of numbers describing each of those line segments.

The non-averaged option originates longer lists. Tenths or hundreds of time points are needed to describe one single partial, each point being assigned a slightly different frequency-amplitude pair. The example below is just the beginning of a typical non-averaged list. Its header tells us that 474 partials are described in this text file. The first partial is described using 126 point in time, from second 0.447 to 1.173 (absolute time in the sound file). The numbers are to be read as follows: second 0.447 has the frequency
204.486 at −38.920 dB; second 0.453 has a frequency of 203.043 at −30.203 dB²; and so forth. After these 126 lines, the first partial (which lasts less than a second) is fully described. Then the list would move to the second partial etc., until all the 474 partials are finished.

<table>
<thead>
<tr>
<th>PARTIALS</th>
<th>POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>474</td>
<td>126</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.447</td>
<td>204.486</td>
<td>-38.920</td>
</tr>
<tr>
<td>0.453</td>
<td>203.043</td>
<td>-30.203</td>
</tr>
<tr>
<td>0.459</td>
<td>201.544</td>
<td>-22.807</td>
</tr>
<tr>
<td>0.464</td>
<td>200.672</td>
<td>-17.831</td>
</tr>
<tr>
<td>0.470</td>
<td>201.297</td>
<td>-14.698</td>
</tr>
<tr>
<td>0.476</td>
<td>202.806</td>
<td>-12.929</td>
</tr>
<tr>
<td>0.482</td>
<td>204.348</td>
<td>-12.024</td>
</tr>
<tr>
<td>0.488</td>
<td>205.490</td>
<td>-10.857</td>
</tr>
<tr>
<td>0.493</td>
<td>205.696</td>
<td>-9.905</td>
</tr>
<tr>
<td>0.499</td>
<td>205.577</td>
<td>-9.155</td>
</tr>
<tr>
<td>0.505</td>
<td>205.647</td>
<td>-8.425</td>
</tr>
<tr>
<td>0.511</td>
<td>205.839</td>
<td>-7.710</td>
</tr>
<tr>
<td>0.517</td>
<td>206.170</td>
<td>-7.031</td>
</tr>
<tr>
<td>0.522</td>
<td>206.715</td>
<td>-6.400</td>
</tr>
<tr>
<td>0.528</td>
<td>207.286</td>
<td>-5.853</td>
</tr>
<tr>
<td>0.534</td>
<td>207.498</td>
<td>-5.423</td>
</tr>
<tr>
<td>0.540</td>
<td>207.224</td>
<td>-5.204</td>
</tr>
<tr>
<td>0.546</td>
<td>205.448</td>
<td>-5.561</td>
</tr>
<tr>
<td>0.551</td>
<td>202.100</td>
<td>-5.033</td>
</tr>
<tr>
<td>0.557</td>
<td>199.911</td>
<td>-4.077</td>
</tr>
<tr>
<td>0.563</td>
<td>200.885</td>
<td>-3.272</td>
</tr>
<tr>
<td>0.569</td>
<td>202.952</td>
<td>-2.656</td>
</tr>
<tr>
<td>0.575</td>
<td>203.979</td>
<td>-2.218</td>
</tr>
<tr>
<td>0.580</td>
<td>204.072</td>
<td>-1.947</td>
</tr>
<tr>
<td>0.586</td>
<td>204.098</td>
<td>-1.890</td>
</tr>
<tr>
<td>0.592</td>
<td>204.247</td>
<td>-2.001</td>
</tr>
<tr>
<td>0.598</td>
<td>204.393</td>
<td>-2.200</td>
</tr>
<tr>
<td>0.604</td>
<td>204.632</td>
<td>-2.407</td>
</tr>
<tr>
<td>0.610</td>
<td>205.080</td>
<td>-2.395</td>
</tr>
<tr>
<td>0.615</td>
<td>205.681</td>
<td>-2.304</td>
</tr>
<tr>
<td>0.621</td>
<td>206.148</td>
<td>-2.202</td>
</tr>
<tr>
<td>0.627</td>
<td>206.539</td>
<td>-2.093</td>
</tr>
<tr>
<td>0.633</td>
<td>207.287</td>
<td>-1.927</td>
</tr>
<tr>
<td>0.639</td>
<td>208.472</td>
<td>-1.705</td>
</tr>
<tr>
<td>0.644</td>
<td>209.870</td>
<td>-1.468</td>
</tr>
<tr>
<td>0.650</td>
<td>211.604</td>
<td>-1.260</td>
</tr>
<tr>
<td>0.656</td>
<td>213.842</td>
<td>-1.108</td>
</tr>
<tr>
<td>0.662</td>
<td>216.341</td>
<td>-1.043</td>
</tr>
<tr>
<td>0.668</td>
<td>218.922</td>
<td>-1.110</td>
</tr>
<tr>
<td>0.673</td>
<td>221.728</td>
<td>-1.372</td>
</tr>
<tr>
<td>0.679</td>
<td>224.993</td>
<td>-2.025</td>
</tr>
<tr>
<td>0.685</td>
<td>228.929</td>
<td>-1.893</td>
</tr>
<tr>
<td>0.691</td>
<td>233.142</td>
<td>-1.843</td>
</tr>
<tr>
<td>0.697</td>
<td>236.016</td>
<td>-2.368</td>
</tr>
<tr>
<td>0.702</td>
<td>237.000</td>
<td>-3.207</td>
</tr>
<tr>
<td>0.708</td>
<td>237.259</td>
<td>-4.093</td>
</tr>
<tr>
<td>0.714</td>
<td>238.031</td>
<td>-4.918</td>
</tr>
<tr>
<td>0.720</td>
<td>239.565</td>
<td>-5.688</td>
</tr>
<tr>
<td>0.726</td>
<td>242.032</td>
<td>-6.313</td>
</tr>
<tr>
<td>0.731</td>
<td>244.132</td>
<td>-6.445</td>
</tr>
<tr>
<td>0.737</td>
<td>244.508</td>
<td>-5.982</td>
</tr>
<tr>
<td>0.743</td>
<td>245.031</td>
<td>-5.423</td>
</tr>
<tr>
<td>0.749</td>
<td>247.030</td>
<td>-5.408</td>
</tr>
<tr>
<td>0.755</td>
<td>250.100</td>
<td>-4.903</td>
</tr>
<tr>
<td>0.760</td>
<td>253.726</td>
<td>-4.431</td>
</tr>
<tr>
<td>0.766</td>
<td>257.475</td>
<td>-4.268</td>
</tr>
<tr>
<td>0.772</td>
<td>261.560</td>
<td>-4.399</td>
</tr>
<tr>
<td>0.778</td>
<td>266.140</td>
<td>-5.011</td>
</tr>
<tr>
<td>0.784</td>
<td>271.693</td>
<td>-5.177</td>
</tr>
<tr>
<td>0.789</td>
<td>277.636</td>
<td>-4.524</td>
</tr>
<tr>
<td>0.795</td>
<td>282.923</td>
<td>-4.528</td>
</tr>
<tr>
<td>0.801</td>
<td>287.091</td>
<td>-5.011</td>
</tr>
<tr>
<td>0.807</td>
<td>290.384</td>
<td>-5.789</td>
</tr>
<tr>
<td>0.813</td>
<td>293.427</td>
<td>-5.442</td>
</tr>
<tr>
<td>0.819</td>
<td>297.077</td>
<td>-5.358</td>
</tr>
<tr>
<td>0.824</td>
<td>301.351</td>
<td>-5.691</td>
</tr>
<tr>
<td>0.830</td>
<td>305.651</td>
<td>-6.264</td>
</tr>
<tr>
<td>0.836</td>
<td>308.482</td>
<td>-6.746</td>
</tr>
<tr>
<td>0.842</td>
<td>309.797</td>
<td>-6.887</td>
</tr>
<tr>
<td>0.848</td>
<td>311.028</td>
<td>-6.931</td>
</tr>
<tr>
<td>0.853</td>
<td>312.673</td>
<td>-6.933</td>
</tr>
<tr>
<td>0.859</td>
<td>314.327</td>
<td>-6.450</td>
</tr>
<tr>
<td>0.865</td>
<td>315.389</td>
<td>-6.179</td>
</tr>
<tr>
<td>0.871</td>
<td>315.594</td>
<td>-6.077</td>
</tr>
<tr>
<td>0.877</td>
<td>315.431</td>
<td>-6.077</td>
</tr>
<tr>
<td>0.882</td>
<td>315.359</td>
<td>-6.112</td>
</tr>
<tr>
<td>0.888</td>
<td>315.108</td>
<td>-6.159</td>
</tr>
<tr>
<td>0.894</td>
<td>314.266</td>
<td>-6.211</td>
</tr>
<tr>
<td>0.900</td>
<td>312.725</td>
<td>-6.290</td>
</tr>
<tr>
<td>0.906</td>
<td>310.725</td>
<td>-5.572</td>
</tr>
<tr>
<td>0.911</td>
<td>308.537</td>
<td>-4.683</td>
</tr>
<tr>
<td>0.917</td>
<td>306.679</td>
<td>-4.059</td>
</tr>
<tr>
<td>0.923</td>
<td>305.398</td>
<td>-3.674</td>
</tr>
<tr>
<td>0.929</td>
<td>304.364</td>
<td>-3.442</td>
</tr>
<tr>
<td>0.935</td>
<td>303.042</td>
<td>-3.285</td>
</tr>
<tr>
<td>0.940</td>
<td>301.234</td>
<td>-3.210</td>
</tr>
<tr>
<td>0.946</td>
<td>299.935</td>
<td>-3.313</td>
</tr>
<tr>
<td>0.952</td>
<td>296.133</td>
<td>-3.708</td>
</tr>
<tr>
<td>0.958</td>
<td>293.065</td>
<td>-4.435</td>
</tr>
<tr>
<td>0.964</td>
<td>290.107</td>
<td>-5.037</td>
</tr>
<tr>
<td>0.969</td>
<td>287.284</td>
<td>-4.278</td>
</tr>
<tr>
<td>0.975</td>
<td>284.618</td>
<td>-3.443</td>
</tr>
<tr>
<td>0.981</td>
<td>282.620</td>
<td>-2.622</td>
</tr>
<tr>
<td>0.987</td>
<td>280.827</td>
<td>-1.961</td>
</tr>
<tr>
<td>0.993</td>
<td>278.612</td>
<td>-1.605</td>
</tr>
<tr>
<td>0.998</td>
<td>275.845</td>
<td>-1.662</td>
</tr>
<tr>
<td>1.004</td>
<td>272.669</td>
<td>-2.222</td>
</tr>
<tr>
<td>1.010</td>
<td>269.054</td>
<td>-3.320</td>
</tr>
<tr>
<td>1.016</td>
<td>265.019</td>
<td>-3.055</td>
</tr>
<tr>
<td>1.022</td>
<td>260.516</td>
<td>-3.270</td>
</tr>
</tbody>
</table>
Exporting averaged partials, on the other hand, produces much smaller lists. The averaged list contains one fixed frequency-amplitude pair for each partial, plus two time points describing its onset and ending time. The example below is the beginning of a list containing 1950 averaged partials. The first three partials are shown. Since they are simple straight lines, each averaged partial needs no more than two points in time to be described.

```plaintext
(PARTIALS 1950
(POINTS 2
 0.868 212.068 -4.527
 1.117 212.068 -4.527
)
(POINTS 2
 0.879 423.652 -13.134
 1.120 423.652 -13.134
)
) etc.
```
Perhaps the most interesting step from a musician’s point of view is the one right before using all data in a “resynthesis” process. We put it between quotes because a realistic resynthesis of the sampled voice often is not the goal of a composer. Rather, it is at this moment that one can make a wide variety of alterations and mutations of the available parameters to get results that range from clearly recognizable speech to sound textures without apparent relationship with the initial voice. The possibility of using speech data as the basis for instrumental writing is also considered within the idea of “resynthesis”. The following is an examination of the two different approaches.

3.2.1 Csound: our Csound implementation for additive synthesis varies according to the use of averaged or non-averaged partials.

a) Averaged Partials: each partial is represented by a simple, steady Csound note. Another IRCAM program (Open Music) has been used to sort the original data and generate the Csound score. Figure 3 is a screenshot of the Open Music patch used. The text file with partial tracking data is imported to the text box on the top left corner. The raw data is then “cleaned up” by a subpatch that sorts the numbers and send them out separated by categories (onset times, durations, frequencies, amplitudes). At this point, any numbers can be altered by any kind of process, e.g. multiplication or permutation. The last two objects in the chain (bottom) are used to generate a Csound score with the information provided.
b) Non-Averaged Partials: this case requires a different paradigm. Partials are still represented as Csound notes, but each note must have a different pair of curves describing its internal frequency-amplitude fluctuations. To accomplish this, such curves are extracted from the data and loaded into separate tables using GEN23. These tables then control the expected internal movement of each partial. A Perl script was especially written for this task (see appendix II). Hundreds or even thousands of function tables are generated this way in the resulting Csound score.
In both cases (averaged or non-averaged partials), an enormous variety of complex sound textures can be obtained, resembling the original voice at different levels. In general, as long as the onsets and durations are kept approximately close to the original values, the resulting sound should resemble the temporal and rhythmic contour of the analyzed voice. On the other hand, sound textures with no resemblance to the human voice can also be obtained by changing parameters beyond certain limits. We have also experimented using instrumental samples instead of sine waves in the resynthesis process, in order to expand the field of timbre possibilities.

The main question to a composer using the processes described above is: where and how to transform the raw data?

In the electro-acoustic domain, factors such as intelligibility of words and of other speech features must be considered according to a given musical objective. The use of the original sample in subtle mixing with the resulting synthesis has proved to be an efficient way to increase intelligibility without losing the interest of the new synthetic texture. Simple function tables with sine waves have been used mostly to create sound “shadows” that follow the original speech, highlighting portions of its rhythmic and harmonic movement. This process can be heard in The Spell of Speech: refer to beginning of line 14 in the score, right before the cue “scene11” (CD 1, Track 1, 20’40”). Another example of a sine wave resynthesis, this time much more distant from the any recognizable speech, appears in the form of an electronic counterpoint for the “cry” section (long synthetic texture that emerges in the middle of the real cry; score line 6, CD 1, Track 1, from 6’30” to 7’40” approximately).
The use of samples as the main timbre for the resynthesis usually yields richer sound textures that may still keep important properties of the original speech. Even though the sonic result can no longer be directly recognized as speech, it can maintain a gestural connection with it. The entire electro-acoustic introduction of The Spell of Speech (score line 1, CD 1, Track 1, from 0’00” to 1’35”) is based on this process of speech analysis and resynthesis. Samples of flute notes were used to generate the sonic material that was eventually sequenced in ProTools. Cue “scene7”, (score line 10, CD 1, Track 1, 10’50”) and the coda “scene11” (score line 14, CD 1, Track 1, 21’08”) are based in speech resynthesis with flute samples as well.

Although it is not used in The Spell of Speech, we should mention that instead of managing the speech analysis data to write a Csound score and create electro-acoustic sounds, one can also make use of the same raw data to develop instrumental compositions. We have composed a piece called Gedankenfabrik (2003) for twelve instruments and electro-acoustic sounds in which the Csound resynthesis was used to generated part of the electro-acoustic material, and the same speech data was used in the creation of some of the piece’s instrumental passages. Some speech features produce a very special shaping of musical ideas in the instrumental domain. We have used two main methods for accomplishing this idea. One is to use Open Music’s “Chord-Seq” module to convert the raw data from the analysis file into a MIDI file. This MIDI file, once converted into a music score, is then used as a rough draft for further development of an instrumental piece. A high level of composer interference is needed when moving from a MIDI file to a more refined instrumental writing. Literal adaptations of converted
data are often impossible, and even undesirable (see an example of this process in Figure 4).

The second method is rather “manual”: the original analysis file is printed out just as it is (a long list of numbers, as shown earlier; see pages 59-61), and the composer’s task is to go through those numbers and “orchestrate” them. As a consequence of this very process, innumerable compositional decisions arise and have to be made, from shaping of musical ideas to issues of orchestration. The resulting instrumental writing is a unique product of this interaction between composer and raw data. This “manual” method is time consuming, but the challenges of dealing with raw data bring, in our opinion, extremely interesting compositional situations that push the limits of a composer’s usual style.

Figure 4: Raw conversion of MIDI data (top) and one possible composed excerpt derived—or rather “sculpted” from it (bottom).
3.3 Techniques II: composition on speech pitches

Another important process of acoustic composition used in the first half of *The Spell of Speech* is the relative control over the pitches of speech. The goal was to reach a level of musical control over the pitch component of speech without letting the result sound too distant from plausible speech. Direct inspiration for this idea came from Schoenberg’s *Pierrot Lunaire* (in which the voice never sounds completely sung nor spoken, which is a particular characteristic of this technique, known as *Sprechgesang*) and from many works by American composer Robert Ashley. Specifically, Ashley describes with some precision his attempt to control pitch (see chapter 2 above, page 48-49), and our composition using the actress’ voice was based on a free interpretation of Ashley’s technique.

The use of a sustained tone in the background serves as reference pitch for the actress. The pitch was chosen after analyzing the vocal range of the actress’ natural speech, in order to find a comfortable region for pitch movement that would keep the naturalness of her speech. The score indicates the pitches that the performer is consciously seeking while saying the corresponding words. As in Schoenberg’s work, the “note” is often abandoned as soon as it is reached; as in Ashley’s work, the pitch composition is developed within the range of nearly normal speech. The theatrical intentions that were sought were mostly composed during rehearsals; the intonation contours required for certain attitudes or meanings were evaluated, and our pitch composition took in consideration all of these aspects. The goal was to reach a sort of natural sounding speech with carefully composed pitch movements “behind” it, a structure that is not apparent at
first but that gives a unique character to that specific sequence of words. The relationship with the ground tone occasionally serves to highlight such construction. The creation of high-level structures such as motifs is also an important strategy for linking speech and music (the “well, well” motif is one of the most obvious; refer to first pages of score and first 10 minutes of the piece).

It should also be mentioned that the ground tone was composed in a way that it would not be just a flat, static sine wave. Internal movement based on the harmonic series come in and out during the time the tone is present. The movement is clear enough to provide some musical variation to this functional background, and discreet enough not to disturb the perception of the actress’ speech.

3.4 Techniques III: the live-electronics

All the live-electronics used in *The Spell of Speech* were created using Max/MSP (Cycling’ 74). All of the pre-composed electro-acoustic parts also were triggered using Max/MSP, with the exception of the “radio” scene (score line 11, cue “scene8”, CD 1, Track 1, 11’55”), which was played from a CD in the May 16th 2004 concert.

The composer controlled the Max/MSP patches during the performance. A total of 13 cue points were defined in advance to trigger pre-composed electronic sounds and/or live processing of the actress’ voice (the voice was being captured by a wireless microphone). The function of each cue will be explained below, along with the corresponding processes used. The cues are shown in the score as boxed texts.

The piece is divided into two patches. *The_Spell_of_Speech_A* (Figure 5) controls the sound flow from score line 1 to score line 10. The radio scene (score line 11), lasting
approximately 5 minutes, was played back from a CD. This gave more than enough time to close the first patch and open the second patch, *The_Spell_ of_Speech_B* (Figure 6). This second patch controls the piece from the end of the radio scene to the conclusion of the piece (score lines 12 to 15). These patches are not combined due to technical problems encountered when we tried to merge them (too much CPU was being used, increasing the risk of a computer crash or simply making the system less reliable). There are probably ways of solving such problems, and an ideal situation would be to have the whole piece within a single Max/MSP patch, but due to time constraints, we preferred to keep them as separate patches for the concert.
Figure 5: First of two Max/MSP patches used in The Spell of Speech.

- **voice1**
- **voice2**
- **voice3**
- **voice4**
- **voice5**

- **harmonizer**

**Recorder**

- $s_1$ = intro
- $s_2$ = cry
- $s_3$ = cry for help
- $s_4$ = standard
- $s_5$ = whispers
- $s_6$ = well well...
- $s_7$ = laugh fade out
- $s_8$ = move the chair
- $s_9$ = radio
- $s_{10}$ = conversation of visitors
- $s_{11}$ = phasing in
3.4.1 Cue “scene1”: beginning of the piece. It triggers the pre-composed electro-acoustic introduction, based on speech analysis and resynthesis with flute samples as described earlier. The introduction ends; the ground tone comes in, also turning up the volume of the actress’ microphone (voice 1 in the patch). The tone serves as background for the actress part (score lines 2, 3, 4 and 5).
3.4.2 Cue “scene2”: end of score line 5, after “well, well” (CD 1, Track 1, 4'55”). This turns on the process of voice accumulation using delay lines (voices 2, 3, 4 and 5 in the patch). The actress will start the “cry” section. Several layers of crying will generate a dense texture of sustained “notes”. Each delay line modifies the original cry according to its own delay parameters. Voices 2 to 5 are basically the same subpatch being driven by different parameters. Theses voices take advantage of the Doppler effect created by continuously changing delay lines in the Max object tapout~. Here is the basic structure of this subpatch called “dopplerdelay”:

![Diagram of dopplerdelay](image)

Figure 7: Patcher dopplerdelay is the core of most of the voices used in the main patch (Figure 5 and 6). 
What happens in this subpatch is the following: once a global delay time is set (valid for both channels), this delay is continuously and independently changed for each channel. A concrete example can help explain what happens. Suppose the global delay time (GD) is 6000 milliseconds (leftmost number in the subpatch). This means that any input sound will generate a delayed copy of itself in the output after 6 seconds. Then the two numbers on the top (called “deviation time” — DT) and the two other corresponding numbers (called “deviation amount” — DA) function in the following way: if we set a deviation amount of 500 milliseconds for each channel (L and R), this means that the global delay time of 6000 seconds will fluctuate between 6000 and 6500 milliseconds for the right channel (that is, GD to GD + DA), and between 6000 and 5500 milliseconds for the left channel (that is, GD to GD – DA). The other pair of numbers to be set is the deviation time (DT). In this example, if we set the deviation time to 1000 milliseconds for each channel, this means that one second is the time period in which those deviations will occur (the time in which the delay time will abandon GD, reach GD±DA and go back to GD). The result of such oscillation in the delay time is the Doppler effect. Delay times increasing in real time make the sound slide down in pitch; decreasing delay times have the opposite effect. The larger the deviation time (DT), the smoother the Doppler effect (less pitch deviation). The smaller the deviation time, the more prominent the Doppler effect. This interacts with the deviation amount (DA): the larger this number, the sharper the pitch deviation will be, and vice-versa. In sum, short deviation times (DT) coupled with large deviation amounts (DA) result in extreme pitch changes due to the Doppler effect. Large deviation times with small deviation amounts render an almost
imperceptible Doppler effect. Numerous intermediary cases can be obtained, not to mention the variations that are possible if we treat the two channels separately (with different DAs and DTs). Since in any case the pitch deviations occur in cycles going back and forth from the center delay time (GD), the sonic results can be compared to a vibrato with variable speeds and pitch ranges.\(^5\)

That is the basic structure of voices 2, 3, 4 and 5 in *The Spell of Speech A* (Fig. 5) and of voice 5 in *The Spell of Speech B* (Fig. 6). The reason for the existence of different voices is to take advantage of the possibility of multiple delay lines using different parameters, all of them accumulating to obtain a specific sound texture. In case of cue “scene2”, these different parameters are stored in the preset object (bottom of main patch). Each voice is set to bring back the input sounds with different transformations after several seconds, like a canon: voice 2 comes in after 15 seconds from the first input sound, voice 3 comes after 22, voice 4 after 30 and voice 5 after 45 seconds. Individual faders placed before the final output control the mixing of all voices plus direct sound.

While the cries accumulate, there is a hidden timer running that was also activated by cue “scene2”. By the end of a little more than a minute, a pre-composed electro-acoustic sound is triggered and is then mixed together with the ongoing voice layers

3.4.3 Cue “scene3”: the voice layers start to fade out after this cue, so that the electronic “cry” becomes prominent (score line 6, CD 1, Track 1, from 6’30” on). This electronic sound (originated from the analysis/resynthesis process explained in section 3.2) naturally fades out leaving the actress alone again.

3.4.4 Cue “scene4”: this triggers a chord derivate from the flute sounds of the beginning of the piece. This chord highlights a break in the acting flow just before the
presentation of a special piece of text, which we call “the mandarin story” (we will talk more about it later). It is interesting to note how many parameters (musical, textual and visual) can work together to convey a specific intention. Since this mandarin story is of fundamental importance to the piece, it is announced by this unexpected chord followed by a significant *diminuendo* of the background tone; also the actress stands up and whispers for the first time (middle of score line 7, CD 1, Track 1, 8’05”)

3.4.5 Cue “4eco”: following the need of emphasizing this moment of the composition, cue “4eco” triggers a pre-composed echo of the actress voice with the same text that was just said, but with a scrambled order of individual sentences and a background of other unintelligible voices (end of score line 7, CD 1, Track 1, 8’37”)

3.4.6 Cue “scene5”: this turns on two simultaneous live-electronics processes. One is the harmonizer subpatch that can be seen below voice 4 on the bottom of Figure 5. This subpatch simply transposes the voice from the microphone input and outputs it on three different pitches, creating a chord that builds up on every word the actress says. The second process activated by this cue is a long 38 seconds delay line that will bring back the sequence just spoken in a modified version due to the Doppler effect settings (score line 8). The actress uses the entry of this delay line to cue the starting point and the duration of her part in score line 9 (CD 1, Track 1, 9’51”).

3.4.7 Cue “scene6”: it is a simple fading out of the above mentioned delay line. It should make the fade out during the laugh; the background tone also fades out at this point, once and for all, leaving the actress in almost silence with her last “well, well” motif.
3.4.8 Cue “scene7”: after a short moment of silence, the activation of this cue plays back a pre-composed electro-acoustic part that is a variation of the introduction (long and low tones with bright high notes during their resonance time; all derivate from flute sounds. CD 1, Track 1, 10’51”). Acting together with these sounds, the actress presents for the first time a longer passage of the text. A slight processing of her voice creates a speech shadow that follows her words. This is made once again by one of the delay line voices. The difference now is that the global delay time is set to be almost instantaneous (less than 100 milliseconds) and a more pronounced Doppler effect is programmed, in a way that subtle and abnormal glissandi emerge from her natural intonation patterns.

3.4.9 Cue “scene8”: this cue was supposed to trigger the long electro-acoustic solo part that we call the “radio” scene (CD 1, Track 1, 11’55”). However, the need of having two separate patches for the first and second halves of the piece made us to take advantage of this moment to close one patch and open and set up the other, while the electro-acoustic sounds for this scene were played back from a CD.

3.4.10 Cue “scene8end”: this is the first cue in the second patch (figure 6, The_Spell_of_Speech_B). As one can see in the DVD, the actress has remained static for about five minutes listening to the radio (the electro-acoustic solo played back from a CD). By the end of the “radio” section, she is about to turn off the radio when cue “scene8end” triggers a sound file with another whispered version of the mandarin story (CD 1, Track 1, 17’20”). The sound file fades out as the text approaches the end, and the actress takes over the whispering and concludes the text by herself.

3.4.11 Cue “scene9”: The actress finally turns off the radio; at this moment cue “scene9” is activated (CD 1, Track 1, 17’55”). From this point on the actress delivers the
longest portion of text of the entire piece (see score line 13). The live-electronics create an accompaniment to it—subtle enough to preserve the intelligibility of what is being said, but present enough to counteract with the live voice. This accompaniment is made of a heavily processed version of the very input text being fed by the actress and delayed by about 10 seconds. All this happens inside the subpatch under the title “voice 6: other delays” that can be seen in the bottom of Figure 5. The original voice from the microphone first passes through comb filters, set to create highlight specific frequency regions; then this filtered signal passes through another harmonizer module, increasing the harmonic density of the speech flow. This is then sent to the main output, but it also feeds voice 5 (the only remaining dopplerdelay subpatch in this main patch), creating another layer of sound that is mixed in the output. The audible result is a sound stream that has clearly some relation to the spoken words that have been just heard, but not in a clear manner all the time. This stream has a continuous harmonic color with a few internal changes in the course of about two minutes. Slight harmonic shifts are made by changing a few parameters in the comb filter and harmonizer modules.

3.4.12 Cue “9break”: near the climax of the text, this cue is responsible for setting off a dynamic adjustment of the background sounds (CD 1, Track 1, after approximately 19’35”). Basically it activates amplitude envelopes that decrease the accompaniment level, though not completely; also the mixing of the accompaniment layers is slightly changed, before a timer finally activates a complete fade.

3.4.13 Cue “scene10”: this should be triggered exactly before the actress says “Am I in love?” (CD 1, Track 1, 20’04”). All the text from that point on will be subject of a “phasing in” process, or some sort of a dynamic stretro. Basically, a delay line is set up to
come five seconds after the actress starts this excerpt. But then the delay line moves gradually from five seconds to zero. In other words, it sounds like the delay line is “chasing” the original voice, until it actually catches up with it by the end of the text (“I am the one who waits”). A simple delay line moving this way would have a higher pitched voice as a result. Since we wanted the original and the delayed line to remain as similar as possible, to emphasize the perception of “the same voice” running after itself, we have added an appropriate amount of transposition before the output of the delayed voice in order to cancel out the Doppler effect at this time. By the end of this “phase in” process, when the actress finishes the text (score line 14, CD 1, Track 1, 20’29”), a pre-composed electro-acoustic passage fades in and connects this section with the next.

3.4.14 Cue “scene11”: simply triggers a sound file that is the “coda” of the piece (CD 1, Track 1, 21’05”). This passage was entirely composed through the analysis/resynthesis process described in section 3.2.

3.5 The Spell of Speech: formal organization

After this technical overview concerning how musical events were structured and implemented, let us summarize now the overall formal shape of The Spell of Speech. A more detailed guide for the succession of events can be found in the score provided in Appendix I. In relation to the text (Waiting, by Roland Barthes. See full text in Appendix IV), the piece progresses from the minimum to the maximum presence of words. The entire section A takes about ten minutes to develop on just a few words; only in the end the first paragraph is fully presented. The last paragraph of the text, to which we refer as the “mandarin story” is a central piece to the overall comprehension of the composition.
It is presented as a parenthesis in the middle of section A, in a way that it sounds “outside” of the main line of events. It is repeated two times more (in the middle and at the end of section B), to make sure that the story is well understood by the listener. This mandarin paragraph is:

“A mandarin fell in love with a courtesan. ‘I shall be yours’, she told him, ‘when you have spent a hundred nights waiting for me, sitting on a stool, in my garden, beneath my window’. But on the ninety-ninth night, the mandarin stood up, put his stool under his arm, and went away.” [our underline]

In all three appearances, the last line (underlined above) is omitted. It is precisely this last line that is used to conclude the whole piece, making the final link with the previous incomplete presentations of the mandarin story.

Section C continues the path started by section A, that of a progressive presentation of the main line of text. If on the one hand section A is slow in its speed of text presentation, reaching the end of a single paragraph only after a long time, on the other hand section C is much faster. It is actually an “explosion” of text that coincides with the final moments (the climax) of the piece.

Below we describe the sub-sections for A, B and C.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>B1</td>
<td>C1</td>
</tr>
<tr>
<td>A2</td>
<td>B2</td>
<td>C2</td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td>C3</td>
</tr>
<tr>
<td>A4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Schematized formal structure of *The Spell of Speech.*
A1: Electro-acoustic introduction. Presents the main sonic material that will be used by the electro-acoustic part. Prepares the entrance of the actress on stage.


A4: First presentation of the mandarin text, as a parenthesis within the main story line.

A5: Return and development of A2 materials. Actress progressively introduces more words in the sentence she is trying to say. End of melodically composed speech. Reference tone disappears.

A6: Episode: first ‘burst’ of text after all the previous musical and theatrical hesitation. Electro-acoustic part presents a variation on A1. Actress prepares stage for the radio scene.

B1: Radio scene. Static harmony, actress immobile. In the middle of this electro-acoustic solo, noisy processed voices appear in the foreground repeating the mandarin story for the second.

B2: Short coda of radio scene. After last chord fades away, the mandarin story is presented for the third and last time, now very clear and without any acoustic or electro-acoustic interference.

C1: Beginning of fast presentation of text allied with more movement on stage and more agitated electro-acoustic background (derived from the text itself).

C2: Conclusion of previous sequence by means of the “phasing in” process described earlier (see section 3.4.13). “I am the one who waits” is the last sentence pertaining to the main story line before the end of the piece.
C3: Coda: long electro-acoustic development of the materials presented in A1. Actress silently removes all furniture from stage. In the last seconds of the electro-acoustic part, she returns to stage to speak the very last sentence, which is the conclusion of the mandarin story. This gives the final “key” for the understanding of that suspended plot that was insistently repeated throughout the piece.

3.6 The concert

*The Spell of Speech* was premiered on May 16th 2004. The concert itself had the same name (see posters and program notes in Appendix V). All the other pieces presented in this concert were related to our research on musical composition and human speech, with the exception of the instrumental piece *Duo #38* and its correlate *Interlude 2*. All pieces were played without applause in between them. The concert was structured in the following way:

- [pre-Prelude]: while the audience took their seats, a composed soundscape of animal sounds (mainly birds and squirrels recorded in White River Junction, Vermont) was being played very softly in the background. This can be seen as an allusion to the origins of speech and communication in a “biological”, evolutionary sense; however, this is never stated anywhere, so it is up to the listener to figure out why those animal sounds were there in the beginning. (CD 2, Track 1)

- *Prelude* (2004): a short electro-acoustic piece based on the analysis/resynthesis process (see section 3.2). The main characteristic of this prelude is the acoustic guitar sound that was used as the basic sample to resynthesize speech data. Also,
in this case the frequencies of the original speech were substituted by the frequencies of a six-note chord that is the same that originated the entire “radio section” in *The Spell of Speech*. First elements of human voice (children at play) enter by the end of this prelude. (CD 2, Track 2)

- **Fonepoemas** (2003): electro-acoustic piece based on vocal improvisations on the phonemes from the name “Tânia”. One of the first results of the analysis/resynthesis process as described in section 3.2 was used in the last movement of *Fonepoemas*. (CD 2, Track 3)

- **Interlude 1** (2004): short electro-acoustic piece based on the actress’ voice. A nonsense text was generated with a “text-scrambler” tool that we have written in Perl script (see Appendix II); the text was then recorded by the actress on specific notes and combined to originate the central harmony presented throughout this piece. (CD 2, Track 4)

- **Duo #38** (2004): instrumental piece for piano and percussion. This piece had no direct relation to the central “theme” of the concert. (CD 2, Track 5)

- **Interlude 2** (2004): short electro-acoustic piece derived from previous versions of *Duo #38*. While this piece was played, the stage was arranged for the beginning of the next piece. (CD 2, Track 6)

- *The Spell of Speech* (2004): the final piece of the concert. (CD 1, Track 1; alternate recordings on Tracks 2 and 3)
3.7 Future work

*The Spell of Speech* has the high risk of being difficult to reproduce. It was composed in a very close collaboration with actress Hannah Chodos during the first months of 2004. Because of the nature of such a work, a great part of it can hardly be transcribed in a traditional score, especially the theatrical elements. However, it is not impossible for the piece to be performed by another actress in the future and without the presence of the composer to assist the process. In order to make this possible, the creation of a very detailed score is needed. Such score does not yet exist. The one presented in Appendix I is to be considered simply as a listener’s guide. It does not contain all the elements necessary to an interpretation of the piece by a performer. Besides the performer’s score, a well-documented guide for the live-electronics is also necessary. At first, it may seem that the cues provided in the performer score can suffice for anyone who is controlling the Max/MSP patches, since the live-electronics task is no more than clicking on the right buttons at the right times. However, an actress that wishes to present this piece in the future will need much more than a “button-clicker” during all the rehearsal process and concert. A technical director (or “electro-acoustic conductor”) with a good understanding of compositional issues is needed in order to deal with musical problems such as possible “tweaking” of numbers inside the Max/MSP patches to adjust timings and volumes that are necessarily particular to each performer and performance, as well as problems of overall balance, levels for the electronic voices adjusted according to the room and the microphone in use, not to mention the diffusion of sounds in the concert space. As to the theatrical elements, two options are possible. The performer’s score may contain a
detailed description of all the scenes, gestures and intentions, and any new performer may attempt to follow these indications as close as possible. However, it is also acceptable for a performance to have a competent theater director with a different view for the creation of scenes, the stage set, the intentions to be given by the actress to the text (provided that they do not conflict with the passages of musically composed speech) and so forth. In this case, a new version of the piece could be created with completely new theatrical elements.

In short, the work of documentation and definition of a score, in a comprehensive sense, still needs to be done to make it possible for the piece to survive for live performance.

---

1 This technique was first presented to us by the Brazilian composer Ignacio de Campos ca. 2001, using Open Music and Csound.
2 This is a dB scale particular to Audiosculpt. It seems to range from approximately –40 to +10 dB. These values were obtained by trial and error experience. The software documentation gives on explanation or hint as to the reason for such range.
3 This Open Music patch (which deals with averaged partials only) is essentially the same developed by Ignacio de Campos in 2001 (see note 1), to which we have added a few new features. The development of the non-averaged analysis/resynthesis option is something new, though. I have implemented it in 2004 with the aid of Eric Lyon and Stefan Tomic.
4 The technique we used to accomplish this was shown to us by Jean-Claude Risset during his stay at Dartmouth College (April-May 2004). It basically consists of superimposing simple waves (with approximately the ten first harmonics at full amplitude) with slight frequency differences (of fractions of a Hertz; for example, 90.20 Hz, 90.25 Hz, 90.30 Hz and so forth).
5 The Doppler effect is a sort of “artifact” that happens with this specific use of the Max/MSP delay objects tapin~ and tapout~. In terms of the physics of such effect, what our subpatch is simulating is actually a variation in the speed with which a moving sound source passes by a fixed listener, as well as its direction. What is somewhat “unreal” in this particular use of the Doppler effect is the fact that the “moving object” makes cycles back and forth in relation to the listener, causing the pitch to slide up and down continuously in a given period, instead of moving in a single direction and fading away as in the classical example of cars passing by.
APPENDIX A (score)
The Spell of Speech
(2004)

music by Bruno Ruviaro (1976)
text by Roland Barthes (1915-1980)

scene1 << boxed texts indicate Max/MSP cues

Slow

electro-acoustic sounds

actress

VOICE ALWAYS
OCTAVE BELOW

2

[steady tone continues]

(impatient, insecure, a little petulant, but with a soft voice; pretending to be self-confident)

Well, well

well, I I well...

(giving up; speaking to yourself)

well, well
[steady tone — Ab]

Well, I am...

(open-mouthed breath)

Well, I am...

(lament; gliss. down after slightly reaching the notes)

Well, I am...

(scene2)

Well, I am...

Well, I am...

(f)

Well, I am...

(mp)
(delay lines: modified versions of the cry will accumulate);
entrance of electronic sounds indicate start of global decrescendo;
tape brings back the Ab tone after this process)

ca. 2 minutes

(chiusa, moving gradually to crying;
long but variable durations;
each new note a little higher in pitch)

(mm...)

(after electronic sounds come in, the cry "calms down"
and actress returns to bocca chiusa centered on the main
note, varying microtones around it)

scene 3

scene 4

I am waiting.

scene 5

Well, well well... will.

A mandarin fell in love with a courtesan.
"I shall be yours", she told him, "when you have spent a hundred nights waiting for me, sitting on a stool, in my garden, beneath my window"
Well, well...  ...will come back modified

I am waiting for an arrival, a return, a promised sign. This can be futile and immensely pathetic. There is a piece, I think it is by Schoenberg, where a woman waits for her lover, alone, at night, in the forest. I am waiting for no more than a telephone call, but the anxiety is the same. Everything is solemn. I have no sense of proportions.

I have received orders not to move. (she turns the radio on)
A mandarin fell in love with a courtesan.
"I shall be yours", she told him, "when you have spent a hundred nights waiting for me, sitting on a stool, in my garden, beneath my window"

(actress turns off the radio after concluding the story)

There is a scenography of waiting. I organize it, manipulate it, cut out a portion of time in which I shall mime the loss of the loved object and provoke all the effects of a minor mourning. This is then acted out as a play. The settings represent the interior of a café. Prologue. We have a rendezvous. I am waiting. I, the sole actor of the play, discern and indicate the other's delay. This delay is as yet only a mathematical, computable entity (I look at my watch several times). The prologue ends with a brainstorm: I decide to take it badly, I release the anxiety of waiting. Act I now begins. It is occupied by suppositions: "Was there a misunderstanding as to the time, the place?" I try to recall the moment when the rendezvous was made, the details which were supplied. "What is to be done?" (anxiety of behavior) "Try another café? Telephone? But what if the other comes during these absences? Not seeing me, the other might leave.." etc. Act II is the act of anger. I address violent reproaches to the absent one. "All the same, he could have... He knows perfectly well... oh, if he could be here so that I could reproach him for not being here!". In Act III, I attain to anxiety in its pure state: the anxiety of abandonment. I have just shifted in a second from absence to death. The other is as if dead. Explosion of grief. I am internally livid. That is the play. "Am I in love?" Yes, since I am waiting. The other never waits. Sometimes I want to play the part of the one who doesn't wait: I try to busy myself elsewhere, to arrive late, but I always lose at this game. Whatever I do I find myself there, punctual, even ahead of time. The lover's fatal identity is precisely: I am the one who waits. I am the one who waits.
(tape fades in with fragments of pre-recorded voice; more excerpts of the text are presented together with synthetic sounds derived from it)

(still electro-acoustic sounds only; strong low note begins; dense texture develops for ca. 1'30")

(right after the beginning of the low note and dense texture, actress stands up and slowly removes the furniture from stage)

(electro-acoustic sounds dissolve into bright, high-pitched grains decrescendo and rallentando)

(upon the last low note that appears in the middle of the "dissolving" electro-acoustic sounds, actress return to stage to speak her last sentence)

But on the ninety-ninth night, the mandarin stood up, put his stool under his arm, and went away. (blackout)
APPENDIX B (Perl scripts)

The Perl script below was written by Stefan Tomic at our request. Its purpose is to deal
with the non-averaged partials data extracted from Audiosculpt in order to generate
appropriate Csound scores for resynthesis.

```perl
# Perl script to convert a partial file from an Audiosculpt analysis and
# put the set of amplitudes and frequencies for each
# partial into separate text files, to be used with Csound using GEN 23.
#
#
# Author: Stefan Tomic, Dartmouth College 3/24/04

# Accept variable arguments to the perl script. If none are entered, print out the order of
# the arguments
# the rest of the if else clauses fill in the variables for the missing arguments
if ($#ARGV < 0) {
    print "Syntax: perl parsepartials.pl <partial file> <score file> <speed> <panning>\n";
    print "Partial file name: ";
    $infilename = <STDIN>;
} else {
    $infilename = $ARGV[0];
}

if ($#ARGV < 1) {
    print "Score file name: ";
    $scorefilename = <STDIN>;
} else {
    $scorefilename = $ARGV[1];
}

if ($#ARGV < 2) {
    print "Speed: ";
    $speed = <STDIN>;
}
```

94
else {
    $speed = $ARGV[2];
}

if ($#ARGV < 3) {
    print "Panning: ";
    $pan = <STDIN>;
} else {
    $pan = $ARGV[3];
}

#Open the input Audiosculpt file and the output Csound score file
open INFILE, "<$infilename" or die "cannot open filename $infilename :$!";
open SCOFILE, ">$scorefilename" or die "cannot create filename $scorefilename :$!";

#initialize partial number and F table numbers in Csound score
$partialnum = 1;
$ampF = 2;
$frqF = 3;

#parse the whole partial file and split it up by words, stored in the array @entries
while(<INFILE>) {
}

#the third entry in the partial file is the number of partials
$npjartials = $entries[2];

#start reading entries at entry #4
$i = 3;
# keep incrementing $i until the total number of entries is read (@entries is the length of the array)
while($i < @entries) {

    # if an entry has the word points, then we need to read in all the data for that partial
    if($entries[$i] eq "POINTS") {
        $i++;
        $numpoints = $entries[$i++];
        $starttime = $entries[$i];

        # open the amp_data and freq_data files for that partial
        $outampfilename = "amp_data" . $partialnum . ".txt";
        $outfrqfilename = "freq_data" . $partialnum . ".txt";
        open AMP_OFILE, ">$outampfilename" or die "cannot create filename $outampfilename :$!";
        open FRQ_OFILE, ">$outfrqfilename" or die "cannot create filename $outfrqfilename :$!";

        # read in all the points for that partial and output them to amp_data and freq_data
        for (1..$numpoints) {
            $time = $entries[$i++];
            $frq = $entries[$i++];
            $amp = csdb($entries[$i++]); # call the csdb subroutine (below) on amp data
            printf AMP_OFILE "%5.3f
",$amp;
            printf FRQ_OFILE "%5.3f
",$frq;
        }

        # close the files and calculate the total time in seconds for that partial
        close(AMP_OFILE);
        close(FRQ_OFILE);
        $partialnum++;
        $endtime = $time;
        $dur = $endtime-$starttime;

        $tablelength = pow2($numpoints);

        # build a string for the output Csound score. This needs to be added for each partial
        $ftablist = $ftablist . sprintf("f%3d 0 %5d -23 \
"%s"\n",$ampF,$tablelength,$outampfilename);
        $ftablist = $ftablist . sprintf("f%3d 0 %5d -23 \
"%s"\n",$frqF,$tablelength,$outfrqfilename);
        $notelist = $notelist . sprintf("i1 %6.3f %6.3f %2d %2d %4d %5.2f %3.2f\n",$starttime,$dur,$ampF,$frqF,$numpoints,$speed,$pan);
        undef $starttime;
    }
#increment F table numbers before going to the next partial
$ampF+=2;
$frqF+=2;
}
i++;
}

#output the strings to the .sco file
print SCOFILE "f1 0 8192 10 1\n";
print SCOFILE $ftablist;
printf SCOFILE "\n\n\n";
print SCOFILE $notelist;
close(SCOFILE);

#subroutine to calculate the next higher power of 2 of a number
sub pow2 {
    my($numin) = shift;
    $comp = 2;
    while ($comp < $numin) {
        $comp = $comp*2;
    }
    return $comp;
}

#subroutine to split up a file into tokens
sub tokenize {
    $_ = $_[0];
    s/
/\n/g;
    s/\n//;
    s/\n/\n/;
    s/([.,!?:;\n])/\n$1/g;
    s/(["'`])([^
])/\n$1$2/g;
    s/([^
])([.,\n])/\n$1$2/g;
    s/\n([A-Z])\./\n$1./g;
    s/\n.([^
]1)/\n$1\n/g;
    s/((?![A-Z])\n)/\n$1\n/g;
    return(split(/
/,$_));
}
# subroutine to convert audiosculpt amp values in dB to Csound amp values in dB
sub csdb {
    my($asdb) = shift;
    $asdb = $asdb + 40;
    $csdb = 54 + $asdb/50*36;
    return $csdb;
}

The Perl script below was written as an extra tool to get an original text as input and make permutations of its letters and sequences of letters (Markov chain process). After the script we present an example of this process. The text for Interlude I (CD 2, Track 4) was generated with the help of this script.

# TEXT_MACHINE
# Author: Bruno Ruviaro, January 2004
#
# Type in a text and choose a level of distortion.
# Small integer numbers yield more chaotic results. Larger numbers keep more and more
# of the original text (scrambled pieces).

$array_text = ();
%hash_text = ();
$final = 0;

print "Type in something: ";
$string = lc(<STDIN>); #lc makes it all lowercase
chomp $string;

# puts the text into an array
foreach $byte (split //, $string)
{
    $array_text[$cnt++] = $byte;
    $final = $final + 1;
}
# Create an appropriate hash (hash of arrays):
# $i is always telling the position of a char in the main array_text.
# Each hash KEY is a character from the text (forloop builds the whole hash).
# The VALUE for a given hash KEY is a reference to an array of numbers.
# These numbers are the position of a char in the main array_text.
# The first value for this array is the first position of that character in the array_text.

for ($i = 0; $i < $final; $i++)
{
    $char = $array_text[$i];
    push @{$hash{$char}}, $i;
}

$again = "y";

while ($again eq "y")
{
    print "Choose synthesis level (any positive integer number): ";
    $level = <STDIN>;
    chomp $level;

    print "Choose approx. length of new text: ";
    $length = <STDIN>;
    chomp $length;
    print "\nLevel => $level\nLength => $length\nNew text:\n"

    if ($level eq "1")
    {
        $i = 0;
        while ($i < $length)
        {
            $index = int(rand($final));
            $last_added = $array_text[$index];
            print "$last_added";
            $i++;
        }
    }
    else    # That means, if level not equal to 1 (all other levels)
    {
        # Choose first character:
        $index = int(rand($final));
        $last_added = $array_text[$index];
        print $last_added;
    }
# And now go into the real other level:
$t = 1;  # 1, since the first letter is "given" and already printed out;
$final = @array_text;
while ($t < $length)
{
    @char_pos = @{$hash{"$last_added"}};
    $index = int(rand(@char_pos));
    $desired_position = @char_pos[$index];
    for ($i = 1; $i < $level; $i++)
    {
        if ($desired_position eq ($final - 1))
            {$desired_position = -1;}
        $last_added = $array_text[$desired_position + 1];
        print "$last_added";
        $desired_position++;
        $t++
    }
}

print "\n\n";

print "Would you like to try again? (same original text) (y/n) ";
$again = <STDIN>;
chomp $again;
}

# who's last_added?
# how many of it are there? (look for it in %hash_text; get the size of array stored in there; the size of array is the number of occurrences of that letter)
# generate a random number within that range (0 - ArraySize).
# Use this number as index for picking a number inside that Array.
# The number obtained is actually the position in @array_text we want to start with
# print it, and print the following character (as many as the chosen $level will require)
# store this last character as $last_added
# repeat the process.
Example of use of the script above

Original Text:

I am waiting for an arrival, a return, a promised sign. This can futile and immensely pathetic. There is a piece, I think it is by Schoenberg, where a woman waits for her lover, alone, at night, in the forest.

Resulting texts (“level” means the degree of intelligibility; the smaller the number, the least intelligibility; “length” is simply the number of characters in the new text):

Level => 1
Length => 200
New text:
iic iairfmh,tor amn w oeengsohma mu,rirruhwgt ,ioonofaoanf cyhfeargncsol,r hgrowewtimbr li m be t,arnn vafelrra ai teonor e,iarimhrbri ui yaohss ece lor,vrbm,aorhs1,.t f klmrsea,eewit hiath h,n

Level => 5
Length => 200
New text:
for alonerg, immely pathest.ign. can am wheretic. thetic. the, athe for arriver, for it ised and for signberght, a woenbere in ther l, a i this c. ther lover loval, is by sc. ther le ane, arriver, in t

Level => 10
Length => 200
New text:
nd immensest.i am woman waithis can for an arromised sink it is waiting forest.i and immens can futis can futurn, a pr an arrival, a ret is by sc. there i think ithis can for an arrn, a prom waiting is by sch

Level => 15
Length => 200
New text:
oman waits for piece, i think it is by schone, at night, woman waits fomised sign. there a woman warrival, a retutile and immensely pathetic. this can futin the forest.iting for an ar, alone, at nink it is by sc

Level => 20
Length => 200
New text:
, i think it is by schoenberg, where a woman waits. i am waiting for the forest. i am waiting and immensely alone, at night, in the forest. a piece, i there a woman waits for her lover, alone.

Level => 25
Length => 200
New text:

it is by schoenberg, where, alone, at night, in the forest. i am waiting for an arrival, a return, at night, in the forest. think it is by schoenberg for an arrival, a return, a promise can futile and immense.
APPENDIX C (Csound files)

Typical Csound files used in the analysis/resynthesis process described in section 3.2.

The basic structure of the orchestra file dedicated to the non-averaged partials case was written with the aid of Eric Lyon.

; ORCHESTRA (averaged partials case; using sine wave for the resynthesis)

sr=44100
kr=4410
ksmps=10
nchnls=2

; Basic use of averaged partials. Use a score file save from OpenMusic.
; Use Instrument 2 to mix in the original sample.
; Amplitudes should come from OM already in linear values.
; p6 value controls panning L-R
; Optional p7 may be used to redirect to different function tables.

instr 1

kenv linseg 1, p3*0.1, 0.9, p3*0.3, 0.3, p3*0.6, 0
kamp = p4*kenv
a1 oscili kamp, p5, 1
outs a1*(1-p6), a1*p6

endin

instr 2

a1 diskin "Short.aif", 1
a1 = a1/2
outs a1, a1

endin

; SCORE (averaged partials case; using sine wave)

f1 0 8192 10 1

<table>
<thead>
<tr>
<th>p1</th>
<th>p2</th>
<th>p3</th>
<th>p4</th>
<th>p5</th>
<th>p6</th>
</tr>
</thead>
<tbody>
<tr>
<td>i2</td>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i1</td>
<td>0.1100</td>
<td>0.5520</td>
<td>4470</td>
<td>171.7760</td>
<td>0.9600</td>
</tr>
<tr>
<td>i1</td>
<td>0.1220</td>
<td>0.5400</td>
<td>3454</td>
<td>346.6200</td>
<td>0.4100</td>
</tr>
<tr>
<td>i1</td>
<td>0.1920</td>
<td>0.2550</td>
<td>1037</td>
<td>482.6590</td>
<td>0.7200</td>
</tr>
<tr>
<td>i1</td>
<td>0.3890</td>
<td>0.1100</td>
<td>756</td>
<td>1227.7600</td>
<td>0.2500</td>
</tr>
<tr>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>-----</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>i1</td>
<td>0.4300</td>
<td>0.1210</td>
<td>1017</td>
<td>1105.7960</td>
<td>0.2000</td>
</tr>
<tr>
<td>i1</td>
<td>0.4470</td>
<td>0.1970</td>
<td>1313</td>
<td>575.7410</td>
<td>0.0300</td>
</tr>
<tr>
<td>i1</td>
<td>0.4640</td>
<td>0.1110</td>
<td>899</td>
<td>753.9700</td>
<td>0.6800</td>
</tr>
<tr>
<td>i1</td>
<td>0.7080</td>
<td>0.1510</td>
<td>5642</td>
<td>225.0350</td>
<td>0.0000</td>
</tr>
<tr>
<td>i1</td>
<td>0.7140</td>
<td>0.1680</td>
<td>3138</td>
<td>675.5700</td>
<td>0.8100</td>
</tr>
<tr>
<td>i1</td>
<td>0.7200</td>
<td>0.1280</td>
<td>1106</td>
<td>905.4430</td>
<td>0.0000</td>
</tr>
<tr>
<td>i1</td>
<td>0.7260</td>
<td>0.1220</td>
<td>1300</td>
<td>1126.0660</td>
<td>0.4400</td>
</tr>
<tr>
<td>i1</td>
<td>0.8770</td>
<td>0.1040</td>
<td>3707</td>
<td>542.6060</td>
<td>0.5100</td>
</tr>
<tr>
<td>i1</td>
<td>0.8880</td>
<td>0.1690</td>
<td>1588</td>
<td>1530.9830</td>
<td>0.2400</td>
</tr>
<tr>
<td>i1</td>
<td>0.8940</td>
<td>0.1280</td>
<td>4300</td>
<td>262.6180</td>
<td>0.8200</td>
</tr>
<tr>
<td>i1</td>
<td>0.8940</td>
<td>0.0990</td>
<td>1825</td>
<td>1293.0730</td>
<td>0.1200</td>
</tr>
<tr>
<td>i1</td>
<td>0.8940</td>
<td>0.2790</td>
<td>978</td>
<td>1771.3660</td>
<td>0.4600</td>
</tr>
<tr>
<td>i1</td>
<td>0.9400</td>
<td>0.2270</td>
<td>1876</td>
<td>1130.3950</td>
<td>0.9600</td>
</tr>
<tr>
<td>i1</td>
<td>0.9640</td>
<td>0.1450</td>
<td>1468</td>
<td>1262.7190</td>
<td>0.9800</td>
</tr>
<tr>
<td>i1</td>
<td>0.9690</td>
<td>0.1690</td>
<td>1227</td>
<td>3483.4050</td>
<td>0.9700</td>
</tr>
<tr>
<td>i1</td>
<td>0.9750</td>
<td>0.0990</td>
<td>1023</td>
<td>3238.7920</td>
<td>0.1900</td>
</tr>
<tr>
<td>i1</td>
<td>0.9930</td>
<td>0.1160</td>
<td>1284</td>
<td>3391.0810</td>
<td>0.0500</td>
</tr>
<tr>
<td>i1</td>
<td>1.0100</td>
<td>0.1050</td>
<td>725</td>
<td>4062.9830</td>
<td>0.7600</td>
</tr>
<tr>
<td>i1</td>
<td>1.0160</td>
<td>0.2380</td>
<td>6874</td>
<td>281.4050</td>
<td>0.4100</td>
</tr>
<tr>
<td>i1</td>
<td>1.0330</td>
<td>0.1920</td>
<td>2366</td>
<td>861.2030</td>
<td>0.9400</td>
</tr>
<tr>
<td>i1</td>
<td>1.0450</td>
<td>0.1280</td>
<td>2980</td>
<td>561.0870</td>
<td>0.8600</td>
</tr>
<tr>
<td>i1</td>
<td>1.0570</td>
<td>0.1330</td>
<td>2178</td>
<td>1461.8780</td>
<td>0.7800</td>
</tr>
<tr>
<td>i1</td>
<td>1.2190</td>
<td>0.1280</td>
<td>1515</td>
<td>1971.1730</td>
<td>0.0000</td>
</tr>
<tr>
<td>i1</td>
<td>1.2190</td>
<td>0.1160</td>
<td>898</td>
<td>3596.0360</td>
<td>0.2500</td>
</tr>
<tr>
<td>i1</td>
<td>1.2540</td>
<td>0.2030</td>
<td>2088</td>
<td>658.0060</td>
<td>0.8800</td>
</tr>
<tr>
<td>i1</td>
<td>1.3870</td>
<td>0.4940</td>
<td>3538</td>
<td>440.2380</td>
<td>0.5400</td>
</tr>
<tr>
<td>i1</td>
<td>1.4280</td>
<td>0.1740</td>
<td>1302</td>
<td>2094.3930</td>
<td>0.4900</td>
</tr>
<tr>
<td>i1</td>
<td>1.4340</td>
<td>0.2200</td>
<td>2135</td>
<td>1745.1620</td>
<td>0.8800</td>
</tr>
<tr>
<td>i1</td>
<td>1.4510</td>
<td>0.1690</td>
<td>8109</td>
<td>737.8650</td>
<td>0.1900</td>
</tr>
<tr>
<td>i1</td>
<td>1.4570</td>
<td>0.1800</td>
<td>4975</td>
<td>1129.3020</td>
<td>0.4800</td>
</tr>
<tr>
<td>i1</td>
<td>1.4570</td>
<td>0.1280</td>
<td>4648</td>
<td>1387.9530</td>
<td>0.4600</td>
</tr>
<tr>
<td>i1</td>
<td>1.4570</td>
<td>0.1100</td>
<td>1916</td>
<td>3061.1250</td>
<td>0.7600</td>
</tr>
<tr>
<td>i1</td>
<td>1.4570</td>
<td>0.1340</td>
<td>2533</td>
<td>3397.6200</td>
<td>0.0200</td>
</tr>
<tr>
<td>i1</td>
<td>1.4630</td>
<td>0.1220</td>
<td>1217</td>
<td>2414.1670</td>
<td>0.5400</td>
</tr>
<tr>
<td>i1</td>
<td>1.4690</td>
<td>0.1270</td>
<td>1101</td>
<td>2783.1140</td>
<td>0.5600</td>
</tr>
<tr>
<td>i1</td>
<td>1.5620</td>
<td>0.1210</td>
<td>913</td>
<td>3170.2370</td>
<td>0.4700</td>
</tr>
<tr>
<td>i1</td>
<td>1.6080</td>
<td>0.1510</td>
<td>1270</td>
<td>854.3560</td>
<td>0.1400</td>
</tr>
<tr>
<td>i1</td>
<td>1.7760</td>
<td>0.2330</td>
<td>2000</td>
<td>9028.8700</td>
<td>0.8300</td>
</tr>
<tr>
<td>i1</td>
<td>1.7760</td>
<td>0.1110</td>
<td>1978</td>
<td>9962.1230</td>
<td>0.6900</td>
</tr>
<tr>
<td>i1</td>
<td>1.7760</td>
<td>0.2330</td>
<td>1883</td>
<td>10299.1020</td>
<td>0.3900</td>
</tr>
<tr>
<td>i1</td>
<td>1.7760</td>
<td>0.1050</td>
<td>3530</td>
<td>10381.4260</td>
<td>0.9100</td>
</tr>
<tr>
<td>i1</td>
<td>1.7760</td>
<td>0.2040</td>
<td>1346</td>
<td>10943.7150</td>
<td>0.0800</td>
</tr>
<tr>
<td>i1</td>
<td>1.7760</td>
<td>0.1980</td>
<td>1018</td>
<td>11057.8800</td>
<td>0.3200</td>
</tr>
<tr>
<td>i1</td>
<td>1.7820</td>
<td>0.1980</td>
<td>1759</td>
<td>8222.5480</td>
<td>0.2500</td>
</tr>
<tr>
<td>i1</td>
<td>1.7820</td>
<td>0.2030</td>
<td>2214</td>
<td>8572.4760</td>
<td>0.8000</td>
</tr>
<tr>
<td>i1</td>
<td>1.7820</td>
<td>0.1510</td>
<td>693</td>
<td>11426.6240</td>
<td>0.6600</td>
</tr>
<tr>
<td>i1</td>
<td>1.7880</td>
<td>0.2030</td>
<td>1134</td>
<td>7908.5850</td>
<td>0.7500</td>
</tr>
<tr>
<td>i1</td>
<td>1.7880</td>
<td>0.1280</td>
<td>1609</td>
<td>8356.5710</td>
<td>0.4200</td>
</tr>
<tr>
<td>i1</td>
<td>1.7880</td>
<td>0.1570</td>
<td>1862</td>
<td>9254.9270</td>
<td>0.1300</td>
</tr>
<tr>
<td>i1</td>
<td>1.7880</td>
<td>0.1740</td>
<td>2017</td>
<td>9714.8730</td>
<td>0.9200</td>
</tr>
<tr>
<td>i1</td>
<td>1.7880</td>
<td>0.0990</td>
<td>801</td>
<td>11304.4340</td>
<td>0.6500</td>
</tr>
<tr>
<td>i1</td>
<td>1.7940</td>
<td>0.1100</td>
<td>672</td>
<td>11513.6820</td>
<td>0.3600</td>
</tr>
<tr>
<td>i1</td>
<td>1.8000</td>
<td>0.1560</td>
<td>958</td>
<td>7672.6560</td>
<td>0.6700</td>
</tr>
<tr>
<td>i1</td>
<td>1.8050</td>
<td>0.1160</td>
<td>1907</td>
<td>9493.9510</td>
<td>0.3600</td>
</tr>
<tr>
<td>i1</td>
<td>1.8050</td>
<td>0.1340</td>
<td>1042</td>
<td>11232.3120</td>
<td>0.9800</td>
</tr>
</tbody>
</table>
### ORCHESTRA EXAMPLE (non-averaged case, using samples instead of sine waves)

sr=48000
kr=48000
ksmps=1
nchnls=2

giglobalamp = 101

instr 1

<table>
<thead>
<tr>
<th>i1</th>
<th>1.8110</th>
<th>0.1510</th>
<th>937</th>
<th>7268.7220</th>
<th>0.7700</th>
</tr>
</thead>
<tbody>
<tr>
<td>i1</td>
<td>1.8230</td>
<td>0.1620</td>
<td>2235</td>
<td>10679.7690</td>
<td>0.7200</td>
</tr>
<tr>
<td>i1</td>
<td>1.8230</td>
<td>0.1040</td>
<td>744</td>
<td>13341.7600</td>
<td>0.0600</td>
</tr>
<tr>
<td>i1</td>
<td>1.8340</td>
<td>0.1110</td>
<td>1136</td>
<td>7802.3110</td>
<td>0.3500</td>
</tr>
<tr>
<td>i1</td>
<td>1.8400</td>
<td>0.1690</td>
<td>1038</td>
<td>7077.4030</td>
<td>0.0300</td>
</tr>
<tr>
<td>i1</td>
<td>1.8460</td>
<td>0.1280</td>
<td>911</td>
<td>6226.6890</td>
<td>0.5400</td>
</tr>
<tr>
<td>i1</td>
<td>1.8520</td>
<td>0.1450</td>
<td>921</td>
<td>7560.1400</td>
<td>0.9400</td>
</tr>
<tr>
<td>i1</td>
<td>1.8750</td>
<td>0.1390</td>
<td>1662</td>
<td>8438.9490</td>
<td>0.9300</td>
</tr>
<tr>
<td>i1</td>
<td>1.8810</td>
<td>0.1280</td>
<td>1692</td>
<td>10033.8780</td>
<td>0.9500</td>
</tr>
<tr>
<td>i1</td>
<td>1.8920</td>
<td>0.1170</td>
<td>2133</td>
<td>8866.4650</td>
<td>0.2000</td>
</tr>
<tr>
<td>i1</td>
<td>1.8980</td>
<td>0.0990</td>
<td>823</td>
<td>6717.9890</td>
<td>0.3300</td>
</tr>
<tr>
<td>i1</td>
<td>2.2580</td>
<td>0.2850</td>
<td>7508</td>
<td>310.2680</td>
<td>0.4500</td>
</tr>
<tr>
<td>i1</td>
<td>2.2760</td>
<td>0.2200</td>
<td>4380</td>
<td>665.5440</td>
<td>0.3500</td>
</tr>
<tr>
<td>i1</td>
<td>2.3100</td>
<td>0.1160</td>
<td>1588</td>
<td>1341.1380</td>
<td>0.9600</td>
</tr>
<tr>
<td>i1</td>
<td>2.3450</td>
<td>0.0990</td>
<td>1981</td>
<td>968.8230</td>
<td>0.2100</td>
</tr>
<tr>
<td>i1</td>
<td>2.3510</td>
<td>0.1220</td>
<td>2663</td>
<td>2552.2380</td>
<td>0.7100</td>
</tr>
<tr>
<td>i1</td>
<td>2.3510</td>
<td>0.1220</td>
<td>891</td>
<td>9191.1790</td>
<td>0.7700</td>
</tr>
<tr>
<td>i1</td>
<td>2.3510</td>
<td>0.1280</td>
<td>780</td>
<td>9305.3990</td>
<td>0.4200</td>
</tr>
<tr>
<td>i1</td>
<td>2.3630</td>
<td>0.0980</td>
<td>2162</td>
<td>3938.0470</td>
<td>0.0200</td>
</tr>
<tr>
<td>i1</td>
<td>2.3630</td>
<td>0.1160</td>
<td>1013</td>
<td>4419.5740</td>
<td>0.8800</td>
</tr>
<tr>
<td>i1</td>
<td>2.3630</td>
<td>0.1160</td>
<td>896</td>
<td>9463.7140</td>
<td>0.0600</td>
</tr>
<tr>
<td>i1</td>
<td>2.3650</td>
<td>0.1160</td>
<td>845</td>
<td>10157.7850</td>
<td>0.9400</td>
</tr>
<tr>
<td>i1</td>
<td>2.3920</td>
<td>0.1040</td>
<td>1864</td>
<td>2764.8600</td>
<td>0.0600</td>
</tr>
<tr>
<td>i1</td>
<td>2.4210</td>
<td>0.1220</td>
<td>2396</td>
<td>536.7870</td>
<td>0.3100</td>
</tr>
<tr>
<td>i1</td>
<td>2.4260</td>
<td>0.1920</td>
<td>922</td>
<td>4341.9580</td>
<td>0.0500</td>
</tr>
<tr>
<td>i1</td>
<td>2.4670</td>
<td>0.0990</td>
<td>1049</td>
<td>2420.7700</td>
<td>0.2500</td>
</tr>
<tr>
<td>i1</td>
<td>2.5540</td>
<td>0.1280</td>
<td>8650</td>
<td>199.5300</td>
<td>0.3900</td>
</tr>
<tr>
<td>i1</td>
<td>2.5720</td>
<td>0.1330</td>
<td>1404</td>
<td>570.6300</td>
<td>0.0700</td>
</tr>
<tr>
<td>i1</td>
<td>2.5720</td>
<td>0.1270</td>
<td>733</td>
<td>948.5660</td>
<td>0.8700</td>
</tr>
<tr>
<td>i1</td>
<td>2.6060</td>
<td>0.1510</td>
<td>2192</td>
<td>354.0030</td>
<td>0.1100</td>
</tr>
<tr>
<td>i1</td>
<td>2.7860</td>
<td>0.3200</td>
<td>5518</td>
<td>161.9270</td>
<td>0.9800</td>
</tr>
<tr>
<td>i1</td>
<td>2.8210</td>
<td>0.3020</td>
<td>3594</td>
<td>478.6360</td>
<td>0.7200</td>
</tr>
<tr>
<td>i1</td>
<td>2.8210</td>
<td>0.2730</td>
<td>2747</td>
<td>630.6000</td>
<td>0.1400</td>
</tr>
<tr>
<td>i1</td>
<td>2.8210</td>
<td>0.1920</td>
<td>2606</td>
<td>797.8840</td>
<td>0.4200</td>
</tr>
<tr>
<td>i1</td>
<td>2.8270</td>
<td>0.3660</td>
<td>3952</td>
<td>344.6370</td>
<td>0.2900</td>
</tr>
<tr>
<td>i1</td>
<td>2.8330</td>
<td>0.1910</td>
<td>2094</td>
<td>951.0280</td>
<td>0.9400</td>
</tr>
<tr>
<td>i1</td>
<td>2.8390</td>
<td>0.2200</td>
<td>1362</td>
<td>1102.8870</td>
<td>0.9900</td>
</tr>
<tr>
<td>i1</td>
<td>2.8910</td>
<td>0.1740</td>
<td>1433</td>
<td>1246.1100</td>
<td>0.0300</td>
</tr>
<tr>
<td>i1</td>
<td>2.9200</td>
<td>0.1100</td>
<td>1560</td>
<td>1395.0790</td>
<td>0.2500</td>
</tr>
</tbody>
</table>

iftableamp = p4
iftablefreq = p5
ipointsiz = p6
ispeed = p7 * (1/p3)
ifuncsize = ftlen(iftableamp)

kphase phasor ispeed
kphase = kphase * ipointsiz / ifuncsize
kamp tablei kphase, iftableamp, 1
kfreq tablei kphase, iftablefreq, 1

apartials loscil3 giglobalamp * kamp, kfreq, 1, 82.407
aleft = apartials * (1-p8)
aright = apartials * p8
outs aleft, aright
endin

; SCORE EXAMPLE (non-averaged case, using samples instead of sine wave; the score is actually
; many pages longer than this; we abbreviated it just for the sake of this example)

f 1 0 0 1 "StringLow.aif" 0 0 0
f 2 0 64 -23 "amp_data1.txt"
f 3 0 64 -23 "freq_data1.txt"
f 4 0 32 -23 "amp_data2.txt"
f 5 0 32 -23 "freq_data2.txt"
f 6 0 32 -23 "amp_data3.txt"
f 7 0 32 -23 "freq_data3.txt"
f 8 0 64 -23 "amp_data4.txt"
f 9 0 64 -23 "freq_data4.txt"
f 10 0 32 -23 "amp_data5.txt"
f 11 0 32 -23 "freq_data5.txt"
(...)
f 1372 0 32 -23 "amp_data686.txt"
f 1373 0 32 -23 "freq_data686.txt"
f 1374 0 32 -23 "amp_data687.txt"
f 1375 0 32 -23 "freq_data687.txt"
f 1376 0 32 -23 "amp_data687.txt"
f 1377 0 32 -23 "freq_data687.txt"

i 2 0 5
i 1 0.023 0.256 2 3 45 1.00 0.50
i 1 0.041 0.180 4 5 32 1.00 0.50
i 1 0.041 0.104 6 7 19 1.00 0.50
i 1 0.041 0.220 8 9 38 1.00 0.50
i 1 0.046 0.134 10 11 24 1.00 0.50
(...)
i 1 4.975 0.122 1372 1373 22 1.00 0.50
i 1 5.004 0.122 1374 1375 22 1.00 0.50
APPENDIX D (l’Attente, by Roland Barthes)

The book *Fragments d’un discours amoureux*, by Roland Barthes, was first published in 1977 (Éditions du Seuil, coll. “Tel Quel”). *L’Attente* (“Waiting”) is one of the chapters. We have used an adapted English version found on the web at

http://family.knick.net/thecastle/barthes.htm

We have also compared this English version with the original in French (Barthes, 1995), making slight changes when needed. The text is not used in its entirety in *The Spell of Speech*. We built a main story line based on the linear flow of the text as presented in the original, with a few breaks and omissions. The last paragraph, the “mandarin story”, gained a prominent role throughout *The Spell of Speech*, functioning as a second, separate story line that remains unconcluded till the very end of the piece. Below we present the complete English version of the text. The pieces of it that were actually used in our composition are fully transcribed in the score (see Appendix I).

**Waiting**

1. I am waiting for an arrival, a return, a promised sign. This can be futile, or immensely pathetic; in *Erwartung* (Waiting), a woman waits for her lover, at night, in the forest; I am waiting for no more than a telephone call, but the anxiety is the same. Everything is solemn; I have no sense of proportions.
2. There is a scenography of waiting: I organize it, manipulate it, cut out a portion of time in which I shall mime the loss of the loved object and provoke all the effects of a minor mourning. This is then acted out as a play.

The setting represents the interior of a café; we have a rendezvous, I am waiting. In the Prologue, the sole actor of the play (and with reason), I discern and indicate the other's delay; this delay is as yet only a mathematical, computable entity (I look at my watch several times); the Prologue ends with a brainstorm: I decide to “take it badly”, I release the anxiety of waiting. Act I now begins; it is occupied by suppositions: was there a misunderstanding as to the time, the place? I try to recall the moment when the rendezvous was made, the details which were supplied. What is to be done (anxiety of behavior)? Try another café? Telephone? But what if the other comes during these absences? Not seeing me, the other might leave, etc. Act II is the act of anger; I address violent reproaches to the absent one: “All the same, he (she) could have…”, “He (she) knows perfectly well…” Oh! if she (he) could be here, so that I could reproach her (him) for not being here! In Act III, I attain to (I obtain?) anxiety in the pure state: the anxiety of abandonment; I have just shifted in a second from absence to death; the other is as if dead: explosion of grief: I am internally livid. That is the play; it can be shortened by the other's arrival; if the other arrives in Act I, the greeting is calm; if the other arrives in Act II, there is a “scene”; if in Act II, there is recognition, the action of grace: I breathe deeply, like Pelléas emerging from the underground chambers and rediscovering life, the odor of roses.
(The anxiety of waiting is not continuously violent; it has its matte moments; I am waiting, and everything around my waiting is stricken with unreality: in this café, I look at the others who come in, chat, joke, read calmly: they are not waiting.)

3. Waiting is enchantment: I have received orders not to move. Waiting for a telephone call is thereby woven out of tiny unavowable interdictions to infinity: I forbid myself to leave the room, to go to the toilet, even to telephone (to keep the line from being busy); I suffer torments if someone else telephones me (for the same reason); I madden myself by the thought that at a certain (imminent) hour I shall have to leave, thereby running the risk of missing the healing call, the return of the Mother. All these diversions that solicit me are so many wasted moments for waiting, so many impurities of anxiety. For the anxiety of waiting, in its pure state, requires that I be sitting in a chair within reach of the telephone, without doing anything.

4. The being I am waiting for is not real. Like the mother's breast for the infant, “I create and re-create it over and over, starting from my capacity to love, starting from my need for it”: the other comes here where I am waiting, here where I have already created him/her. And if the other does not come, I hallucinate the other: waiting is a delirium.

The telephone again: each time it rings, I snatch up the receiver, I think it will be the loved being who is calling me (since that being should call me); a little more effort and I “recognize” the other's voice, I engage in the dialogue, to the point where I lash out furiously against the importunate outsider who wakens me from my delirium. In the café,
anyone who comes in, bearing the faintest resemblance, is thereupon, in a first impulse, *recognized*.

And, long after the amorous relation is allayed, I keep the habit of hallucinating the being I have loved: sometimes I am still in anxiety over a telephone call that is late, and no matter who is on the line, I imagine I recognize the voice I once loved: I am an amputee who still feels pain in his missing leg.

5. “Am I in love? — Yes, since I'm waiting.” The other never waits. Sometimes I want to play the part of the one who doesn't wait; I try to busy myself elsewhere, to arrive late; but I always lose at this game: whatever I do, I find myself there, with nothing to do, punctual, even ahead of time. The lover's fatal identity is precisely: *I am the one who waits*.

(In transference, one always waits--at the doctor's, the professor's, the analyst's. Further, if I am waiting at a bank window, an airport ticket counter, I immediately establish an aggressive link with the teller, the stewardess, whose indifference unmasks and irritates my subjection; so that one might say that wherever there is waiting there is transference: I depend on a presence which is shared and requires time to be bestowed—as if it were a question of lowering my desire, lessening my need. *To make some wait:* the constant prerogative of all power, “age-old pastime of humanity.”)

6. A mandarin fell in love with a courtesan. “I shall be yours”, she told him, “when you have spent a hundred nights waiting for me, sitting on a stool, in my garden, beneath
my window.” But on the ninety-ninth night, the mandarin stood up, put his stool under his arm, and went away.

by Roland Barthes
APPENDIX E (posters and program notes)

Program notes and posters designed for the concert. There were two different posters (size 11x17 inches). The same design of the posters was used for smaller postcards (presented below in original size). Program notes were printed in papers of different colors (Letter size, landscape) and folded. The text for the program notes is reproduced below.

The Spell of Speech

electro-acoustic and instrumental music
by
Bruno Ruviaro

Prelude

Fonopoemas

Interlude I

Duo #38
Fabio Oliveira, percussion
Bruno Ruviaro, piano

Interlude II

The Spell of Speech
Hannah Chodos, actress

There will be no intermission.
The pieces will be played without interruption.
Please remember to turn off your cell phones.
Prelude, Interlude I and Interlude II are short electro-acoustic pieces composed in 2004. The materials come from other pieces on this program.

Fonepoemas is an electro-acoustic piece composed in 2003. Its source material is a recording of vocal improvisations on the phonemes of the name “Tânia“. The first section is based on the /t/ and its suggested rhythms; the second part condenses and “harmonizes” the Portuguese nasal /a/ and the consonant /n/; the last section reflects the high-frequency qualities of the /i/. The final /a/, almost not pronounced, suddenly concludes the piece.

Duo #38 is a collaborative work with percussionist Fabio Oliveira. Since 1998 I have been working on contemporary music improvisation with different instrumental ensembles. Our work on improvisation has little to do with the illusion of “total freedom”; rather it seeks to connect indeterminacy and preciseness to create musical structures in real time. The collaboration with Fabio begun in 2004.

The Spell of Speech is a musical theater piece for actress, live-electronics and electro-acoustic sounds. It is the musical result of my Master’s degree research, which deals with the use of speech in music composition (or, perhaps, the use of music in speech composition). The text used is an English translation of L’attente ("Waiting"), a chapter from Roland Barthes’ Fragments d’un discours amoureux.

* * *

Bruno Ruviaro

Bruno Ruviaro was born in 1976 in São Paulo, Brazil. In 2000 he graduated in Piano and Composition at the State University of Campinas. He composes both instrumental and electro-acoustic music and also works with improvisation and collective composition. After finishing his Master’s degree in Electro-Acoustic Music at Dartmouth College, he is moving to California to continue his musical studies at Stanford University.

Fabio Oliveira was born in Brazil’s capital (Brasília) in 1980. He received his Bachelors degree in Percussion Performance from the Sao Paulo State University in 2001, and has recently completed his Master’s Degree at the University of Massachusetts. Fabio focuses on the performance of New Music in collaboration with young composers to comission and premiere new works.

Hannah Chodos was born in Santa Monica, CA, in 1984. In 2002 she graduated from the Los Angeles County High School for the Arts, where she majored in theatre for four years. This spring will complete her second year as an undergraduate at Dartmouth College.

Bruno Ruviaro’s home page:
http://eamusic.dartmouth.edu/~bruno

E-mail:
brunoruviaro@gmx.net
Many thanks to those who helped making this concert possible: Jon Appleton (advisor), Theresa Ciambra, Yuri Spitsyn, Masaki Kubo, Owen Osborn, Bill Haslett, Irina Escalante, Steve Pierce, Hannah Chodos, Fabio Oliveira, Brent Reidy, the Bregman Studio of Electro-Acoustic Music and the Music Department.

the spell of speech

electro-acoustic and instrumental music
by
Bruno Ruvioso

May 16th 2004 (Sunday)
8:00 pm — Rollins Chapel
free admission

guest artists:
Hannah Chodos, actress
Fabio Oliveira, percussionist

Graduation Concert — Master's degree in Electro-Acoustic Music at Dartmouth College

[poster and postcard design 1]
the spell of speech

new music
by
Bruno Ruviaro

May 16th 2004 (Sunday)
8 pm — Rollins Chapel
free admission

guest artists:
Hannah Chodos, actress
Fabio Oliveira, percussionist

graduation concert — Master’s degree in Electro-Acoustic Music at Dartmouth College

[poster & postcard design 2]
APPENDIX F (pictures)

The pictures below were taken during the first rehearsals in Faulkner Auditorium (this is a classroom; it is not the actual concert space) at Dartmouth College, March/April 2004. Actress Hannah Chodos wears an AudioTechnica wireless microphone to have her voice amplified and sent to a Macintosh iBook G4 running Max/MSP through a MOTU 896 Firewire interface. Most of the theater part was composed during these rehearsals. Refer to the enclosed DVD to see the final result including costume and stage settings.
APPENDIX G (CD and DVD)

CD 1:

<table>
<thead>
<tr>
<th>TRACK</th>
<th>PIECE</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>The Spell of Speech</em></td>
<td>Dry recording: direct out from the computer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May 16\textsuperscript{th} 2004 (concert)</td>
</tr>
<tr>
<td>2</td>
<td><em>The Spell of Speech</em></td>
<td>Recording with room ambience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May 16\textsuperscript{th} 2004 (concert)</td>
</tr>
<tr>
<td>3</td>
<td><em>The Spell of Speech</em></td>
<td>Recording with room ambience</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May 17\textsuperscript{th} 2004 (extra take)</td>
</tr>
</tbody>
</table>

CD 2:

<table>
<thead>
<tr>
<th>TRACK</th>
<th>PIECE</th>
<th>COMMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[soundscape] (2004)</td>
<td>Sparse sounds of birds and squirrels played</td>
</tr>
<tr>
<td></td>
<td></td>
<td>during the entrance of the public</td>
</tr>
<tr>
<td>5</td>
<td><em>Duo #38</em> (2004)</td>
<td>Bruno Ruviaro, piano</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fabio Oliveira, percussion</td>
</tr>
</tbody>
</table>

DVD:

1. *The Spell of Speech* (May 17\textsuperscript{th} 2004)

2. *The Spell of Speech* (May 16\textsuperscript{th} 2004): Concert recording. Due to technical problems in the camera during the concert, the last two minutes of the piece were not recorded.

3. *Duo #38* (May 16\textsuperscript{th} 2004): piano and percussion improvisation; concert recording.
REFERENCES


Eimert, Herbert (1963?). *Epitaph für Aikichi Kuboyama; Komposition für Sprecher und Sprachklänge. 6 Studien; elektronische Musik* (LP and booklet), WERGO 60014. [we couldn’t find the exact year of release of this LP, but it was certainly after 1963].


