ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

MUSIC OF SPEECH: TRANSFERRING SPOKEN DIALOGUE TO MUSIC

M.A. THESIS

E. Haydar Cengiz

Department of Music

Master Program In Music

September 2018

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF ARTS AND SOCIAL SCIENCES

MUSIC OF SPEECH: TRANSFERRING SPOKEN DIALOGUE TO MUSIC

M.A. THESIS

E. Haydar Cengiz
(409131101)

Department of Music

Master Program in Music

Thesis Advisor: Assoc. Prof Jerfi AJI
Thesis Co-Advisor: Dr. Jane HARRISON

September 2018

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ SOSYAL BİLİMLER ENSTİTÜSÜ

KONUŞMANIN MÜZİĞİ: KONUŞMA DİYALOĞUNUN MÜZİĞE TRANSFERİ

YÜKSEK LİSANS TEZİ

E. Haydar CENGIZ
(409131101)

Müzik Anabilim Dalı

Müzik Yüksek Lisans Programı

Tez Danışmanı: Doç. Dr. Jerfi AJI
Eş Danışman: Dr. Jane HARRISON

Eylül 2018

E. Haydar CENGIZ, an M.A. student of ITU Graduate School of Arts and Social Sciences student ID 409131101 successfully defended the final project entitled "MUSIC OF SPEECH: TRANSFERRING SPOKEN DIALOGUE TO MUSIC ", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Dr. Jerfi AJI** Istanbul Technical University	.....
Co-advisor :	**Dr. Jane HARRISON** Istanbul Technical University	.....
Jury Members :	**Assoc. Prof. Dr. Ozan BAYSAL** Istanbul Technical University	.....
	Assoc. Prof. Dr. Robert REIGLE	.....
	Prof. Dr. Ertan YURDAKOŞ Altınbaş University	.....

Date of Submission : 4 May 2018
Date of Defense : 6 September 2018

To my wife,

FOREWORD

Dear readers,

I would like to thank:

Jerfi Aji, Jane Harrison, Ozan Baysal and Robert Reigle for their guidance,
Laçin Şahin for his infinite support in every project that I do,
Amy Salsgiver, Emin Fındıkoğlu and Alexandros Charkiolakis for being great music teachers.

Lastly, I would like to thank my wife for being lovely and supportive.

September 2018 E. Haydar CENGIZ

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

MUSIC OF SPEECH: TRANSFERRING SPOKEN DIALOGUE TO MUSIC
SUMMARY

This master's thesis focuses on oral conversation. *Music of Speech* aims to discuss emotional and interactive aspects of oral interchange through musical terminology, based on the suggestions that speaking is the most frequently used improvisational instrument among humans and each dialogue is a formed-piece in which introduction, development, and conclusion are executed, based on selected topic(s). I suggest that analyzing dialogues in such way could be useful for composers, theorists and musicians in order to bring new perspectives to their musical ideas.

The main part of the project is based on a sound recording experiment. In this experiment, I perform a casual conversation with Neslihan Aker, a professional actress. This conversation is recorded with a single microphone. The analyzed part is selected from a whole conversation, which is over 50 minutes. Selected dialogue is transcribed into musical notation and observed not only through theoretical and rhythmic perspectives, but also in terms of compositional qualities.

KONUŞMANIN MÜZİĞİ: DİYALOĞUN MÜZİĞE DÖNÜŞTÜRÜLMESİ

ÖZET

Bu yüksek lisans tezi sözlü diyaloğa odaklanmaktadır. *Konuşmanın Müziği*, konuşma eyleminin insanlar arasında en çok kullanılan doğaçlama enstrüman oluşu ve her konuşmanın, seçilen belirli bir konu/konulara dair girişi, gelişmesi ve sonucu olan birer eser olduğu önermelerini temel alarak, karşılıklı sözlü değişimin duygusal ve interaktif yönlerini tartışmaktadır. Diyalogları bu şekilde analiz etmenin, müzikal fikirlerine yeni bir bakış açısı getirmesi açısından bestecilere, teorisyenlere ve müzisyenlere faydalı olabileceği kanaatindeyim.

Tezin ana bölümü bir ses kaydı deneyi üzerinde temellendirilmiştir. Bu deneyde ben ve aktris arkadaşım Neslihan Aker gündelik bir konuşma icra etmekteyiz. Bu konuşma tek bir mikrofonla kaydedilmiştir. Seçilen sohbet bölümü, tümü 50 dakikayı aşan bir bütün konuşmadan seçilmiştir. Seçilen diyalog müzik notasına transkripsiyonu yapılmış ve sadece ritmik ve teorik açıdan değil, kompozisyonel kaliteleri açısından da incelenmiştir.

1. SPEECH MUSIC

1.1 Purpose of This Research on "Music of Speech"

In this thesis, I propose some new methods for analyzing speech as music.

In her book *The Music of Everyday Speech*, Ann Wennerstrom indicates that the more she worked with the discourse, the more it has become clear that prosody- intonation, timing, and volume- is central to the interpretation of spoken language, but that, unfortunately, it is often ignored in actual analysis of discourse (Wennerstrom,2001,p.7)

Therefore, *Music of Speech* aims to discuss emotional, interactive and prosodic aspects of oral interchange through a musical terminology, based on the suggestion that speaking is the most frequently used improvisational instrument among humans. In *Music of Speech*, no different than conversation analysis, a sudden act of interactive speech is taken as a "dramatic dialogue"; a piece of spoken text that is intended to be performed and listened to. Instead of being re-performed by speech, the piece of dialogue is reproduced by musical instrument(s). During this research, a random dialogue with a close friend of mine was recorded as the main entity. A small portion of this conversation was chosen and taken in the context of a dramatic dialogue in order to be observed through the musical terminology. Spontaneity of the dialogue is crucial in order to stay close to the act of improvisation.

There are several reasons for choosing spontaneous speech. One of them is the use of concrete symbols (words), whereas sounds (notes) and their combinations could be semantically assigned more relatively and broadly. Since we have a solid dialogue with words, topic(s) and their evolution in the dialogue is pre-determined. Another reason is the uncertainty of the musical creation process. From *Ursprung* (musical work's origin) to "Ideal object" (the entity that can be passed on to others with the help of memory, speech and writing inscription) of Edmund Husserl (Benson, 2003, p. 33),

the path of musical creation is a unique and chaotic one, including composing and improvising which are mostly and unconsciously interwoven by the "music creator". There are various musical phenomenology theories about the "creation process". By using spontaneous dialogue as a frame, we had the chance to observe an "improvisational composition". Christian Fuchs asserts that "communication is a reciprocal process between at least two humans, in which symbols are exchanged and all interaction partners give meaning to these symbols" (Fuchs, 2017, p. 6).

The chosen dialogue will be analyzed as a formed-piece in which some sort of introduction, development and conclusion are executed, based on selected topic(s) in a literary context similar to a musical context. I suggest that analyzing dialogues in such a way could be useful for composers, theorists and musicians in order to bring new perspectives to their musical ideas. What would happen if the improvised speech would be imitated with musical instruments? How would instant human information trafficking on a specific topic sound like in musical context? What would be remarkable rhythmic and melodic features? This thesis aims to find answers to those questions.

1.2 Overview of Chapters

This thesis aims to introduce not only its own ideas and evolution, but also several multidisciplinary approaches to and speech and it's music to clarify it's own path. At first, in order to reflect the perspective through the topic, the purpose of the thesis, the importance of information traffic and interaction will be stated. Then, an overview on concepts of language and music will be explained. Nowadays, technology plays a crucial role in our communication routines. Therefore, technological developments are stated in order to give some current and future examples that have direct and dramatic effects. The second chapter focuses on music of speech through an interdisciplinary and intercultural perspective. The first section of the second chapter cites research made by linguists, musicologists, ethnomusicologists and psychologists from the 18th century to modern day. The second section gives some contemporary musical examples. In the third chapter, the methodology is given followed by an analysis. The

fourth section focuses on the analysis and musical transcription of the recorded dialogue. The detailed harmonic, rhythmic and formal analysis of the dramatic passages chosen through the dialogue are cited. Finally the fourth chapter presents a conclusion and further possibilities based on the analysis

1.3 General Theoretical Background of Speech and Language

1.3.1 Information and Information Traffic

According to the scientific approach, a living organism usually has two principal elements: A set of instructions that tells the system how to sustain and reproduce itself, and a mechanism to carry out the instructions (http://www.hawking.org.uk/life-in-the-universe.html). In order to sustain and reproduce as mentioned above, sets of instructions have developed since the very beginning of the universe and were transferred to younger generations of living organisms with DNA-or arguably RNA in the early ages of evolution. Therefore, even in this process, one can mention two terms, communication and interaction, when we talk about carrying out the instructions. These two phenomena are both used during the process since not only is there a definite act of information sharing, but also this sharing of information affects others in order to sustain life and reproduction. From single cells to whole species, all beings use different types of communication systems.

In a habitat like our planet where many different types of species survive, there is a huge amount of useful data systems. But as far as we could observe, only the human-species is able to write books about emotion and information to use them as knowledge and wisdom. At this point, human's capacity to communicate in such complex ways takes a crucial role. The closest degree that animals have come to language has been in studies in which pygmy chimpanzees learn a simple vocabulary and syntax based on interactions with humans (Patel, 2010, p. 349). These chimpanzees are able to learn a few hundred words which is far less than an ordinary human child, who is capable of learning thousands of words with complex grammatical structures. After all, until now this great tool called "language" has been discovered only by Homo sapiens, through means of symbolic and linguistic creativity.

1.3.2 Theories of Communication

Communication is studied from several different scientific perspectives which makes it some sort of "functional inclusion field". In 1999, Robert T. Craig summarized different theoretical strands in communication scholarship into seven "traditions:" rhetorical, semiotic, phenomenological, cybernetic, sociopsychological, sociocultural and critical (Schultz and Cobley,2013, p.31). While making such a classification, he also attracts attention to the "disciplinary health" of communication scholarship.

On the other hand, Schultz and Cobley's deduction asserts that "there may be no such thing as COMMUNICATION THEORY, but there be many communication theories, each proceeding from a different understanding of communication phenomena and each contributing to scholarship from that understanding" (Schultz and Cobley,2013, p.32).

1.3.3 Evolutionary Origins of Language

When and how did humankind differentiate from their ancestors? The answer lies in a complex sociologic collaboration process. As anybody would guess, chimpanzee group hunting is a clear example of ape cooperation toward a shared goal, in which one could search for common grounds with our ancestors' cooperative and collaborative attitudes. Psychologist and linguist Michael Tomasello thinks that such comparison puts too much of a human face on chimpanzees. He asserts that: "apes understand the intentions of other apes, but they seem to react to these competitively rather than cooperatively, unable to enter into an intentionality shared and sustained among the group" (Tomlinson, 2015, p. 97).

On the other hand, archeological records suggest a different picture of the family Hominidae of Lower Paleolithic society. Their ability to see a broader picture leans them towards immediate cooperation with those around them, division of labor along the members of the group and finally attaining their goals. The ability to cooperate gives greater rewards (Tomlinson, 2015, p. 98).

Cognitive scientists aim to ask a critical question in order to illuminate this creative process: Have human bodies and brains been shaped by natural selection for language? Two major views stand on the scene. The first view is an evolutionary perspective from "language adaptationists" like Pinker and Jackendoff (2005), arguing that language acquisition involves a set of cognitive and neural specializations tailored by evolution to develop such rich communicative systems. A second view claims that

selection acted to create certain unique social-cognitive learning abilities in humans, such as "shared intentionality" to construct languages. Briefly, as Patel remarks, "The debate between language adaptationists and language constructivists keeps on giving us new arguments if we are evolved to create language or not" (Patel, 2010, p.359).

The power of speech is such that it is often considered nearly synonymous with being human (Zatorre and Gandour, 2007, p. 1). Therefore, over the last century, there is a tremendous effort in the scientific society to research such a mechanism. So far, there are two main ideas trying to explain speech and pitch processing. First of them is the domain-specific model, originated in 1905s by Alvin Liberman with the help of spectrograph. Basically, the domain-specific model proposes that speech perception and production depends on specialized mechanisms that are dedicated exclusively to speech processing. This particular theory led to motor theory of speech perception, proposing that speech bypasses the normal pathway for sound analyzing and passes through a system exclusively dedicated to speech (Zatorre and Gandour, 2007, p. 2).

The second intellectual trend is the cue-specific model, claiming that speech is perceived by the same system that other auditory functions are processed. Both models have experimental findings for their argumentation. However, there is still a long way to see the whole process.

Any system of communication could be called as language. Watson (2015) asserts that language "relies on a suite of cognitive capacities that make it far more complex than any other communication system. One notable capacity is our ability to label external objects and events with acoustically distinct, referential words" (p. 495).

Several different hypotheses of language origin exist, which developed in time with the help of linguistic science. From early speculations gathered by Max Müller, those hypotheses try to enlighten the past as much as they can, but regarding the fact that the "unrecorded" past is just "mystery", the history and origin of languages remain as controversial matters. Human kind created countless languages, some of which are completely lost, or have no written literature and no conventional orthography such as many of the American Indian languages, some other which are only known by writings like manuscripts such as Latin, Ancient Greek, Assyrian and Ancient Egyptian and some, which are still written and spoken such as English, French, German and Turkish.

One should remember that this particular tool developed to "label in order to classify", provided grounds for almost every accomplishment in our short history of existence; arts, sciences, medical sciences, political systems, economics etc. However, it would be useful to keep in mind that the main focus remains the same; information and its traffic.

1.3.4 Evolutionary Theories About Music and Language

On the other hand, a human's palette of interaction capabilities is not only limited with labelling concrete objects and/or symbolizing them. However, at the very beginning of the human journey, probably there was only voice to express. The utterances of human voice gave vent to basic emotions such as veils, laughs, howls, and cries before the first word (or label) was decided. Darwin claimed that human's communicative abilities could be based on some communication intermediate between modern language and music. Therefore, it would not be wrong to say that the evolution of music was heading simultaneously with the evolution of languages, in order to express emotions such as sorrow, joy or taking a crucial role in praying and healing sessions. Our need to express our emotions and imitating nature positioned music as crucial. We know that people attend to a wide range of feelings as a source of information. Emotions, taking a role as a filter, manipulate and affect our perception of reality.

However, when it comes to music, humans are not the only species who can produce it. Patel remarks that songbirds and certain whales are notable singers, and in some species, such as European nightingale, an individual singer can have hundreds of songs involving recombination of discrete elements in different sequences. Furthermore, songbirds and singing whales are not born knowing their songs, but like humans, learn by listening to adults (Patel,2010, p.355). The difference between animals and humans in music-making comes from diversity of meanings. Most of the animal songs are sung in a limited variety, especially for territorial warnings, flirting and social status, sung mostly by males. On the other hand, humans –due to complex socio-cultural structures, tend to produce music on a much broader range of issues.

1.4 Impact of Technological Advances on Human Communication

Today with all the systematic knowledge inherited from their ancestors, humankind created a civilization based on complex law, economic, scientific and socio-cultural systems. This complex web gives constant birth to new discoveries and solutions to illuminate the unknown. Especially in the last three centuries, with historical events such as the industrial revolution, two world wars and the cold war, technological improvements have accelerated with immense pace. We are able to visit outer space, observe from sub-atomic particles to distant stars with various tools, cure diseases, examine DNA and much more. In this context, it's useful to mention some vital technological improvements and their influence on the subject of the thesis.

In order to observe technological improvements that are crucial in our subject, one should begin with the process of sound recording. Sound recording technologies have come a long way since Thomas Edison's phonograph (Şahin ,2015, p.21). While stereo imaging of sound enables us to hear recorded sounds in detail, "synthesizing" as a part of computer technologies, allows us to reproduce analog sounds with minimum error or the opposite. Those improvements gave great options to artists to express themselves in wider perspective and scientists to observe sounds since the beginning of the 20th century, as seen in Figure 1.1. Manipulation of recorded sounds made a critical effect on musical creativity. Since any desired sound could be recorded and stored, music that was kept only written until the late 1800's was also recorded. This particular achievement caused the greatest expansion of music since the invention of Western musical notation system.

Recording and capturing technologies have another revolutionary effect on humanity. One should be aware that in the last century, humanity acquired the tool that could capture and record pretty much anything that one desires. Even nowadays, anyone could keep desired amount of time in audio, visual or audio-visual format. One could easily observe that younger generations born in the beginning of the 21st century adopted those technologies so easily. That is a crucial triumph for mankind for two reasons: first reason is that in a sense, it is his first real chance to store and manipulate time –only the present becoming the past-, even though that is physically not possible with our physics knowledge. The other reason is also a time-related but anatomical problem: Human, accessing a small part of his brain is constantly challenging with his own memory function which is pretty narrow and selective.

Figure 1.1 : Frances Desnmore recording Blackfoot chief on a cylinder phonograph for the Bureau of American Ethnology (1916).

Only 30 years passed since neurology used positron emission tomography (PET) for the first time, and only 15 years since the first use of functional magnetic resonance imaging (fMRI). With the help of those functional imaging studies giving immense pace to scientists, new theoretical and empirical studies will help them to see more.

As expected, brain activities while music making and speaking are another curiosity. Same processes mentioned above are used in order to analyze differences and similarities in speech and music making. Fortunately, in recent years academic branches interested in human-speech and music-making activities have started collaborating on multi-disciplinary platforms, which could accelerate scientific discoveries on music and language related areas.

At this point, one would find useful what kind of scientific improvements technology caused in terms of speech and music making. Recent acoustical analyses suggest that the probability distribution of the amplitudes of harmonics present in human speech can be used to predict the structures of musical scales, in terms of pitch intervals that are most commonly used across cultures (Zatorre and Baum ,2012, p. 1).

Information trafficking requires the storage of necessary information in order to pass it through, as in the example of DNA, where the genetic information is perfectly stored. Therefore, through the history, in order to store, calculate and analyze big chunks of information, humanity developed computing systems and internet.

Internet is one of the major inventions of the 20th century. Apart from being the largest library that easily gives access to enormous data shared widespread, it is also the biggest communication platform that revolutionized our ways to interact live with long distances. However, with the current development of related technologies such as virtual reality or even augmented reality, people tend to spend much more time on interactive games, platforms like social media and instant messaging. Regarding the fact that the "invisible cloud" becomes one of our primary communication and informative tools, one would speculate that human's communication routines are evolving. Such a tool to shape our information dynamics deserves our attention in order to understand the future.

2. SPEECH MUSIC

Even though they share some similarities, speech and music also have lot of differences. One of the main differences, as mentioned in the previous chapter, is that speech contains a set of complex modulation called prosody, in which pitch variations take part. On the other hand, pitch variations in musical melodies are discrete, using limited set of tones. Therefore, being "out of tune" does not really apply to speech. In other words, the former focuses on contour and the latter focuses on pitch encoding and production.

Speech music looks into speech from a musical perspective.

2.1 Different Speech-Music Perspectives

In this section, different approaches about music and speech will be investigated. As mentioned in previous chapters about communication theory and discourse analysis, music scholars, composers and other scholars examine the intersection between speech and music from different perspectives. Therefore, approaches and musical examples demonstrated below are from intellectually different angles. However, contributions of each might be useful in various levels, in order to shape the vision of this thesis.

2.1.1 Intercultural and Interdisciplinary Approach

This thesis takes two vital elements into consideration to distinguish their similarities and differences: Music and speech. Therefore, as also seen in previous chapters, it is inevitable to bring most of the related scientific branches to obtain an interdisciplinary

approach. However, including music as a cultural activity to communicate emotion and information, it is also inevitable to nurture an intercultural approach.

In Western European musical culture, one can observe speech-like elements in various eras and genres. From *bel canto* opera to contemporary art music and jazz, semantics and phonology[1] of speech provoked artists towards using it in their creative processes. Therefore, those elements have been studied by many scientists over time. In fact, one of the earliest and most pioneering studies was written in 1775, *An Essay Towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols* by Joshua Steele. In his essay, Steele asserts his motivation in the introduction of the essay;

> "I had long entertained opinions concerning the melody and rhythmus of modern languages, and particularly of the English, which made me think our theatrical recitals were capable of being accompanied with a bass, as those of ancient Greeks and Romans were, provided a method of notation contrived to mark the varying sounds in common speech" (Steele,1775, p. 176).

In his method, Steele reproduces speech with a bass viol, by marking all the chromatico-diatonic stops or frets, suitable to that bass, from the bottom to the top, as represented in Figure 2.1.

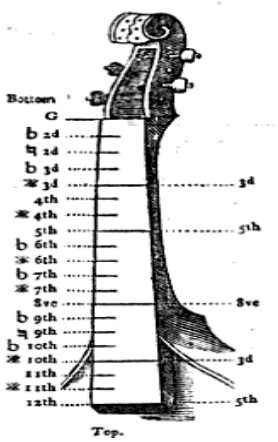

Figure 2.1 : Steele's bass viola markings for speech imitation.

Later in his essay, Steele embodies his theory with his own notation in Figure 2.2.

[1] The system of contrastive relationships among speech that constitute the fundamental components of a language.

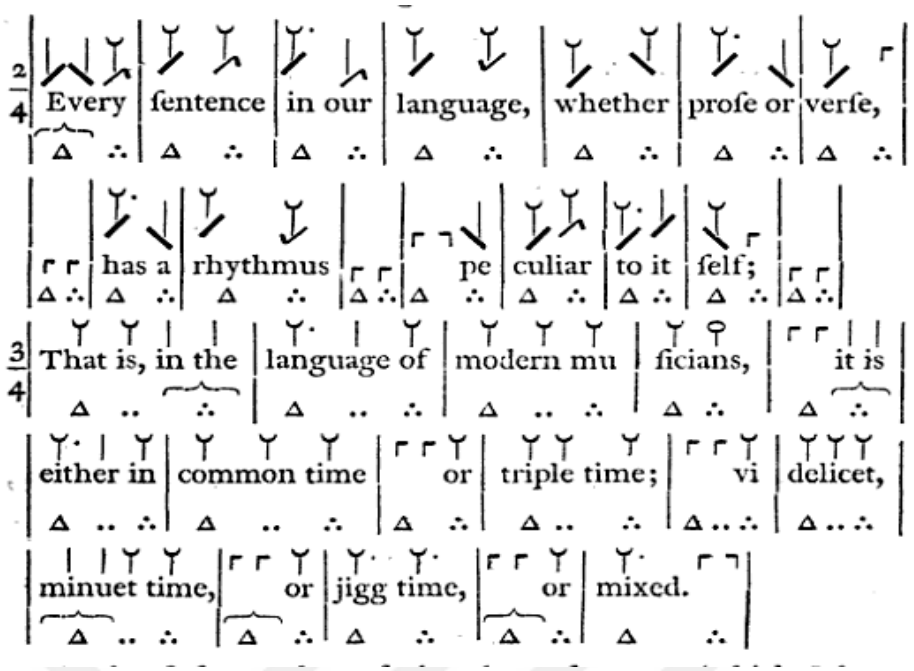

Figure 2.2 : Steele's theory with his own musical notation.

2.1.2 Speech to Song Illusion

"Speech to Song Illusion" was discovered by psychology professor Diana Deutsch in 1995. While she was fine tuning one of her spoken commentaries on her CD, she noticed that the repetition of her phrase "Sometimes behave so strangely" sounded as though sung rather than spoken. She comments that: "When you listen to this sentence in the usual way, it appears to be spoken normally – as indeed it is. However, when you play the phrase that is embedded in it: 'Sometimes behave so strangely' over and over again, a curious thing happens. At some point, instead of appearing to be spoken, the words appear to be sung, rather as in the Figure 2.3.

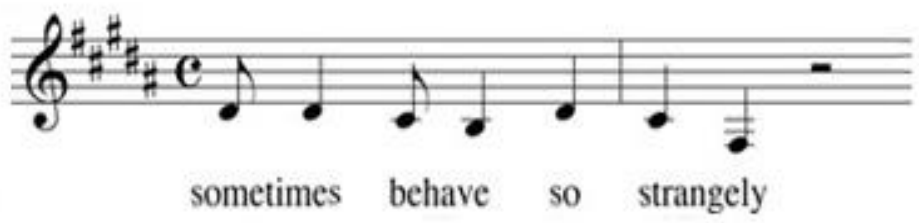

Figure 2.3 : Musical transcription of the phrase "sometimes behave so strangely" after several repetitions. (http://deutsch.ucsd.edu/psychology/pages.php?i=212)

Deutsch and her team started to investigate this effect in several experiments. At first, they tested three matched group of subjects, and presented each group with a different condition. Subjects listened to the full sentence and then ten presentations of the phrase. In each pause between different versions, subjects decided on a five-points scale whether they heard speech as if it is sung or spoken as observed in Figure 2.4. In all conditions, the first and last presentations of the phrase were identical and they examined two manipulations on subject's judgements. One manipulation is a slight transposition and the other is the presentation of the syllables in jumbled ordering. The results are demonstrated in Figure 2.5.

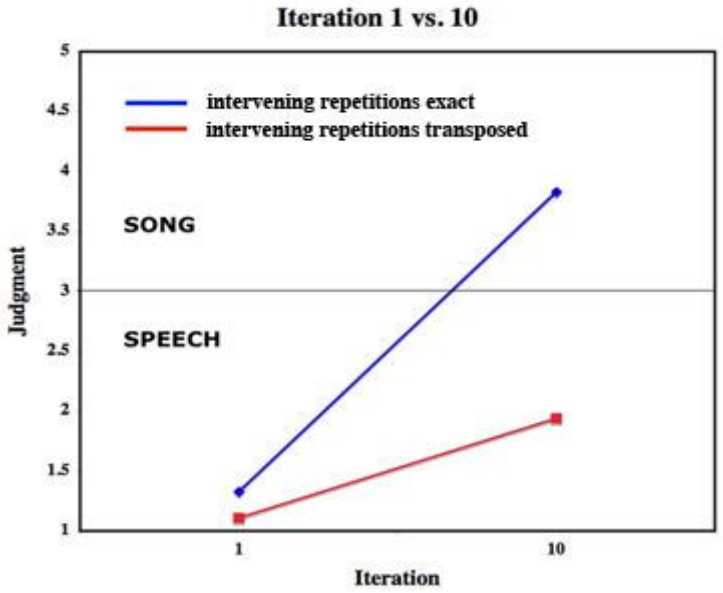

Figure 2.4 : Subject's judgements of the spoken phrase after ten repetitions. (http://deutsch.ucsd.edu/psychology/pages.php?i=212)

When the representations were exact, the phrase was heard solidly as song. When the phrase was transposed slightly on each repetition, the phrase continued to be judged as speech.

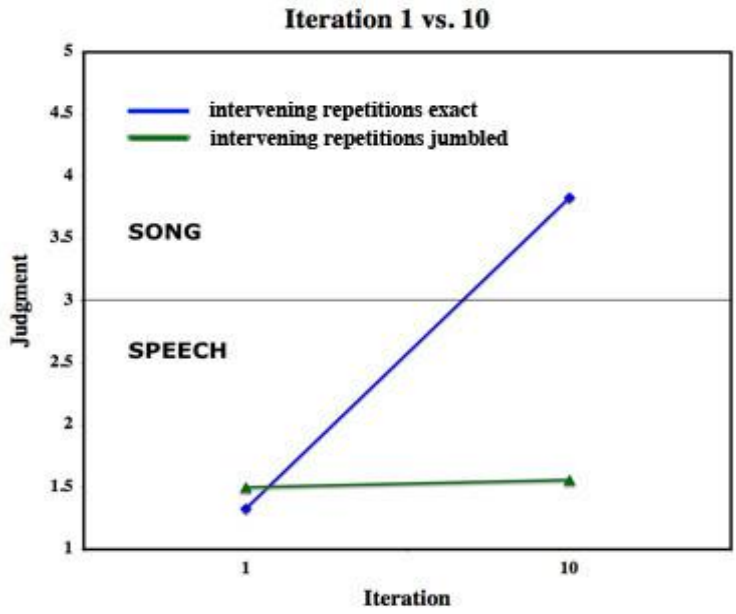

Figure 2.5 : Subject's judgements when intervening repetitions are jumbled. (http://deutsch.ucsd.edu/psychology/pages.php?i=212)

The experiment continues with the reproduction process. The team chose 11 female subjects, who had experience singing in choirs, in order to reproduce what they hear, when isolated from each other. The subjects deal with three types of presentations: they hear the spoken phrase once, they hear the sung phrase once and they hear the spoken melody ten times. As a result, it can be seen that the multiple repetition of the spoken phrase is heard as sung, seen in Figure 2.6.

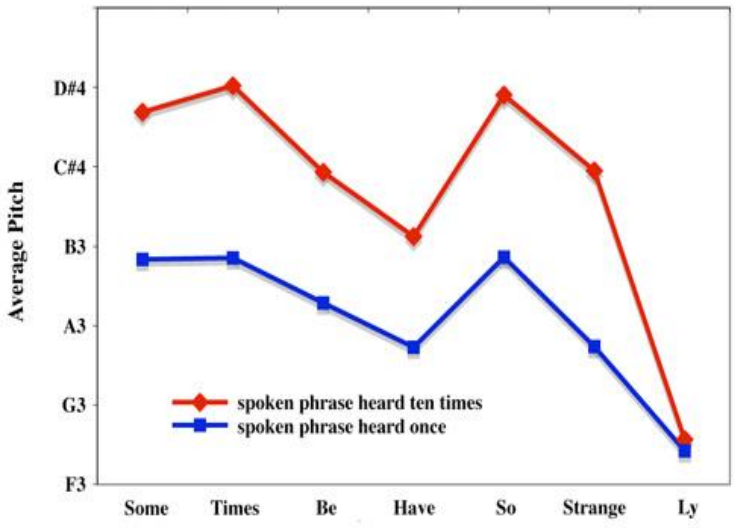

Figure 2.6 : Spoken phrase with and without repetition. (http://deutsch.ucsd.edu/psychology/pages.php?i=212)

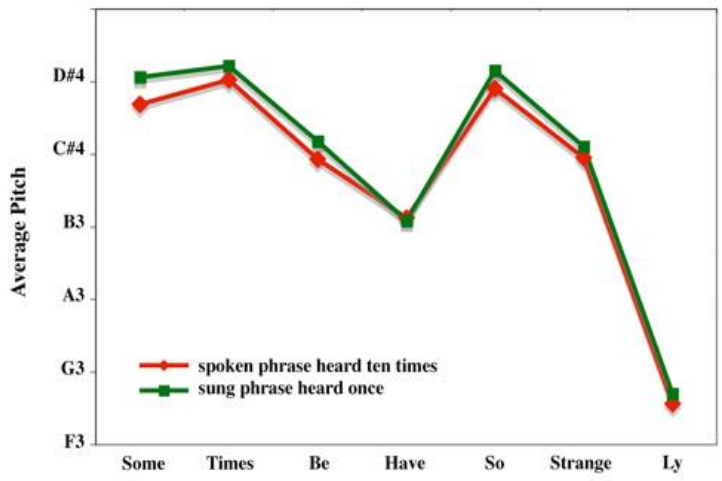

Figure 2.7 : Repeated spoken phrase and unrepeated sung phrase. (http://deutsch.ucsd.edu/psychology/pages.php?i=212)

Figures 2.6 and 2.7 demonstrate that while repetition of the spoken phrase is rerceived and reproduced by subjects as sung, the sung phrase is instantly reproduced with the same pitch relationship.

Deutsch concludes her experiments as follows:

> "…However, the present experiment shows that for a phrase to be heard as spoken or as sung, it does not need to have a set of physical properties that are unique to speech, or a different set of properties that are unique to song. Rather, we must conclude that, assuming the neural circuitries underlying speech and song are at some point distinct and separate, they can accept the same input, but process the information in different ways so as to produce different outputs. As a further point, this illusion demonstrates a striking example of very rapid and highly specific perceptual reorganization, so showing an extreme form of short term neural plasticity in the auditory system"(http://deutsch.ucsd.edu/psychology/pages.php?i=212).

On the other hand, Leon Jakobovitz's study of the repetitions of a selected phrase produces a different conclusion. He used the term "semantic satiation" for the phenomenon in which repeated words or phrases lose their semantic possessions and become meaningless chunks of sound. According to the satiation hypothesis, excessive activation of a node (mental structures in which words are represented within semantic memory) should result as fatigue, and as priming effect should be attenuating (Black, S., H., 2001, p. 494). At this point, both perspectives are useful to our research in operation of speech transcription into a musical entity. However, since the main focus

is to obtain a musical transcription of a dialogue, repetition will be used with the consciousness of avoiding semantic satiation. However, what this thesis mainly makes use of is Speech to Song Illusion.

2.1.3 Sine-wave Speech

Sine-wave speech is a form of artificially degraded speech first developed at Haskins Laboratory (https://www.mrc-cbu.cam.ac.uk/people/matt.davis/sine-wave-speech/). Basically, sine-wave speech can be described as synthesized sine waves on the center of the formant[2] frequencies of the selected utterance. The production process of sine-wave speech is demonstrated in Figure 2.8.

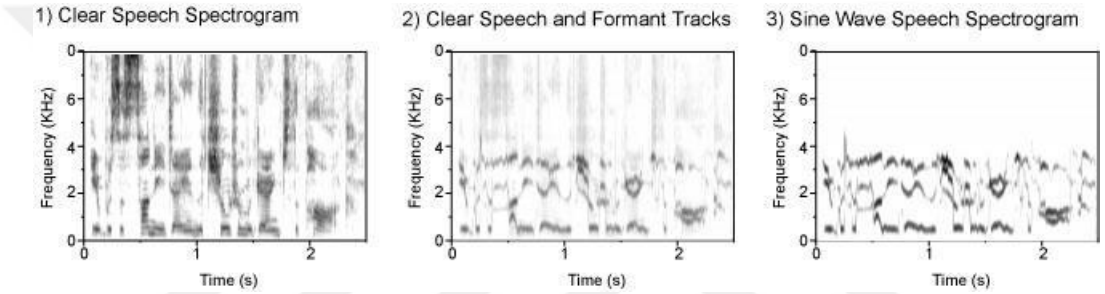

Figure 2.8 : Production of the sine-wave speech **(**https://www.mrc-cbu.cam.ac.uk/people/matt.davis/sine-wave-speech/**).**

In 1981, Remez, Rubin, Pisoni and Carell published an article titled *Speech Perception Without Speech Cues*. This article demonstrated the dramatic change of sine-wave perception of listeners, depending on their specific prior knowledge. Listeners hear no more than meaningless sounds when they first meet with the degraded speech. However, having the information about what they heard or if they were exposed to the speech itself first, listeners tend to hear sine-wave speech as a fully intelligible spoken sentence. This particular phenomenon is an example of "perceptual insight".

In the following years, in addition to sine-wave speech, other types of distorted speech have been used such as noise-vocoded speech. Experiments presented a range of evidence suggesting that perceptual grouping of speech is driven not only by primitive grouping cues, such as similarity of pitch, timbre and timing, but also by powerful

[2] Formant is a characteristic component of the quality of a speech sound; specifically: any of several resonance bands held to determine the phonetic quality of a vowel (https://www.merriam-webster.com/dictionary/formant).

experience-driven mechanisms sensitive to high-level, linguistic, characteristics of speech such as lexicality, context and expectations (Davis and Johnsrude ,2007, p. 136).

Evidences mentioned above are important to understand the way that people tend to perceive sound derived from real speech when the lexical information is removed. Sine-wave speech research may play a critical role for further investigations. This role will be argued in the next chapter.

2.1.4 Segnini and Ruviaro Analysis

In 2005, Rodrigo Segnini and Bruno Ruviaro from Stanford University published a paper called "Analysis of Electroacoustic Works with Music and Language relations". They proposed a theoretical model for music language intersections in contemporary music. They chose 15 different pieces as shown below in Table 2.1.

Composer	Piece	*
SCHOENBERG, Arnold (1874-1951)	*Pierrot Lunaire* (1912)	1v + ins
SCHWITTERS, Kurt (1887-1948)	*Ursonate* (1922-32)	1v
SCHAEFFER, Pierre (1910-95); HENRY, Pierre (1927)	*Symphonie pour homme seul* (1950)	ea
STOCKHAUSEN, Karlheinz (1928)	*Gesang der Jünglinge* (1955-56)	ea
BERIO, Luciano (1925-2003)	*Visage* (1961)	ea
EIMERT, Herbert (1897-1972)	*Epitaph für Aikichi Kuboyama* (1960-62)	ea
LIGETI, György (1923)	*Aventures and Nouvelles Aventures* (1962-66)	3v + ins
LUCIER, Alvin (1931)	*I am sitting in a room* (1970)	ea
DODGE, Charles (1942)	*Speech Songs* (1972)	ea
BERIO, Luciano (1925-2003)	*A-Ronne* (1974)	6v
APERGHIS, George (1945)	*Récitations* (1977-78)	1v
LANSKY, Paul (1944)	*Six Fantasies on a Poem by Thomas Campion* (1978-79)	ea
BODIN, Lars-Gunnar (1935)	*On Speaking Terms II* (1986)	ea
MOSS, David (1949)	*Direct Sound: Five Voices* (1989)	5v
WISHART, Trevor (1946)	*Tongues of Fire* (1994)	ea

Table 2.1: Segnini & Ruviaro's list of musical examples. [*] v = voice; "ins" = acoustical instruments; "ea" = electroacoustic (tape) pieces (https://ccrma.stanford.edu/~ruviaro/texts/Ruviaro_2005_Segnini_Analysis_Music_Language_Intersections.pdf)

Their intention was to analyze those pieces in terms of text intelligibility in an audio signal and represent the listener's perception of music-like and/or speech-like features. The conclusion is demonstrated in the Figure 2.9. below:

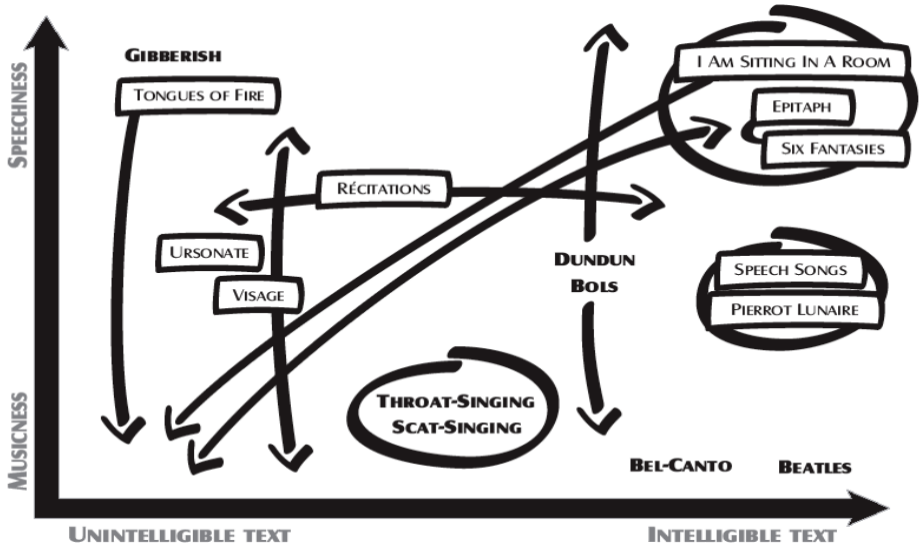

Figure 2.9 : Some pieces in S&R experiment placed within the music-language sonic space (https://ccrma.stanford.edu/~ruviaro/texts/Ruviaro_2005_Segnini_Analysis_Music_Language_Intersections.pdf).

The placement of pieces shows us their text clarity and musical qualities. Moreover, downward arrows demonstrating transitions of pieces between axis in the table are remarkable, since arrows indicate evolutions of choosen pieces through text intelligibility and musicness/speechness levels.

Alvin Lucifer's *I Am Sitting In A Room* is appropriate to clarify Segnini and Ruviaro's way of creating the table and their terminology. The piece begins as a pure speech, marked as speechness containing high intelligibility, and is placed at the top right side of the table. However, the downward arrow indicates the evolution of the piece from right top right area to bottom left area, where the structure of piece changes from pure speech to pure sound.

Kurt Schwitter's *Ursonate* is another example in Segnini and Ruviaro's table. The piece is marked as stable in the table, standing between musicness and speechness , having an unintelligible text. There is no structure changes in the piece. Those examples are important in order to illustrate the methodology of my thesis.

The paper provides us the conclusion that with the help of technologic improvements made in the last 50 years, composers tend to use more human voice and its manipulated derivatives in their composition and that dealing with the boundaries of both music and language could give useful insights for understanding the nature of perception in those domains.

2.2 Examples of Musical Pieces Using Speech as Key Element

Especially with the technological developments mentioned in the previous chapter, contemporary and jazz musicians started to explore more possibilities of the use of speech in their pieces in the 20th century.

Peter Ablinger's works reproduce speech, street noise or music from acoustic phonograph for computer-controlled player piano. He works on what he calls *phonorealism,* which consists of sound of speech transduced into musical tones via the musical automata of the player piano (Barrett, G., D., p.5). His curiousity began by wondering "what the concept of photographic realism could mean for music" (Barrett, G., D., p. 158). Barrett demonstrates Ablinger's process as follows:

> "…Ablinger compares the temporal and frequency grids with tecniques used in graphic arts in which photographs are rendered into prints. The series of pieces similarly creates "best possible fits", so to speak, between recordings and what effectively become the sonic analogue of pixels: musical tones. Ablinger explains, however, that a "genuine *phonorealism* would only be possible if the instruments had no overtones and their playing speed could be taken beyond the limit of the continous, namely16 beats per second, and if series of changing parameters could be rendered at that speed"

A Letter From Schoenberg (2006) is an installation for the *Haus am Waldsee* exhibition. The high volumed and fragmented piano piece produces intelligible speech, a letter written by Schoenberg to Mr. Ross Russell, accusing him to to have published one of his pieces without his consent. In accordance with Schoenberg's letter, his original speech produced with piano tones sounds furious.

Steve Reich's *Different Trains* uses recorded speech from interviews that he made with war victims from USA and Europe before, during and after the second world war. Speech element is used as the melody, seen in Figure 2.10, by transferring them into a

Casio FZ-1 digital sampling keyboard. His speech melody elements are also developed in *The Cave* and *City Life*.

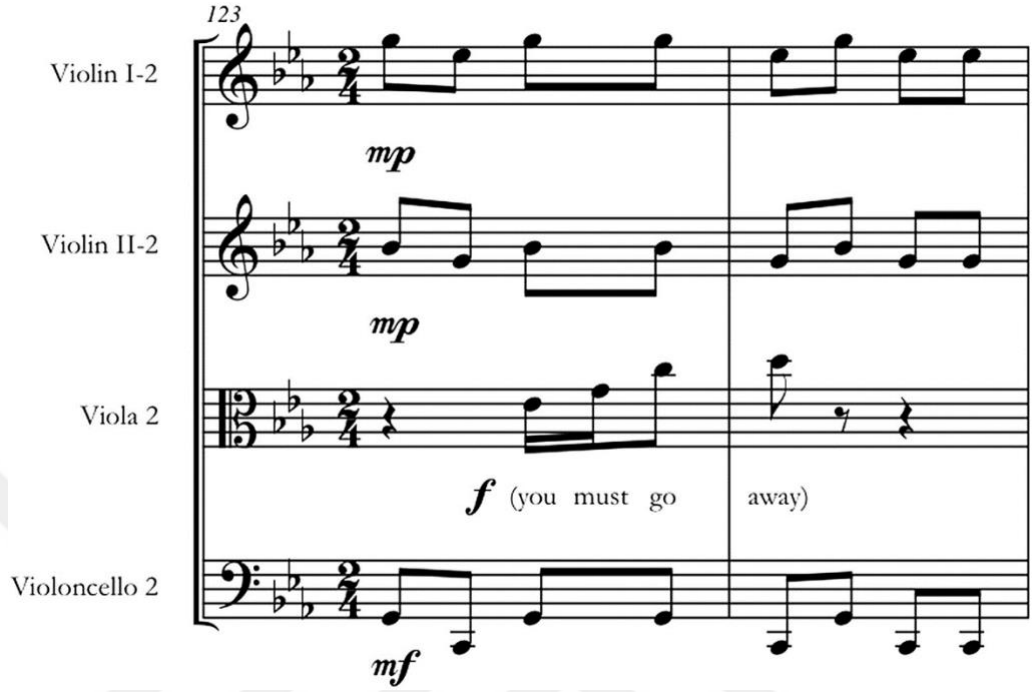

Figure 2.10 : Steve Reich, *Different Trains*, Rachel Speech melody "You must go away", mvt.2, mm. 123-124.

Maybe one of the most interesting examples is Jason Moran's *Ringing My Phone (Straight Outta Istanbul)*. Moran is a jazz pianist and composer and also the musical adviser for jazz at the Kennedy Center. The piece is based on a sampled phone conversation between two women in Turkish, although we hear one of them. Moran uses the speech in his composition as melodic, rhythmic and even thematic blueprint, not only in his recordings but also live, as seen in Figure 2.11. His trio is elaborating various parts of the sampled speech with improvisations as an ensemble and as individuals. The result is a complex, free-flowing and organic whole of improvisation tied to human speech. *Ringing My Phone* is not his only speech-used composition and Turkish is not the only language that he uses. His composition *Thief Without Loot* contains transcriptions of various phrases in Japanese.

I had the chance to interview him after his show in Istanbul and asked about his process of using speech as a musical element. His "evolution" began when he was 20, when he heard Hermeto Pascoal (mentioned below) for the first time "playing" animals,

teachers and politicians. That was certainly an evolution for him, since this particular experience made him rethink how he was listening to the language and the sound. At this point he remarks that: “In music school they teach you about music, but they don’t teach you about sound, they don’t teach you how to listen ”.

The particular journey of Moran starts when he hears the speech implementation for the first time. Then he starts to practice this concept. He points an important aspect of performing speech: “ What it does is also it challenges the technique, it does make your hands do differently. It is not like Bach or Bud Powell and that really excites me”.

Another question that I asked was about Moran’s transcription process of speech to music, in order to find similarities with Diana Deutsch’s “Speech to Song Illusion” theory mentioned above. He asserted that he listened to speech for a long time, months, sometimes all day long, in order to have a general understanding of “melody”. Then, he would go on Logic and transcribe it section by section. He quotes: “I put the section shortly you know, two seconds at a time, and when you listen to it wow it is done! I have written it out but you really have to listen to it to find where it goes”. What he means is that he repeatedly listens to some short section of two seconds, until he hears music out of it. In his case, several repetitions of small parts of speech makes them musically meaningful. Repetition is the key element while transcribing speech to melody.

Regarding the fact that Moran is not only a virtuoso player and a composer but also a music teacher, it is pretty obvious that his process of speech musicalization including internalization of the whole speech, spoken words and their gestures is a solid and aware one.

Putting speech in the center of a composition for a jazz trio which is highly improvised is hard work. The orchestration process for Moran began quiet clearly: “ I taught it to them. That was kinda hard for us for a while but we just played the song. Whether we make mistakes we played it and all of a sudden we got the gesture”.

The idea of dialogue interpretation is another angle in Moran’s music. Even though he has advanced speech-music perception and interpretaion, he for the moment, did not

get into “dialogue”. When asked, he asserted: “ I think that’s also what makes it good, you know... Every person has conversation with their mother on the phone, or every person has heard someone having a conversation and only hear one part of it. So the imagination goes to the otherside into space, like there is this negative space I think also music needs”. At this point, Moran chooses to leave the room for the music, the music that not only shapes, but also gets shaped by his main theme or “melody”, the speech. Finally, the last question was about the rhythmic observation on speech. He asserted that: “ The point is to not think there is rhythm though. I never count when I am talking to you. It actually has more freedom. It is when you understand that a phrase of music doesn’t have to be locked”.

Figure 2.11 : Jason Moran playing *Ringing My Phone (Straight Outta Istanbul)* at Jazzbaltica Festival 2004.

In addition to Moran, Robert Glasper’s *Got Over* is built on a monologue by Harry Belafonte. The piece was recorded live at Capital Studios in 2014. Belafonte’s recorded and digitally manipulated speaking voice fits with the tonal center of the piece. In addition, it has a question-response relationship with the progression played by Glasper and the rest of the trio, consisting of double bass and drums.

Another significant example is from Japanese electronic music duo Asa Chang&Junray’s *Human Without Shadows*. The piece is formed of synthesized elements and a live *tabla* performance. The interesting point is that the complex *tabla*

part is composed based on modelled human voice, whereby syllables that are recorded independently from each other form the compositional element.

Bass player Dwayne Thomas Jr. demonstrates a lot of speech imitation and re-interpretation as short video clips on Instagram. What he does is important in order to see more than just an interpretation of speech, but also the development of rhythmic and harmonic structure from spoken words. In addition, he demonstrates possibilities of instrumental performability of spoken words.

Finally, Brazilian jazz musician Hermeto Pascoal used the speech of a former Brazilian president as his foundation for his composition *Pensamenta Positivo* in his album *Festa Dos Deuses*.

To sum up, among the examples cited above, Moran's *Ringing My Phone* and Pascoal's *Pensamenta Positivo* are closest attempts to imitate impromptu/written speech with musical instruments. However, one should not forget that the examples above are purely artistic experiments and that their main purpose is not completely imitating human speech but to use speech in authentic compositions and/or performances. Furthermore, it's worth mentioning that all the examples cited in this chapter –even Moran's- are based on "one side" of the dialogue, on "monologues". Therefore, it's not possible to see an implementation of an interactive dialogue.

3. METHODOLOGY, TRANSCRIPTION AND ANALYSIS

3.1 Methodology

At this point, it is essential to mention a multidisciplinary approach which forms the foundation of this thesis. Sidnell (2016) asserts that conversation analysis is an approach to the study of social interaction and talk-in-interaction that, although rooted in the sociological study of everyday life, has exerted significant influence across the humanities and social sciences including linguistics. Apart from other linguistic sciences that take the human mind as the home of language, conversation analysis is based on the assumption that the primary environment of a language is co-present interaction. Conversation analysis seeks to discover and determine the characteristics of conversation order through not only theoretical studies but also with many recording-based observations.

Four main important domains of organization within conversation analysis are turn-taking, repair, action formation and ascription and finally, action sequencing. First two domain is important in this research. As Sacks asserts: "Turn-taking is basically the distribution of opportunities to participate in interaction. It should be organized by the participants themselves. It is locally managed, partly-administrated, interactionally controlled."(Sidnell, 2016)

Repair is an organized set of practices that participants address troubles of speaking, hearing and understanding. Repair is used when "trouble source" or "repairable" is present.

Another important study in conversation analysis apart from the domains mentioned above is paralanguage. Wennerstrom asserts that "paralanguage is the variation of pitch, volume, tempo and voice quality that a speaker makes for pragmatic, emotional and stylistic reasons and to meet the requirements of genre (Wennerstrom ,2001, p. 60). In other words, it is the way of saying things to express one's own emotional state

and its intensity level to listener(s). Therefore, paralanguage is one of the domains in conversation analysis that intersects with the main purpose of music.

However, when it comes to analyzing conversations, every scientist has different perspectives and conclusions, and even employ different symbols to explain their analyses. Therefore, one could mention subjectivity (or diverse methodology) in conversation analysis which will also be present in the musical analysis of speech in this thesis.

As an example of differences in analysis style, Figure 3.1 demonstrates a "data-internal evidence" usage between Shelley and Debbie by Jack Sidnell.

12 Shelley: district' attorneys office.
13 Debbie: Shelley:¿
14 Shelley: Debbie¿=
15 Debbie: ^what is tha dea::l.
16 Shelley: whadayou ^mean.
17 Debbie: yuh not gonna go::?
18 (0.2)
19 Shelley: well -hh now: my boss wants me to go: an: uhm finish
20 this >stupid< trial thing, u[hm

Figure 3.1: Analysis of Data-Internal Evidence (linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001/acrefore-9780199384655-e-40#acrefore-9780199384655-e-40-bibItem-0031)

A second analysis example by Ann Wennerstorm uses more symbols in order to demonstrate pitch accents and pitch extremes in Figure 3.2:

1 S see I'M on the ÓTHER END↓
2 / FÍVE / KÍDS / Í / CÁN / NÓT/ [+WÁIT+↓
3 ALL [ah ha ha ha!
4 ((extended laughter))
5 S ↱I mean it's like- (.2) +PLE:::ASE+ GOD↓ I've +DONE+
6 (.5) the DUTY↓
7 ALL ha ha ha ha ((more laughter)) ah [huh huh huh
8 S [+PULLÉASE+↓ (.5)
9 / Í / ÁM / (.2) / +DÓNE+↓/ (.3)
10 I don't wanna +É:::VER+ be +Á:::BLE+ to HAVE children
11 AGÁIN↓ I MEAN I LOVE my KIDS↗ but
12 ↳/ Í / DÓN'T / WÁNT / Á / NÓ / THÉR
13 / Bə [/ HÁY / BÉE ((“baby”))
14 A [hn hn

Figure 3.2: Analysis of an excerpt to demonstrate pitch variables.

Wennerstrom defines prosody as: “A general term encompassing intonation, rhythm, tempo, loudness, and pauses, as these interact with syntax, lexical meaning, and segmental phonology in spoken texts” (Wennerstrom, 2001,p.4). It has language specific and universal qualities. For example, an increase of pitch levels normally indicates warnings but on the other hand, Mandarin and English languages have different intonation systems. Intonation is basically the melody of the voice. Again, Wennerstrom indicates that intonation is not derived automatically from the stress patterns or syntax of an utterance; instead, a speaker decides to associate particular intonation patterns with particular constituents, depending on the discourse text (Wennerstrom, 2001,p.17). Therefore, one could mention personal choice and creativity during the use of intonation in spoken discourse.

However, looking at the examples above, one could notice the difficulty of transcribing discourse. Grammatical punctuations that are commonly used in orthography cannot sustain such a deep analysis that aims to clarify all the prosodic features of spoken interaction. Therefore, prosody has a “sideline status” in discourse analysis, since spoken discourse lacks prosodic details when converted to written form. This thesis aims to provide some additional tools found in musical language for the benefit of prosody, in order to get a more balanced perspective in conversation analysis.

In order to avoid that particular loss of information, discourse analysts create symbols for the theoretical tools that they use. One example is the use of capitals, arrows and font sizes indicating pitch accents and pitch boundaries.

Pitch accents are the various tones associated with lexical items that a speaker finds especially salient in the information structure of the discourse (Wennerstrom, A., 2001, p.18). High pitch accent indicates addition of new information to discourse, low pitch indicates addition of information that is not new or in other words already accessed. To clarify the terms mentioned above, a short example is demonstrated in Figure 3.3.

D . . . for the BÍCYCLE in (.) in the U.S. versus the BÍCYCLE in CHÍNA.

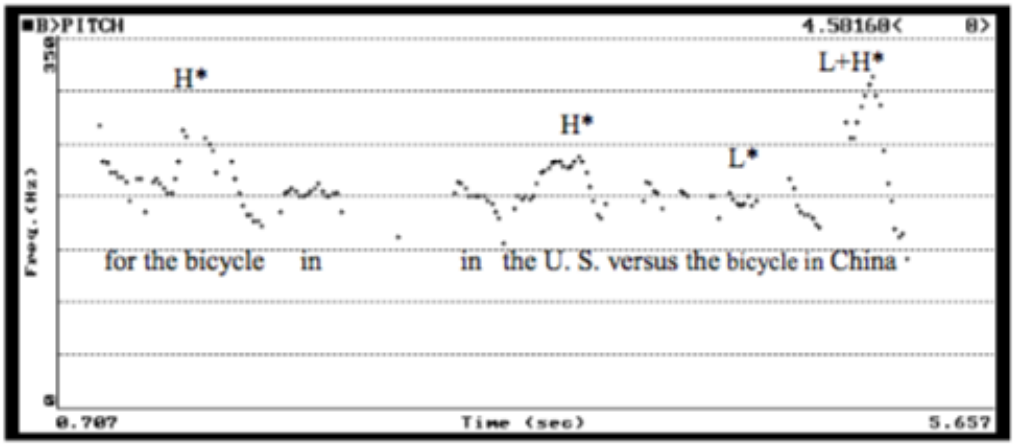

Figure 3.3: Pitch Accents Interacting with the Information Structure of the Discourse.

The table justifies the terminology mentioned above. The high-pitched word "bicycle", which provides new information is low pitched when repeated for the second time. The word "China" at the end of the sentence indicates a comparison to the prior information, "the bicycle in the U.S.".

Another term mentioned is pitch boundaries. Pitch boundaries are the pitch configurations at the end of phrases, accompanied by a lengthening of final syllables (Wennerstrom, 2001, p.18). Pitch boundary categories are: High pitch boundary, indicating anticipation from another speaker; low pitch boundary, indicating end for the current speaker; low-rising boundary, indicating anticipation of subsequent discourse for interpretation; partially falling boundary, indicating interdependency with the same speaker's utterance; and finally plateau boundary, also anticipating subsequent discourse for interpretation, but also used for hesitation. Wennerstrom uses arrow symbols to indicate such phenomena, which can be found in Figure 3.2. Especially pitch boundary examples are interesting, since they may be associated with *diminuendo* and *accelerando* terms in the musical notation system.

At this point it is useful to note that all the terms and examples mentioned in the methodology section use English as language, and it may contain several differences with different languages.

As a musician, I would want to learn if a cross-disciplinary exchange of knowledge could be made. In other words, would it be possible to create more detailed musical notation elements or concepts with the help of conversation analysis?

As a broad and multi-disciplinary study area, conversation analysis contains many more elements that cannot be covered in this thesis. In addition, doing a conversation analysis is beyond the scope of this thesis. However, some components and essential perspectives of conversation analysis will be employed in this thesis. Since our main reference point is recorded dialogue, the fundamental knowledge of turn constructional, repairable, or paralinguistic features of conversation would be as important as an analytical basis of a musical one. Discourse analysis might be useful for musicians interested in speech and dialogue in order to get deeper knowledge to visualize and to implement the structure of prosodic aspects of speech in a musical way.

A wish from Wennerstrom defines, not only her book's, but also one of this thesis *Ursprungs*; "Perhaps some day all fields involving human behavior can routinely regard the "music" of speech as a part of the foundation of communication."

The realization of the material used in this thesis required a multi-layered approach. Several techniques have been used for analysis and transcription in order to be able to move away from subjectivity. The entire path could be given in a chronological order to make visualization of the process easy.

We recorded 57 minutes of conversation that is interrupted once in the 47th minute. The continuity of the dialogue was an advantage in terms of the interactors' concentration on topics and speech itself, to put as a dramatic dialogue. During the chat, five several topics were discussed. Even though it contains outdoor sounds, it is possible to say that the recording was made in a quiet environment, in order to avoid any chaos that could obstruct our main focus, namely conversation. With the help of several listening sessions, a distinct smaller piece was extracted from the recorded whole. The distinctive aspects of the chosen piece are: semantic integrity (indivisibility of meaning), mid-level interaction between speaking sides, duration, stable topic, distinctive dynamics, and an introduction-development-conclusion progression.

As a result, an appropriate piece of 1 minutes and 25 seconds was selected.

The transcription process brought the critical problem of subjectivity within. Even trained ears would interpret the same speech totally different from each other. In order to have a "relatively" solid point of reference, first we put the dialogue in a tuning program called "Melodyne" and let it assign notes.

Melodyne is a tool developed by the company Celemony, which gives access to edit and manipulate intonation, melody, harmony, rhythm, groove, dynamics and formant of recorded sound data. Melodyne detects the tempo, scale and tuning, letting one edit musical aspects of the material as comprehensively as the notes. We configured Melodyne so that it could assign given sounds to the closest tonal pitches between semitones. The first 28 seconds of the transcription are given in Figure 3.4 below.

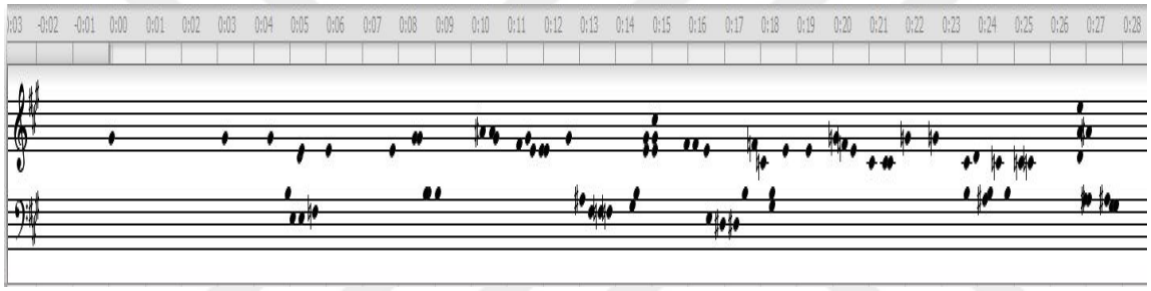

Figure 3.4: The first 28 seconds notation of speech using Melodyne.

One should notice that first of all, there is no metronome marking or any rhythmic indication. The staff is given with seconds to make the timing of the notes more precise. However, the rhythmic analysis required work on smaller scales. Every pronounced word or group of words was combined to make smaller rhythmic entities in order to come up with a more meaningful conclusion.

Another ambiguity was that Melodyne puts some of the formants of the sound separately into the notation. Even though we know that only two people joined the conversation, at the beginning of the 15th second and at the end of the 26th second there are three or four-note chord shapes. Therefore, we can assume that "chords" contain some formants of the interactor's voice. At this point, we should mention the timbral quality of human voice of speech. As a further investigation, timbral aspects of interactors could be observed in order to compare with musical instruments

(acoustic and synthesized), and may result with the replacement of human voice with an appropriate instrument.

While identifying pitch, one could observe that Melodyne also identifies a tonal center. In the analysis section, accidentals used will not be put onto the staff but the pitch analysis will be taken as it is produced by the programme.

Melodyne is mostly used to correct vocal tracks in the music industry. Therefore, one should be aware that it is not exactly a scientific tool for pitch analysis, and that the results that it provides might not be sufficient for our analysis. But on the other hand, using Melodyne is important to understand the comparison of different programs identifying and manipulating pitch.

The second tool which yields a similar but more detailed analysis is Tony. Tony is a software program developed by researchers from Queen Mary University and New York University for high quality scientific pitch and note transcription. The program aims to make scientific annotation of melodic content, and especially a more efficient estimation of pitches in singing (Mauch et al, 2015,p.1). This program automatically analyses the audio to visualise pitch tracks and notes from monophonic recordings. Tony also enables its users to manipulate given data, and it also contains a spectrogram. Tony's transcription between seconds 40 and 50 is shown in Figure 3.5 below.

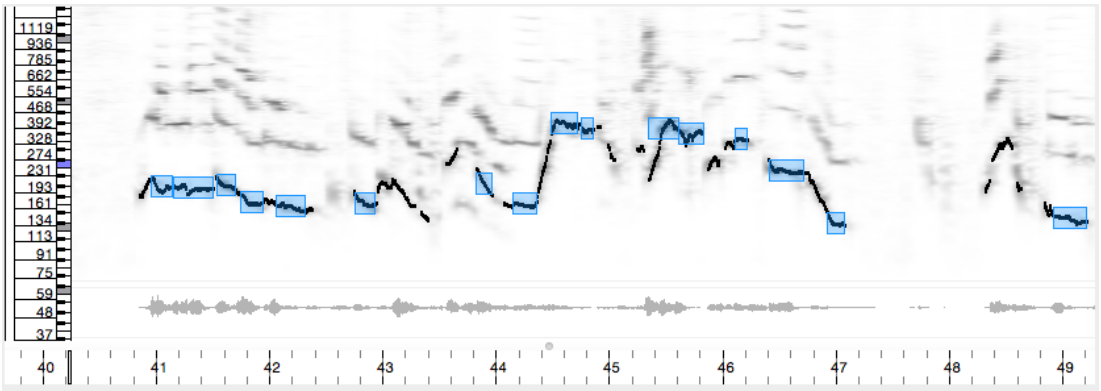

Figure 3.5: Transcription from Tony, between seconds 40 to 50.

Note that the left column shows frequency, the bottom panel indicates time in seconds, and above it stands the waveform panel. Most important features are the pitch track and exact note values, including microtones, illustrated with blue regions. Pitch track information could tell something about one of the theoretical foundations of this thesis: In the third chapter, terms and examples of discourse analysis from different researchers were explained. As one of the terms mentioned, let's consider low pitch boundary that provides closure at the end of a set of interdependent constituents. Between 45th and 47th seconds, Neslihan ends her sentence with the words "…ama, mesela şımartılmamış" which can be translated "…but, for example, he's not been spoiled" referring to how her friend has been raised. The word "şımartılmamış" begins at 46.2 and ends in 47.1. The downward motion of the pitch tracker demonstrates the low boundary. Such basic analysis of conversation would be also useful in order to shape the use of musical transcription and musical writing.

The third tool used is Praat. Praat is an open coded and free software tool for linguists to analyze speech or any acoustic analysis. It is written and maintained by Paul Boersma and David Weenink of the University of Amsterdam. Praat offers a wide range of standard and non-standard procedures, including spectrographic analysis, articulatory synthesis, and neural networks. The main reason for using Praat is to get specific information on speech. Information on male and female speakers will form a solid basis to the musical analysis.

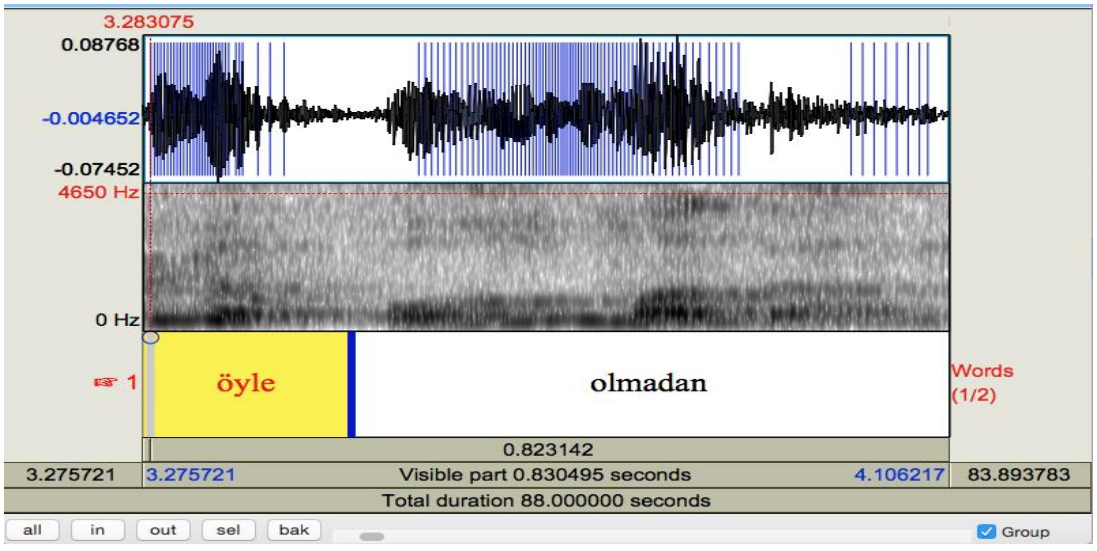

Figure 3.6: Amplitude pulse and spectrogram views of words "öyle olmadan" by male speaker between 3.27 and 4.1 seconds. Blue lines indicate pulses.

Praat's spectrogram demonstrates a more detailed analysis than Tony's. In the spectrogram view shown in Figure 3.6, the horizontal axis indicates the timeline and the vertical axis indicates the frequency level. When the color is darker, more of that frequency level is featured at that moment. Stripes in the spectrogram represent vibrations of the vocal chords.

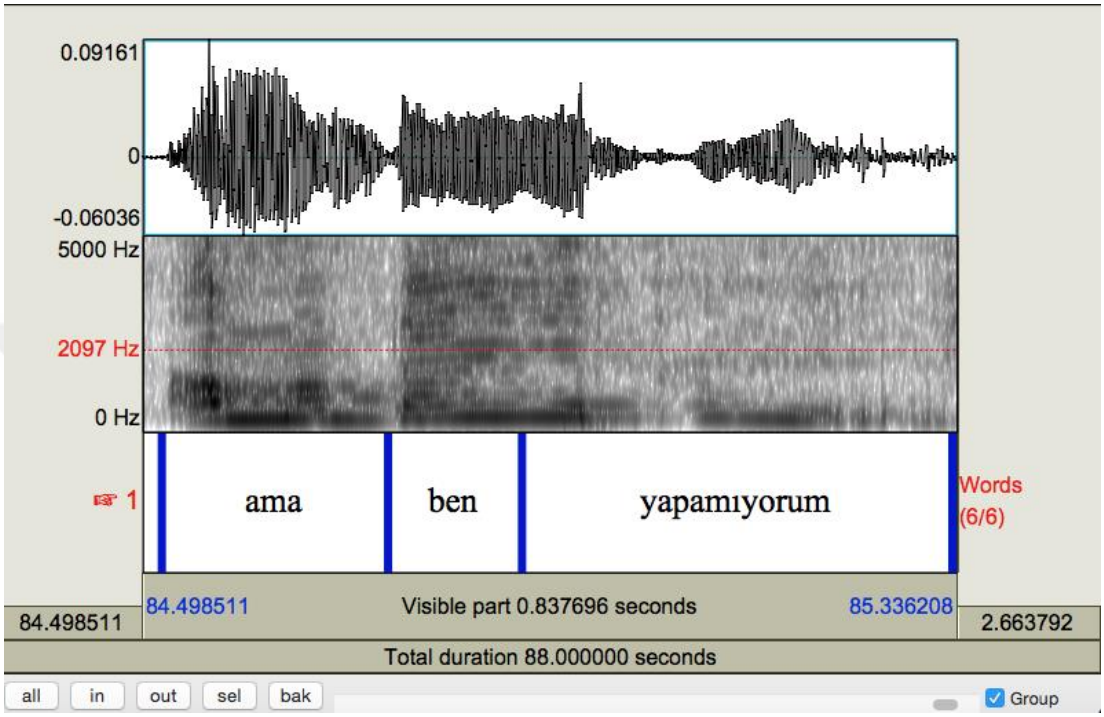

Figure 3.7: Amplitude and spectrogram views of words "ama ben yapamıyorum" by female speaker between 84.5 and 85.3 seconds.

There are several differences to notice between the spectrogram displays of male and female speakers. The male speaker has a distinctive amount of stop between words and he uses low frequencies. On the other hand, the female speaker's pronunciation has more singing qualities and uses a wider range of frequency. It is hard to visualize separate words on her amplitude table. Words "ben" and "yapamıyorum" are almost pronounced as a single word in Figure 3.7.

In order to see other similarities and differences between male and female voices, it is useful to create another demonstration containing pitch, intensity and formant information of the two interactors. In Figures 3.8 and 3.9 below, note that the blue lines indicate pitch, the green lines indicate intensity and finally, the red dots indicate formants.

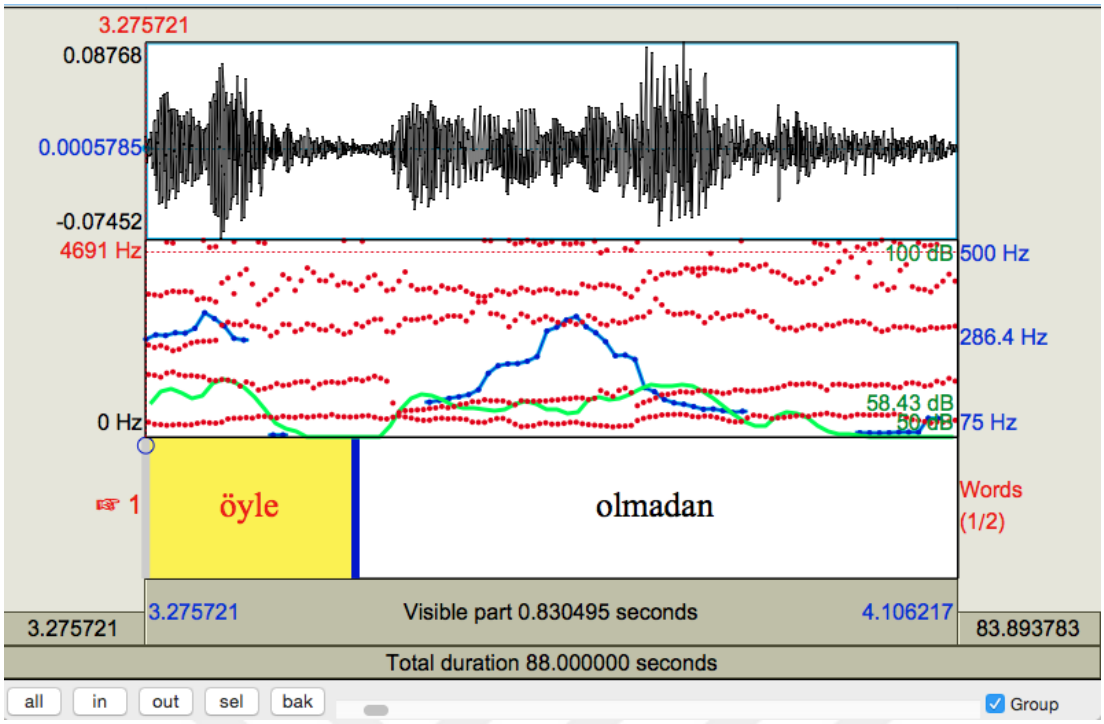

Figure 3.8: Amplitude, pitch, intensity and formant visualization of words "öyle olmadan" by male speaker between 3.27 and 4.1 seconds.

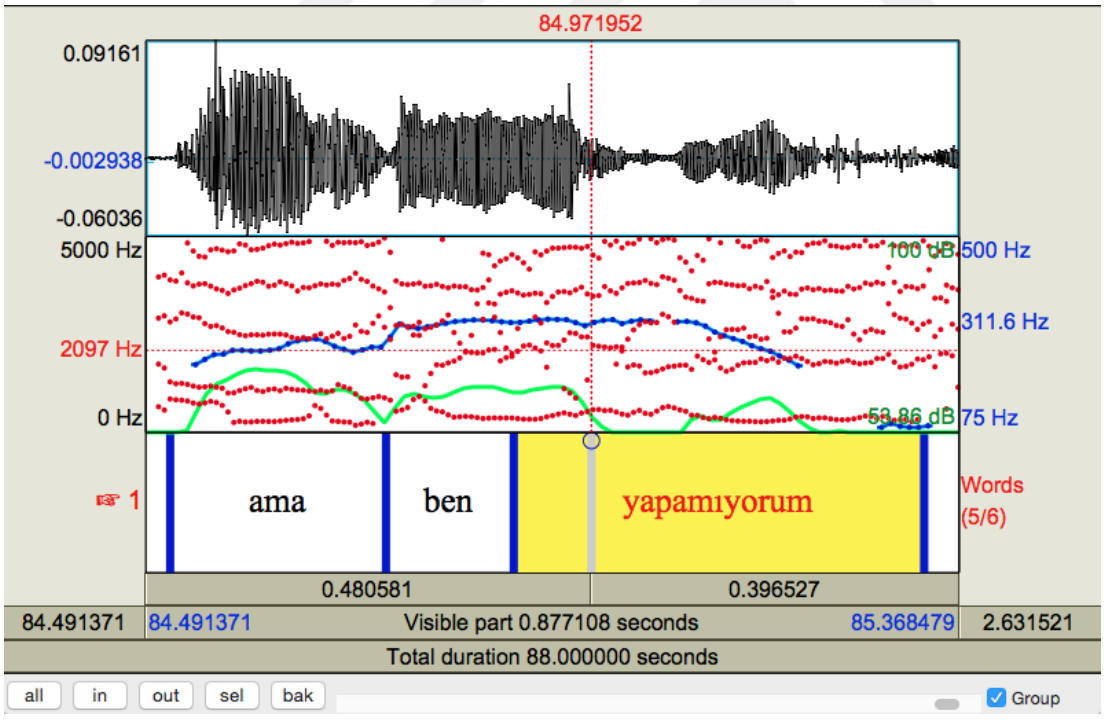

Figure 3.9: Amplitude, pitch, intensity and formant visualization of words "ama ben yapamıyorum" by female speaker between 84.5 and 85.3 seconds.

The information and analysis presented above could be useful in order to conduct a sine-wave speech experiment for further research. As a musician, I would wonder if an entire song could be made through the reproduction of a dialogue. A synthesized version of the recording could be recreated to get tested in terms of listener perception and to answer the following questions : Given the priori information or exposing listeners to speech, would it be possible if musical values such as rhythm and pitch, could also contain a semantic value or could reflect the correct semantic value with a priori information? Or in other words, would it be possible to obtain semantic data from reproduced sine-wave speech? If yes, how would this phenomenon affect the listener's perception?

3.2. Analysis

After the transcription process, especially with Tony and Melodyne, a manual transcription has been made by listening in order to get a deeper musical analysis. One should remember that this transcription may even contain some missing elements and subjective voicings since it is done by ear. Somehow, it is impossible to completely reproduce the exact transcription of speech without using a microtonal approach. In a future study, a chosen system of microtonal approach could be adopted in speech-music research in order to broaden the pitch palette that this thesis benefits from.

The rhythmic aspect of the dialogue is transcribed in order to get useful information related to our topic. Therefore, some rests between each interactor's words may not be notated exactly, since such a transcription would complicate the process of analysis and human speech is nearly impossible to replicate with musical instruments, at least with human performance, as observed in Peter Ablinger's case. However, the focus of this thesis is not to imitate exact speech but to investigate different aspects of a dialogue through a musical perspective. Therefore, the recorded dialogue has been analyzed in order to find mentionable quality.

As a result, a tonal, re-interpreted version of the spoken dialogue will be exhibited, in order to find musical entities with the help of the knowledge manifested in the previous part of this thesis.

The selected dialogue begins with a suggestion from the male speaker. He reflects his thought with a short sentence, as demonstrated in Figure 3.10.

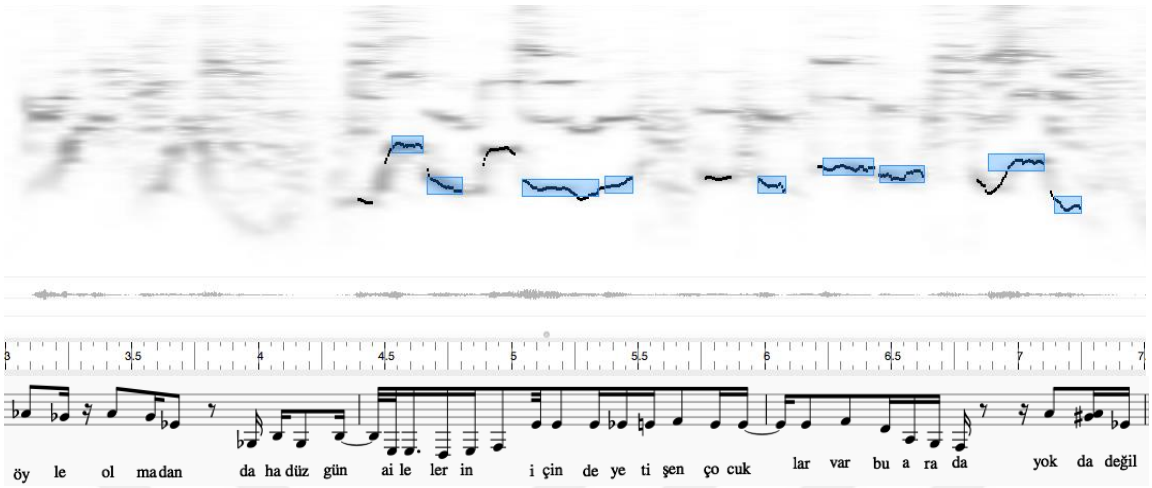

Figure 3.10: Recorded dialogue, secs. 3-7.

In the figure above, waveform, pitch track information, note information with blue boxes and the timetable indicated are acquired from Tony. The sentence by the male speaker "Öyle olmadan, daha düzgün ailelerin içinde yetişen çocuklar var bu arada. Yok da değil.", can be translated as "By the way, there are children who grew up in proper families". This particular suggestion is articulated in such a way that one could think of it as a complex musical entity by itself. Even Tony couldn't analyze its pitch track and note values; the first two words of the first sentence "Öyle olmadan" are transcribed as A flat – G flat and A flat – G flat and E flat respectively. The two-word block shows similarity with a classic five-note idea which could be heard in countless improvisations and frequently used in almost every musical culture. Conversation analysis and the musical transcription both demonstrate that the two-word block contains a *partially falling boundary*, from G flat to E flat, indicating that the information flow will continue from the same speaker.

Followed by a stop which is equivalent to a comma, the rest of the sentence is performed. The second part of the phrase is harmonically and rhythmically more complex. Beginning of the second part of the phrase, "daha düzgün" is transcribed as two notes, G flat and B respectively. Listening to the recording, and following the pitch track information from Tony, one could notice that low and high pitch accents are associated with both of the words, which contrasts with the antecedent two-word block

“öyle olmadan”. However, in the rest of the second part of the phrase, rhythmic pace accelerates through the end of it. After the word “düzgün” that ended with a high pitch, “ailelerin” is pronounced in such a way that the former two syllables have downward pitch motion and the latter two have upward. The word is transcribed as E flat – D – E flat – F. It is interesting that upward pitch motion relates the second part of the word (–lerin) to the next word in such a way that it is almost heard as a conjunctive. Raising from the low pitch, “ailelerin” is followed by high pitched information in a fast rhythm. “İçinde yetişen çocuklar var” sustains the note E for most of the time, except the word “yetişen”, which has chromatic E flat – E – F notes. The use of chromaticism in such speed and context might exemplify the borrowing from improvised speech into musical ground. The following words “bu arada”, translated as “by the way” are transcribed as D – A – G – F respectively. Lowering pitch indicates the end of information traffic from Haydar. That information is also obvious for Neslihan, therefore, she instantly throws the name Memo to embody Haydar’s argument. Memo is transcribed as A sharp and E flat. However, Haydar decides to add an extra verbal content to strengthen his statement. At this point, Neslihan’s word “memo” and Haydar’s “yok da değil” overlap and this makes it difficult to understand exact pitches. The overlap is observed in terms of turn-taking aspect of conversation analysis.

Another remarkable point is that while listening to the recording, it is seen that despite the overlap , including hearing Neslihan’s word, it takes exactly one second for Haydar to react to Neslihan’s claim after finishing his sentence. As one might observe, another aspect of the speech that has been recorded is the lack of thematic progression. One could think that a musician, while improvising such a passage would break, expand and repeat the melodies demonstrated above.

The next phase of the recorded dialogue demonstrates another main domain of conversation analysis, namely repair. As a dramatic example, the musical transcription between 10 and 15,5 seconds is demonstrated in Figure 3.11.

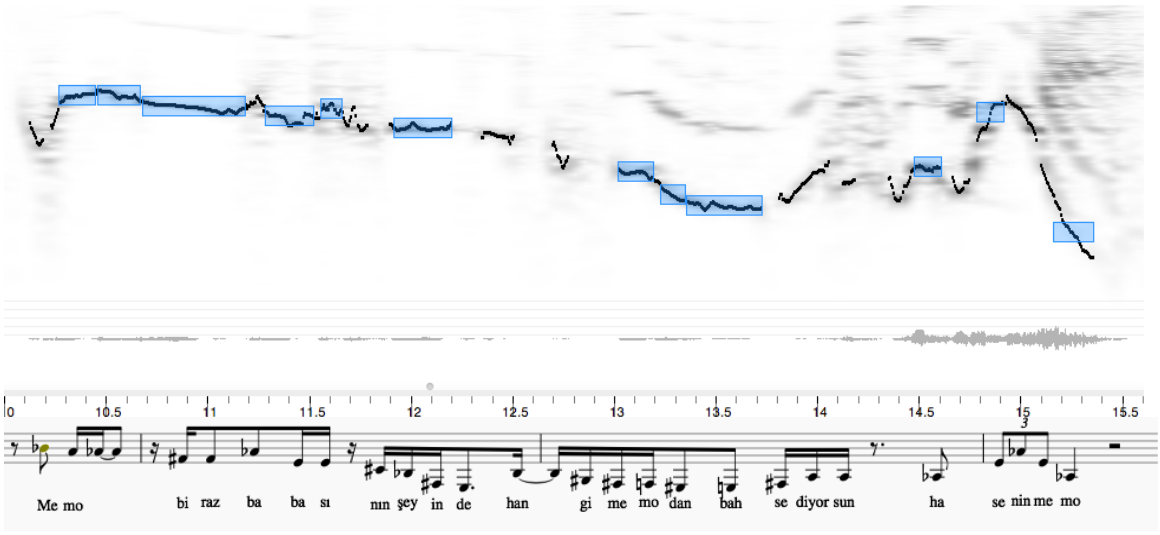

Figure 3.11: Recorded dialogue, secs. 10-15,5.

"Memo" (the local abbreviation of the name Mehmet) pronounced in the first measure, clearly indicates a disagreement. In other words, "Memo" is the *repairable* of the conversation at this point. The high pitched pronounciation of the name and chromatic descent from B flat to A flat make it obvious that Haydar does not think that "Memo" fits into the subject discussed. In addition, he does not complete his sentence. That particular type of intonation and sentence is also noticed from Neslihan and she rapidly asks the question to ensure that both of them are talking about the same person. Beginning at the 13th second, she asks "hangi memodan bahsediyorsun?". The question could be divided in two parts as descending and ascending. The second part which is ascending, ends with an A and indicates a clear type of oral question. It contains typical elements of speech prosody: rhythmically fast and dynamically regular.

The question is followed by a correction of the misunderstanding, starting at the end of the third measure, a half-step lower from the female interactor's last note. "Ah senin Memo", is basically emphasizing the correct person. One could see that is an arpeggiated triad, consisting of A flat and E. This sentence, listened alone, could be heard almost like a musical cadence. In addition, in terms of musical dynamics, it is an obvious *forte* sentence.

The recording continues with the female speaker's friend's childhood story which is demonstrated in Figure 3.12. Note that in this figure Neslihan is located at the top staff and Haydar at the bottom, labelled with N and H respectively.

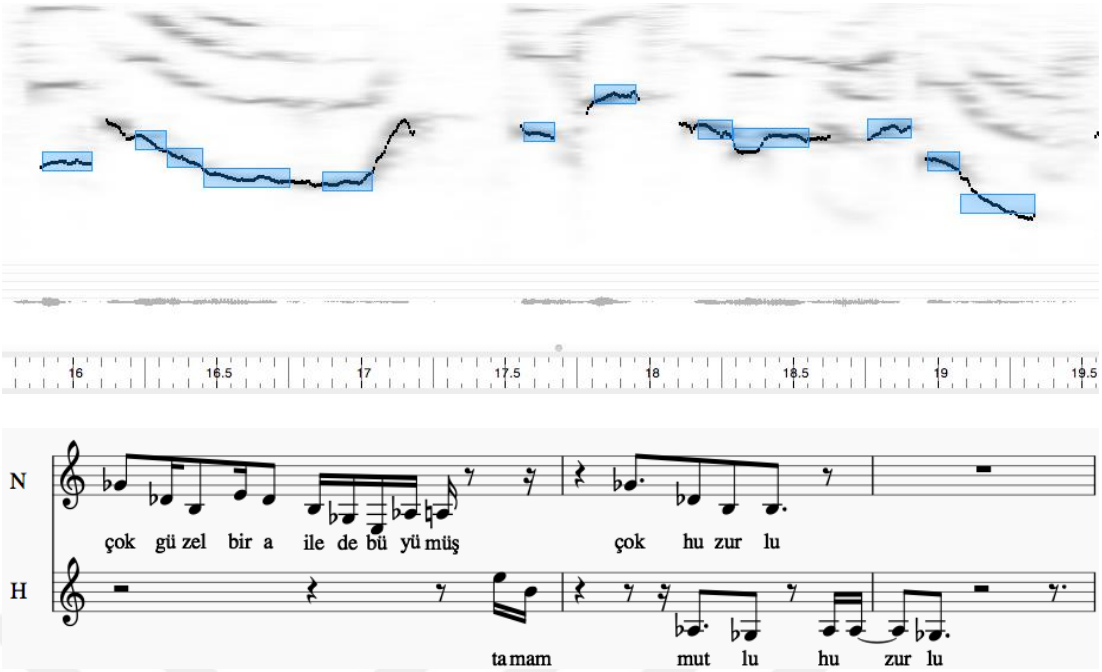

Figure 3.12: Recorded dialogue, secs. 15-20.

This example demonstrates not only another overlap, but also fast approval elements from Haydar. Haydar's interaction is not made to create turn-taking, but to justify that he understands the story that has been told. After the first five words "Çok güzel bir ailede büyümüş" from Neslihan, Haydar instantly justifies her with the word "tamam", constituted of notes B and E. One should note the lack of rest between the two interactors' parts. That instant reaction is remarkable in order to analyze dialogues because such an instant response asserts that the male speaker has already decided to announce that he understood what the female speaker says before she completes her part of the sentence. Therefore, he doesn't wait for the end of information from Neslihan, and keeps low dynamic. At this point it is useful to say that Haydar's approval is not limited with his words, but also his gestures (head shaking) and facial expressions are equally dominant in the conversation. However, the importance of gesture in human interaction is beyond the scope of this thesis.

This passage is a two-part writing in terms of musical perspective. Therefore, it requires observation on sudden pitch relations between interactors. Neslihan's passage, "Çok güzel bir ailede büyümüş, çok huzurlu" which can be translated as "He grew up in a beautiful family, very peaceful". The two words "çok güzel" are transcribed as G flat, D flat and B. The ascending characteristic of the word

"büyümüş" which is followed by a stop, is the sign of upcoming information. The next adjective "very peaceful", still defines the female friend's family.

In the second measure, the overlap of the male and female speaker's parts is also remarkable. The word "Çok" that Neslihan pronounces seems to be complemented by Haydar's word "mutlu", just like the first example. However, the female speaker tends to finish her sentence with the word "huzurlu" and the male speaker instantly repeats the adjective "huzurlu" to justify her statement. The overlap creates a rich polyrhythmic view. The pitch duo that the male speaker uses for the words "mutlu" and "huzurlu" are A flat and G flat.

Another dramatic example from the recorded excerpt contains a verbal description of some human characteristic. Between seconds 33 and 35, the female interactor gives a dramatic description about her friend's general emotional profile. The musical transcription of this section is demonstrated in Figure 3.13.

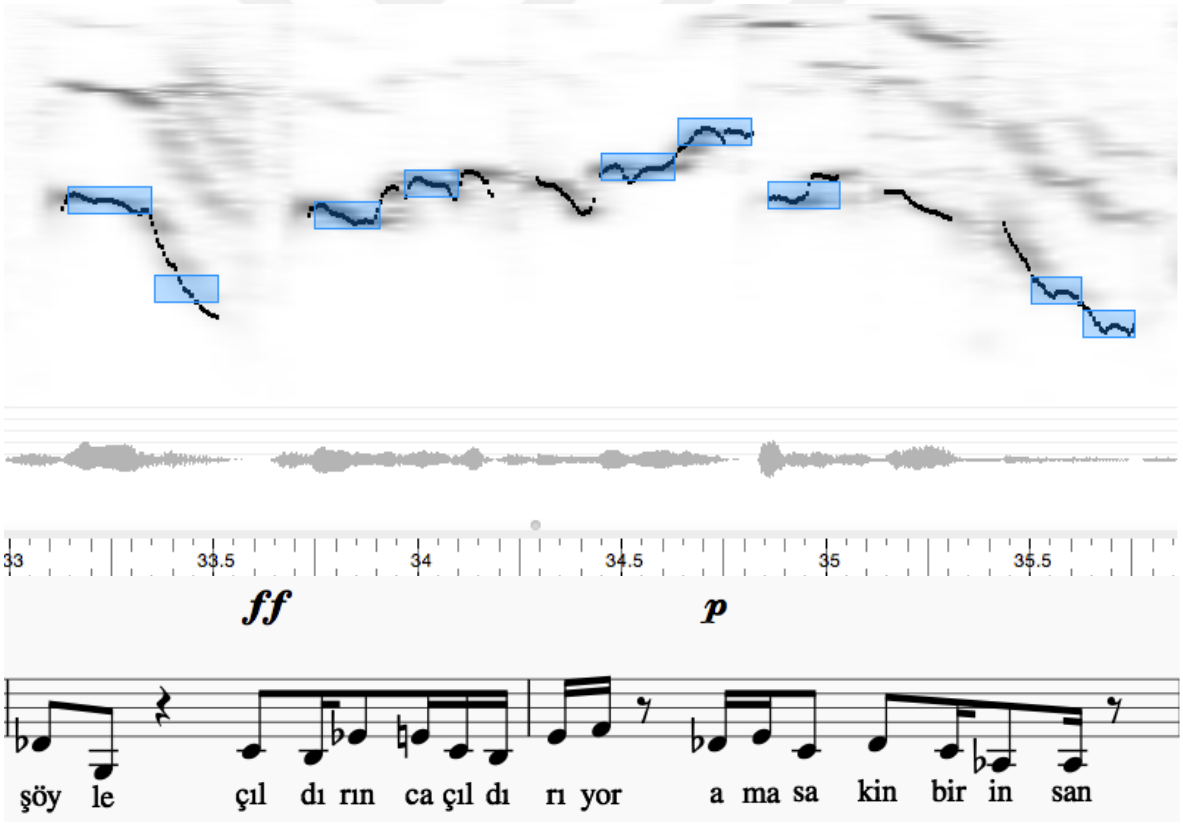

Figure 3.13: Recorded dialogue, secs. 33-35.

The phrase above could be translated in English as: "Namely, he loses it when he gets angry, but he is a calm person in general". The interesting point is the way that

Neslihan uses a *forte* dynamic when she says: “he loses it when he gets angry”, but instantly calms down and uses a *piano* voice when she mentions that her friend is a calm person in general.

The examples mentioned above have been chosen to demonstrate rhythmic, dynamic and melodic qualities of expression in interaction. However, where exact note values are ignored, talking about pitch groupings would be no more than speculation. Therefore, as mentioned earlier, the use of a microtonal system in looking at exact pitch values will also fill the lack of presence in sound groupings.

4. CONCLUSION AND FURTHER RESEARCH

Speaking is a particular action that defines us as human. The process of speaking is the most common type of interaction so far, and it is mostly made spontaneously starting at a very early age, even without any formal education. However, spontaneity in music requires at least the same amount of effort as the action of speaking. Therefore, there is no doubt that musical improvisation or "talking fluently" with a musical instrument is a tremendously hard work. Improvisers and composers spend their lifetime to learn how to reflect their own musical creativity through instrument(s). They practice for months before they premier a piece. One could say that music making, unlike speaking, is hardly an uninterrupted reflex.

Imitation of speech in music has been a familiar process among musicians for some time. However, despite its importance, imitation of speech dialogue in music seems intact. Studies are predominantly based on semantical, phonological or anthropological aspects of speech.

This thesis represents an analysis of a random and spontaneous dialogue using musical terminology. However, some scientific facts about both music and speech make this process intricate for researchers. At first, the concept of rhythm is much more determined and plays an important role in music and in its terminology. Rhythm in music is basically the time pattern, enabling musicians to see some kind of path and follow it through. Even though it can be complex and variable through compositions and improvisations, rhythm and other rhythmic features are precise and determined. However, analyzing speech demonstrates us that rhythm in dialogue is not a pre-determined or "locked" one. It can be seen more like a flow, collaboratively constructed, controlled and conducted by interactors. In the example presented in this study, the female speaker tends to talk much faster and *passionato* while the male speaker tends to talk more *piano*, as if he wishes to balance the general mood of the conversation.

The particular flow mentioned above is effortlessly performed while interactors speak with each other, since one of the main focuses in a conversation is still words and their semantic values. People don't tend to think about rhythm, and this phenomenon creates the flow in their conversations. How could this naturally flowing element occuring in dialogues be adapted to music ? There are three different questions that as a researcher I want to investigate as further projects. First, would it be possible to have a entity of sound, created with the sine-wave reproduction of a dialogue that provides lexical data, given prior information ? In other words, is it possible to make a wordless song from spoken words, in which one would understand what words are said, topic or mood by listening to it. The second question is what kind of material would this process provide musicians that they can adopt and improve as creative and performative concepts? Finally, from a musical perspective, how is the relation and interaction of speech dialogue with the other sounds present in the environment?

It can be seen that the rhythmic complexity of a dialogue provides a wide range of opportunities, not only for improvisers but also contemporary composers. Moreover, the reactions of the interactors depending on their mood, perception and even instrument quality, change their way of rhythmic and dynamic approach as a collaborative sum.

The flow mentioned above can not only determine the dynamic range of the piece/conversation but also lead to a different topic that might be appropriated to *movements* in music terminology.

Another argument that is worth mentioning is the intersection of dynamic range and semantics. Naturally, various levels of tension on various topics have a sudden impact on dialogues. The frustration or even any description of it –as seen in Figure 3.13 – could suddenly change dynamics dramatically. Moreover, those dynamics could be also affected by outside factors, as mentioned in the introduction.

Several types of interactions between speakers could give some ideas about phrasing and the compositional structure of the music. In speech, sudden interruptions, confirmations or disagreements occur in a natural way. Those elements are crucial and underrated elements of the flow. They could be laughs, sighs, howls or even silence. Those reactional sounds shape dialogues and give additional emotional information

about the speakers. Analyzing and implementing such figures could develop not only a musician's way of playing but also their way of hearing music.

One should be aware that this thesis tends to create several questions, some of which have been asked above. Especially, sine-wave speech experiments are promising in order to reveal some speech-music semantic value. Listening experiments could tell us where the music of speech is in the listener's perception.

However, music does not only contain rhythm or pitch, but also other important acoustical features such as timber. The perspective mentioned above is only a particular one that should be used with many others.

REFERENCES

Audio Engineering Society. (2005). Recording Technology History. Retrieved January 21, 2017, from http://www.aes.org/aeshc/docs/recording.technology.history/notes.html

Barrett, G., D. (2010). *Window Piece: Seeing and Hearing the Music of Peter Ablinger*. Retrieved from http://ablinger.mur.at/engl.html

Barrett, G.,D. (2009). *Between Noise and Language: The Sound Installations and Music of Peter Ablinger.* Retrieved from http://ablinger.mur.at/engl.html

Benson, B. E. (2003). *The Improvisation of Musical Dialogue, A Phenomenology of Music.* New York. Cambridge University Press.

Berliner, P. F. (1994). *Thinking In Jazz, The Infinite Art of Improvisation.* Chicago. The University of Chicago Press.

Black, S., H. (2001). *Semantic Satiation and Lexical Ambiguity Resolution*. The American Journal of Psychology, Vol. 114, No.4.

David, M. A., Johnsrude, I. S. *Hearing Speech Sounds: Top-down influences on the interface between audition and speech perception.* Elsevier, 2007, Retrieved from www.sciencedirect.com

Deutsch, D., Henthorn, T., and Lapidis, R. *Illusory transformation from speech to song*. Journal of the Acoustical Society of America, 2011, 129, 2245-2252.

Feld S., Fox, A. A. (1994). *Music and Language.* Annual Reviews. Annual Review of Anthropology, Vol. 23. Retrieved from http://www.jstor.org/stable/2156005.

Fuchs, C. (2017). *Social Media: A Critical Introduction*; Sage.

Hanning, B. R. (2010). *Concise History of Western Music* (4th Edition). W.W. Norton & Company, USA.

Kent, R. G. (1939). *Reconstructing the History of a Language.* Retrieved from http://www.jstor.org/stable/4340586

Lee, D. & Hatesohl, D. (2017). *Listening, Our Most Used Communication Skill.* Retrieved November 1, 2017 from http://extension.missouri.edu/p/CM150

Monson, I. (1996). *Saying Something, Jazz Improvisation and Interaction.* Chicago. The University of Chicago Press.

Oxenham, A. J. *The Perception of Musical Tones.* The Psychology of Music 3rd Edition.

Patel, A. D. (2008). *Music, Language and the Brain*. Oxford, New York: University Press.

Peterson, G. E. (1955). *An Oral Communication Model.* Language, Vol. 31. No. 3. Retrieved from http://www.jstor.org/stable/410809

Ross, D., Choi, J., Purves, D. (2007). *Musical Intervars in Speech.* Proceedings of The National Academy of Sciences of The United States of America. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1876656/

Scheirer, E. & Slaney M. *Construction and Evaluation of a Robust Multi-feature Speech/Music Discriminator.* Retrieved from 0-ieeexplore.ieee.org.divit.library.itu.edu.tr

Schultz, P. J, Cobley, P. Theories and Models of Communication, 2013, p.31

Segnini, R. & Ruviaro B. (1989). *Analysis of Electroacoustic Works with Music and Language Intersections.* Retrieved from https://ccrma.stanford.edu/~ruviaro/texts/Ruviaro_2005_Segnini_Analysis_Music_Language_Intersections.pdf

Sidnell. J. (2016). *Conversation Analysis.* Retrieved from http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-40

Steele, J. (1775). *An Essay Towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols.* Retrieved from https://play.google.com/books/reader?printsec=frontcover&output=reader&id=wgFFAAAAcAAJ&pg=GBS.PR1

Şahin, L. (2015) *Genre Based Accordion Recording Techniques: Balkan, Caucasian, Tango, Turkish Folk Music.*

Tierney, A., Aniruddh, P. (2016). *Acoustic and musical foundations of the speech/song illusion.* International Conference for Music Perception & Cognition. San Francisco.

Tomlinson, G. (2015). *A million years of music: the emergence of human modernity.* Brooklyn, New York. Zone Books.

Watson, S.K., Townsend, S.W. West, W. & Slocombe, K.E. (2015). *Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees.* Retrieved from http://dx.doi.org/10.1016/j.cub.2014.12.032

Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis.* Oxford University Press.

Wlodarski, A. (2015). The composer as witness: Steve Reich's Different Trains. In *Musical Witness and Holocaust Representation* (Music since 1900, pp. 126-163). Cambridge: Cambridge University Press. doi:10.1017/CBO9781316337400.006

Zatorre, R. J. (2002). *Structure and Function of Auditory Cortex: Music and Speech.* Trends in Cognitive Sciences Vol.6

Zatorre, R. J., Gandour J. T. (2007). *Neural Specializations for Speech and Pitch: Moving Beyond the Dichotomies.* Philosophical Transactions of The Royal Society. Doi: 10.1098/rstb.2007.2161

Zatorre, R. J., Baum, S.R. (2012). *Musical Melody and Speech Intonation: Singing in a different Tune?* Plos Biology, vol. 10, issue 7.

Zeller, H. R. (1960). *Mallarmé and Serialist Thought.* Die Reihe. Speech and Music, Vol. 6. Leeds. John Blackburn Ltd.

BIBLIOGRAPHY

Bamdad, T., Okada, B. M. & Slevc. L. R. (2016) *High Talkers, Low Talkers, Yada, Yada, Yada: What Seinfeld tells us about emotional expression in speech and music.* International Conference for Music Perception & Cognition. San Francisco.

Bas Cornelissen, B. Makiko Sadakata, M. & Honing, H. (2016). *Categorization in the speech to song transformation (STS).* International Conference for Music Perception & Cognition. San Francisco.

Battcock, A. & Schutz, M. (2016). *Signifying Emotion: Determining the relative contributions of cues to emotional communication.* Paper presented at the International Conference on Music Perception and Cognition. San Francisco, CA.

Cohen, S. D., Wei T. E., Defraia, D. C. & Drury, C. J. (2011). *The Music of Speech: Layering Musical Elements to Deliver Powerful Messages.* Retrieved from http://isites.harvard.edu/fs/docs/icb.topic657116.files/The%20Music%20of%20Speech_Distribution.pdf

Dzhambazov, G., Senturk, S. & Serra, X. *Automatic Lyrics-To-Audio Alignment in Classical Turkish Music.*

Foucault, M. (1972). *The Archeology of Knowledge and The Discourse of Language.* New York. Tavistock Publications Limited.

Pabst, A., Vanden Bosch der Nederlanden, C., Washburn, A., Abney, D., Chiu, E. M. & Balasubramaniam, R. (2016). *Perceptual-Motor Entrainment in the Context of the Speech-to-Song Illusion: Examining Differences in Tapping Patterns to Perceived Speech vs. Song.* International Conference for Music Perception & Cognition. San Francisco.

Pfeiffer, R.E., Williams, A.J., Shivers, C., Gordon, R.L. (2016). *Using a Rhythmic Speech Production Paradigm to Elucidate the Rhythm-Grammar Link: A Pilot Study***;** Podium presentation at International Conference for Music Perception & Cognition 14. San Francisco.

Phillips, N., Amick, L., Grasser, L., Meujeur, C., & McAuley, J.D. (2016). *The Silent Rhythm and Aesthetic Pleasure of Poetry Reading***.** The 14th International Conference on Music Perception and Cognition, San Francisco, CA.

Robledo J. P., Hawkins Sarah, Cross Ian, Ogden R., (2016). *Pitch-Interval analysis of 'periodic' and 'aperiodic' Question+Answer pairs*. Speech Prosody

Conference. Boston.

Schultz, B. G. & Kotz, S. A. (2016). *Finding the beat in poetry: The role of meter, rhyme, and lexical content in speech production and perception*. International Conference for Music Perception & Cognition. San Francisco.

Schwarz, N. (2010). *Feelings as Information Theory.* Michigan. Handbook Theories of Social Psychology.

Stockhausen, K. (1960). *Music and Speech.* Die Reihe. Speech and Music, Vol. 6. Leeds. John Blackburn Ltd.

Weidema, J. L. Roncaglia-Denissen, M. P. & Honing, H. (2016). *Top-Down Modulation on the Perception and Categorization of Identical Pitch Contours in Speech and Melody.* International Conference for Music Perception & Cognition. San Francisco.

Woodruff, J. (2014). *Tonal Discrepancies and Cognitive Dissonance in Peter Ablinger's Voices and Piano: Angela Davis*. Kunst Muzik, Vol. 16.

APPENDICES

APPENDIX A: Jason Moran's interview transcription

APPENDIX A

H.C.: So, I am doing my master thesis on speech-music. That's basically what you did on *Ringing My Phone,* but I am also interested in observing the dialogue. I would like to ask you when did you first meet with the idea of speech-music? When did you think speech could be translated into music?

J.M: I heard Hermeto Pascoal. I heard a record. He plays with like animals, and he plays teacher in the school, a politician and I just thought wow! That was just fascinating.

H.C.: What was your age?

J.M: I was 20 or 21 when I heard him.

H.C.: I see that in your other concerts *Ringing My Phone* is not the only song that you imitate speech. How did the process go through? How did you start to play speech?

J.M: I heard and then I started to play, because it made me rethink about how I was listening the language or the sound you know. In music school they teach you about music but they don't teach you about the sound. They teach you just about the music so such and such wrote this music and you read the music. They don't teach you how to listen. I think Hermeto was listening to the world, Messiaen listens to the world and you hear it in the music. Also hip-hop listens to the world. The first one actually we played tonight we haven't played in a long time. We don't use the audio, it's a song on my second album *Thief Without Loot.* So, we played it tonight which he wanted to play, we never play that song anymore but that's Japanese and it plays the language just the phrases. What it does also is that it challanges the technique. It makes your hands do things differently. While you trying to find the pronounciation of it, your hands play differently than they play Bach or Bud Powell you know. That was really exciting, felt different when you played something that someone says.

H.C.: So did somehow the perception of speech as music manipulate your playing and even your hearing?

J.M: Yes

H.C.: Another question… I am prepared by the way because I don't want to take much of your time. How many times did you listen to before you started playing the speech in *Ringing My Phone*?

J.M.: I listened to it all day. I get on the subway and I just listen to it over and over again, all day long. Then I thought it to them I was like "come over the house! I am gonna teach you this stuff" and I have written it out but you really have to listen to find where it goes you know… Those kinda hard for us for a while but we didn't just play the song. Whether we made mistakes we just played it on. Then all of a sudden "now we know the gesture", how the "melody" sounds. But I would listen to it for months before I started. Then I got my *Logic* program, just go section by section, put my section shortly, you know, two seconds at a time.

H.C.: That's exactly the way I do it now.

J.M.: (Laughs) Then you listen to it, you are like "Wow! It is done!". Then you have to play the whole thing, that is the hard part.

H.C.: Also you add your own creativity in addition.

Did you ever think about interpreting a dialogue? I know that the Turkish girl is talking on the phone but what about her mother?

J.M.: I think that is also what makes it good, you know. That piece in particular is good because every person has had a conversation with his mother on the phone or every person has heard someone having a conversation and only hear one part of the conversation. So the imagination goes to the otherside in the space like there is this negative space that I think also music needs. Because also the way we play, we three musicians are also talking, right? So that is how we talk to each other for a long time. I have never done two people talking together, I am lazy (laughs). That is more to do with it.

H.C.: Maybe a three dimensional interpretation of whole room (laughs). We are talking and they are talking.

J.M.: They are good, they are laughing now (laughing).

H.C.: What did you encounter rhythmically?

J.M.: Well, it is to not think that is rhythm though. I never count when I am talking to you. I dont think like “oh that is 75 bpm”. I don’t think that. So, it actually has more freedom. When you understand that a phrase of music doesn’t have to be locked. Because we walk around in our lives not locked, we are not robots, it’s fluid you know. Learning these songs and having figured it out, we play a lot of different languages by the way, we have just never recorded them. While we’re learning them we also try to learn our improvisational language too. All of this was happening at the same time. We were trying to figure out our group sound. So that language also helped us like “Oh! This is very different feel, what if we make some other songs that feel like this?”

H.C.: Well I believe those are all of my quetions. Thank you very much.

J.M.: You are welcome. I loved what you do by the way. Do you play piano too?

H.C.: No, I play guitar.

CURRICULUM VITAE

Name Surname **: Esen Haydar Cengiz**

Place and Date of Birth **: Istanbul 27/08/1984**

E-Mail **: ehcengizister@gmail.com**

EDUCATION **:**

- **B.Sc.** **:** 2009, University of Marmara, Economics, Economics

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2012-2013 Diversey Sales Representative
- 2011-2012 Peugeot Key Account Responsible
- 2009 Borusan Holdıng – Land Rover Sales Team Member