Daniel Buner Formo

The Music of Language and Language of Music

Norwegian Artistic Research Fellowship Programme

Trondheim, June 2018

Norwegian University of Science and Technology
Department of Music

◼ NTNU

Table of Contents

The Music of Language and Language of Music

PROJECT OUTLINE

Abstract

The artistic research project *The Music of Language and Language of Music* explores the relationship between music and speech, in particular between *improvised music* and *everyday conversation.* As a tool in this exploration, a digital musical instrument has been developed for "orchestrating" musical features of speech into music in real time. This has evolved into an electroacoustic performance concept entitled *The Orchestra of Speech,* where musical features of speech and conversation have been investigated and explored as sources for creating improvised music.

Topics in this work include music and language, prosody, speech genres, improvisation, performance, speech analysis, signal processing, machine learning, instrument development, interface design, sound, perception, semiotics, and form and meaning.

Introduction

This documentation presents the ideas, development, results and reflections of an artistic research project by Daniel Formo entitled *"The music of language and the language of music"*. This doctoral project was conducted as a part of the Norwegian *Artistic Research Fellowship Programme*, and carried out at the Music Department of the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. The main supervisor for this project has been Øyvind Brandtsegg, professor in Music Technology at NTNU. In addition, Diemo Schwarz, researcher at the *Institut de Recherche et de Coordination Acoustique/Musique* (IRCAM), and composer Trevor Wishart, have acted as secondary supervisors.

Research context and documentation format

This project was carried out within the Scandinavian context of artistic research, emphasising *artistic* over *theoretical* practice, and at the core of the project is the creative artistic work from the practitioner's point of view. In the still young academic field of artistic research the norms and forms have not quite settled yet, but with regard to documentation formats a pattern is starting to emerge – at least in the Norwegian context – with artistic research *expositions* being presented on dedicated web sites, or on web platforms like the *Research Catalogue* (researchcatalogue.net) or the peer reviewed *Journal for Artistic Research* (jar-online.net). In order to seamlessly integrate the kind of sound and video examples that are crucial for understanding a project about music and performance, a *web site* format has also been chosen as the main documentation format for this project, available at orchestraofspeech.com

Although this web site is the recommended format due to the important media content, a portable (and printable) digital version of the text content is provided as a compact downloadable file as well, with links to the sound and video examples available online.

To guide my priorities as to how artistic research should be conducted, I have tried to look into the very foundations of the artistic research field: in general, how *research* is justified politically as the primary way of producing *new knowledge* within a field, and in particular how the field of artistic research came into being (in Europe at least) as a direct result of EU free market standardisation of higher education. I discuss this in greater detail in an essay on artistic research included in the appendix, concluding that if we are to take seriously that artistic research is meant to produce new knowledge in the field of art practice, then it has to be relevant for fellow artists regardless of academic affiliation. On this background, I do not view artistic research as separate from artistic practice in general, and this project has been carried out with the understanding of artistic research as the actual practical investigation that is carried out *through* artistic practice (what Frayling termed research *for* art but Borgdorff calls research *in* the arts (Borgdorff, 2012; Frayling, 1993)) and not as an academic theoretical research that is

based on this practice. In line with this, it is the *artistic outcomes* and the potential insights gained from these that represent the results of this research, and not the accounts and reflections contained in this documentation, which, in the words of Adorno, merely is meant to provide the *"thoughts and concepts [that] assemble around art in such a way that the artworks themselves begin to speak"* (Borgdorff, 2012, p. 122).

Content and structure

This documentation is structured as a set of main chapters giving an overall account of the different background ideas, working processes and actual outcomes of this project, as well as essayistic reflections on topics that have been central to the artistic development.

In addition, the web version of this documentation includes the original research blog (http://orchestraofspeech.com/blog) capturing the project's progress during its development. These blog posts serve as stills from the development process with descriptions of events and performances, and as such provide a valuable supplement of details to the overall perspective provided by the main chapters.

I have organised the content of this documentation into three main categories: the first part covers the project's **background**, giving an account of the *ideas* behind the project in relation to the *context* of other musicians in the field. Then the practical **results** are presented, describing the creative musical processes and outcomes of this project. The last part consists of a series of wider **reflections** on different topics related to the overall themes of speech and music as well the results of this project in relation to relevant fields.

Acknowledgements

I would like to thank the Norwegian University of Science and Technology for giving me this opportunity for in-depth study and development, and especially the Department of Music and its unit for Music Technology for an inspiring working environment. I would also like to thank my supervisors for all the help and constructive input during the project, and not least my family for all the support and understanding.

Background

This section presents the ideas behind the project, including the development of its aims and areas of interest, and an account of the chosen approaches and artistic focus in relation to the context of other artists that have been working with similar ideas.

Project topic

The topic for this project is the relationship between music and speech, in particular *improvised music* and *everyday conversation*. From a creative musician's point of view, it explores how features of speech can be used as a source for making improvised music. The main methods for this exploration include the development of a new digital musical instrument system and performance concept for *"orchestrating"* speech into music in real time, the actual use of this instrument in practice through performance, and the method of *improvisation* as an intuitive way of approaching this by ear.

Key questions

To define concise research questions does not always seem relevant in artistic research. Artistic exploration is not hypothesis-led but could rather be seen as discovery-led (Rubidge, 2005), about pursuing an artistic hunch, an intuitive search on the basis of observation and sensitivity towards certain phenomena that seems interesting to explore. For such a search, as Henk Borgdorff notes, the prevailing format for research design is basically inadequate (Borgdorff, 2012, p. 164).

Nevertheless, a certain delimitation of area of interest is necessary to focus the exploration, and the initial aim of this project was defined somewhat loosely as this:

> *To develop an improvisational foundation for making music that is closely related to the genuine human musicality inherent in spoken language.*

During the work with the project, it has gradually become clearer which aspects of the speech/music relation I am actually exploring, and which aspects are not part of this particular investigation. For instance, when grappling with the many narratives and contexts that came into play when introducing speech into the music, I had the growing insight that I perhaps was as much trying to explore the communicative implications of improvised musical utterances as I was exploring the musical potential of speech. This led to a shift away from speech as spoken *word,* with its implicit references to the conceptual reality of semantics and language, to a more specific focus on the nonverbal aspects of spoken utterances that are similar to improvised musical gestures, and to how these gestural aspects of speech work in actual everyday conversation.

Following this shift, the initial and somewhat general aim to explore speech in relation to improvisation has been developed further to be more specifically about the kind of speech going on in real life social interactions, how such social relations are reflected in stylized *speech genres,* about the shape of spoken utterances primarily as vocal 'musical' gestures, and about how such gestures work in relation to the interaction and interpretation going on in improvised music.

This can be expressed through a range of new artistic questions, both practical and more philosophical:

> *What is the musical and communicative potential of vocal gestures of everyday speech when used in a setting of musical improvisation?*

> *How can the musical exploration of everyday speech be integrated in a performance concept based on improvisation?*

> *How does this musical use of such speech gestures affect the perception of both music and speech in the music?*

So, in a sense, what I have been trying to achieve is to juxtapose the act of engaging in musical improvisation with the everyday activity of engaging in spontaneous conversation in a way that hopefully can shed some light on connections between possible meanings and functions of sound gestures in both speech and music.

Background ideas: language, speech and music relations

In a very general sense, the project is based on the idea that spoken language and music are closely related and probably share evolutionary origins, and that it is reasonable to believe that some aspects of creating and experiencing music can be related to the *communicative* role of musical features in speech. Such ideas have been explored from several perspectives in a growing literature on the evolutionary origins and functions of art and ritual in recent decades, as in the interdisciplinary field mapped out by Ellen Dissanayake and others through compilations like for instance *Communicative Musicality* (S. Malloch & Trevarthen, 2009).

The topic then, is not *what* we say, but *how* we say it – how the intonation, register, tempo, rhythm, dynamics, and voice quality form a communicative layer of its own in speech. This is what linguists call *prosody* (from Greek: *towards song*), and this *music of everyday speech* constitutes a huge semantic potential that with or without our knowing expresses our state of mind, our intentions, expectations, attitudes, relations, feelings, notions and views, and which in hermeneutical ways affect how our utterances are interpreted.

As the linguistic fields of prosody and conversation analysis show, these features have obvious pragmatic functions for helping structuring conversation (Szczepek Reed, 2011; Wennerstrom, 2001). But while linguist look at how such prosodic features are used for negotiating turn taking or highlighting new information, it is from a *musical* point of view interesting to see if these structures – nuanced and intuitively meaningful vocal gestures – also can make sense as recognizable patterns in music. In this regard, *improvised* music can be viewed as a particularly close parallel to conversation, as both involve a continuous dialogical negotiation of content and development, conveying intentions with many of the same means and mechanisms.

Another interesting aspect of speech from a musical point of view is how prosodic styles can express *social relationships*. This is what the language philosopher and literary critic Mikhail Bakhtin referred to when discussing *speech genres* – stylistic templates that we tend to use as formal frameworks when constructing utterances on the fly (Bakhtin, 1986). Just like literary genres, they include choice of style and wording, but speech genres also include specific prosodic traits like the use of certain registers, dynamic ranges, vocal effort, tempi etc. Taken together such traits can be seen to form *musical characters* that communicates something important about the social situation and thus provides the context for interpreting the actual words uttered. For instance, the degree of metric regularity of speech conveys something about the social distance, with very dynamic *'tempo rubato'* signifying a close relationship, subjective opinion, private conversation etc., while more even, regular timing is used when referring something objective, impersonal and formal (Leeuwen, 1999). Other significant genre characteristics typically include speech rate (tempo), register (mean pitch), voice quality, loudness, phrase and pause length, melodic contour, dynamic range, etc. All these prosodic traits affect the interpretation of the possible meaning and intention behind any utterance.

This basic gestural layer of meaning is deeply embedded in spoken language, and we intuitively use different genres in different social situations, such as talking to children, to a judge, to a lover, to an audience or a reporter on live TV. The genres are a natural part of the everyday social characters we take on, and only stick out when used differently from what is expected, like for instance the patronizing way of talking to adults as if they were children. According to Bakhtin, there are as many potential speech genres as there are potential social relations. *Small talk, pillow talk, baby talk, interrogation, public address, report, confession,* etc. are only some examples of speech genres where the *form* provides an important part of the meaning of an utterance.

Bakhtin's emphasis on genres derived from his view on language that *words* do not have any meaning by themselves – it is how they are used in a particular *utterance* with a specific social context, that actually provides the

meaning. Speech genres is part of what expresses and generates this social context and thus convey a kind of social meaning of intention that we seem very attentive to. Interestingly enough, this is a kind of meaning that is expressed mainly through *musical* features like rhythm, melody and dynamics.

A musical exploration of the characteristics of such speech genres has therefore been one of the main themes in this project, and is one of the reasons for its focus on prosody as the main musical material of speech.

Musical context

In a historical context, to make the connection between music and language is nothing new. In Europe during the 17th century in particular, music was increasingly seen in connection with Antiquity's highly developed art of *rhetoric* (Bartel, 1997; Bonds, 1991). Music theory books from the period show how much these ideas of rhetoric influenced German baroque music (Mattheson, 1739), and that this music *speaks* is something that *early music* pioneers later have pointed to as a key for interpreting and performing this kind of music (Harnoncourt, 1982). Speech has never been far away in the recitatives of Opera either, and features in the instrumental music of some composers like the speech-melodies transcribed by Janáček, and in the sprech-gesang by championed later by composers such as Schönberg, Berg and Webern.

Nevertheless, it is mainly during the last 60-70 years that the availability of sound recording technology has made possible a much more extensive musical exploration of speech and the voice. Cathy Lane has given a thorough overview of many compositional approaches and contributors in this field (Lane, 2006), many of which also feature in her compilation *"Playing with words"* (Lane, 2008), and are also covered by Michael Vincent (Vincent, 2010). Lane identifies several distinct compositional approaches and techniques using speech and voice in music, from pure documentaristic pieces, montages *(e.g. the radiophonic pieces of Glenn Gould)*, performative explorations of language and the voice *(Aperghis, Berio, Ligeti)*, sound poetry *(Schwitters, Jaap Blonk)*, different ways of electronically transforming recorded speech and song *(Herbert Eimert, Stockhausen, among others)* and the use of speech fragments as melodic motives *(Steve Reich)*. Trevor Wishart in particular has explored many aspects of the voice and speech in his compositions, such as sonic transformations *(Red Bird)*, the voice as icon of personality and identity *(Two Women, American Triptych)*, phonetic units as musical material *(Tongues of Fire, Globolalia)* etc., and has also written extensively on composing using the expressivity of the human voice (Wishart, 1994) (Wishart, 1996) (Wishart, 2012). Other approaches include the connection between sound and text explored by the Swedish tradition of Text-Sound composition after the likes of Lars-Gunnar Bodin (Brunson, 2009). Many have used speech directly as a melodic source, such as Paul Lansky, Paul DeMarinis, Robert Ashley, Scott Johnson, Florent Ghys, Jacob ter Veldhuis, Michael Vincent as well as jazz

pianist Jason Moran. Others have made instrumental music based in various ways on speech, like the spectral analyses transcribed by Jonathan Harvey in his 2008 orchestra piece "Speaking". On the video sharing web site YouTube, there is even a whole sub-genre of musicians "playing" the speech melodies in sync with videos of well-known speeches or TV shows.

Interesting technological approaches relating speech to *gesture* have also been explored in recent years, such as the analysis, modelling and transformation of speech expressivity by Grégory Beller and others in the speech research community at IRCAM (Beller, Schwarz, Hueber, & Rodet, 2005) (Beller, 2009). Relevant for my project's emphasis on *improvisation* is the music and research of pianist Sten Sandell, who from the perspective of a performer uses the act of speaking as an integral part of improvised piano performances (Sandell, 2011, 2013). Another relevant reference is the music of Peter Ablinger, especially his cycle of "Voices and Piano" pieces and his use of a mechanical piano to render speech.

These are only some of the multitude of ways that speech has been used in relation to music. Since speech and music are universal human phenomena and thus can be related to almost any aspect of human experience, a large number of interesting perspectives are possible. So even if the subject of speech and music is common, the particular *focus* of each individual approach can be quite different.

The focus of Ablinger for instance is on the *representation of reality*. He has described his use of the mechanical piano as imposing a *grid* on the sonic reality, a *phonorealistic* music as an analogy to photorealistic painting (Ablinger, n.d.). His voice pieces have the additional character of musical portraits of famous historical persons, placing the emphasis on personal idiosyncrasies, individual stories and shared cultural history. Sten Sandell on the other hand focuses on the act of speaking primarily as a *performer*, like a performing poet equating speaking with playing music as two possible outcomes of the same improvisational impulse. While Wishart has treated a particularly wide range of aspects of speech in his compositions, the focus is often on the *sound* and the *voice* as a much wider phenomenon than just speech. In the piece *"Encounters in the Republic of Heaven"*, which with its focus on everyday speech comes close to the approach of this project, there is also the overall concept of a *voice portrait* of the local community in Yorkshire.

To explain my own musical approach to this topic of speech, it is perhaps necessary to detail my musical background. Educated as a performer of the piano and Hammond organ, I have worked mostly with improvised music in jazz and contemporary genres. One direct influence for doing this project has been the experience as a performer that many of the things going on in improvised interplay are quite similar to the dynamics of spoken conversation. Not just analogous or metaphorically similar, but at times actually the same, like for instance the linguistic concept of *backchannels* –

short responses such as *"uh huh"* or *"yeah"* to affirm and acknowledge that one follows the line of thought of fellow speakers, similar to the *"comping"* figures often used for the same purpose in jazz improvisation. In a previous project of developing a personal contemporary idiom for the Hammond organ, I used the concept of staging improvised *musical dialogues* with unusual instrument combinations as a method for provoking new ideas and come to new musical conclusions, not unlike how one can reach new insights through spoken dialogues with different people.

Sound example: Musical dialogue as a method:

http://orchestraofspeech.com/wp-content/uploads/hammond-dialogues-example.mp3

Excerpt from the record *Hammond Dialogues vol 2: Twined*
(Hammond organ with string trio)
http://particularrecordings.com/shop/daniel-formo-twined/

That experience led to the idea of using actual speech as material in improvised music, to further explore this connection and see how this could affect the perception of both speech and music. Rather than using stylized forms like recited poetry or public speeches, I wanted to pursue the connection of dynamic interaction and dialogical interpretation being present in both improvised music and spontaneous conversation, and the focus of this project has therefore been on the improvised speech going on in real life conversations. An additional approach has been to explore *speech genres* as social context and musical character, and one aim has been to highlight the connection between conversation and musical improvisation as similar modes of communicative interplay. Another important concern that emerged during the project was how these topics can be integrated into an appropriate performance concept, bridging the *sound realms* of acoustic instrumental performances and virtual electric soundscapes of recorded speech.

Seen in relation to the historical and musical context described above, this represents a slightly different musical approach to speech and spoken conversation primarily as gestural improvised interplay, highlighting *improvisation* as discourse and language-like process both in music and conversation. A more detailed review of how the actual artistic results of this project relates to the different contexts will be discussed later in the chapter "Reflections on Musical Results". But before that, an account will be given of the results themselves and a description of the processes leading to these results, as well as some wider reflections on related issues, thoughts and ideas.

WORK AND RESULTS

This section gives an account of the musical outcomes of this project. Different aspects of the music making process are described – from what constitute the material, what kind of features of speech have been explored musically, to how these features have been translated into musical structures and sound, and how these structures have been organised as music. This includes approaches to improvisation, composition, repertoire development, orchestration, sound sources and performance methods. In short – all the practical artistic methods used in this project for fulfilling the aim of creating and performing improvised music based on speech.

A fairly detailed overview of the software instrument system developed for realizing this musical exploration is also included. This is provided not just as a plain technical description, but – as all the technical choices in the design are a consequence of musical decisions – as an important and integral part of the background to understand the concrete methods adopted for making music in this project.

Speech gestures as material

What exactly constitutes the *material* when using speech as a source for music? Since speech includes language and language conveys ideas, it could from a conceptual point of view be almost anything in the sphere of human experience that could be spoken of or inferred from speech – the historical context, the site, the identities, any ideas or topics of conversation, the narrative, the poetic qualities of words, the voice as instrument, or as a metaphor (*voice* concern, give voice to, vote), and so forth. Speech is of course first of all experienced physically as *sound*. Above all, highly *structured* sound, a feature it shares with music. Sound is vibrations – movement – and to produce sound one must have physical movement of some kind. Following this simple fact one could even say that *"musical experience is inseparable from the sensation of movement"* (Godøy & Leman, 2010, p. 3). This is a particularly relevant observation for the experience of improvised music, where any action and sound contributed by an improvising musician becomes a *gesture* – a potentially significant act, an utterance that not only is subject to active interpretation but depending on this also can change and actively shape the further development of the musical discourse. This function of improvised musical ideas as *gestures* is also something Njål Ølnes has emphasized in his thorough investigation into how musical signs or *gestures* are used to establish and develop musical ideas through dialogical processes into larger musical forms or *gestalts* (Ølnes, 2016).

In this regard, it is reasonable to consider the musical *gesture* as the basic musical unit in the improvised exchange of sonic ideas. In other contexts, the concept of *gesture* may have other meanings and can refer to a metaphor or just describe an action, but it is in the sense of a communicative act – *a movement to express meaning* – it is used here.

My wish to explore speech from the perspective of instrumental improvised music meant that such musical gestures – wordless musical utterances in improvised interplay, formed the musical background and approach of this endeavour. That resulted in a decision to primarily focus on the abstract prosodic qualities of speech, and not on the words and semantic content or other aspects and concepts associated with language, voice, identity, site, story, etc.

This is the main approach to speech I intuitively have adopted throughout this project, viewing spoken utterances primarily as musical gestures: non-conceptual, but with potential musical meaning. Conversation not as a discussion of ideas and concepts, but of non-verbal actions and interactions, parallel to how musical gestures make up the wordless discourse of improvised music. This approach led me first to look into the linguistic fields of prosody and conversation analysis, to learn what functions and significance this musical foundation of prosody might carry in language seen from a linguistic perspective.

Prosodic phenomena

In *"The music of everyday speech",* the linguist Ann Wennerstrom gives a thorough account of how prosodic features are actively used to structure utterances and convey information in conversations (Wennerstrom, 2001). Some examples of such features include how **strong accents** are typically used to **highlight** the most **important words**, while *high*-pitched syllables are used to mark *new* information. *On the other hand*, the modulation to a higher *"key"* is often used to signify a change of subject, *(and similarly, a lower mean pitch is used to signal supplementary comments, as if in parentheses)·*

Though linguists typically operate with *phonemes* as the lowest level of segmentation, the *syllable* is regarded the basic unit of *rhythm.* A syllable is usually based around a *voiced vowel*, and having a pitch it can be viewed as corresponding to the concept of a *note* as a musical unit. The pitches of subsequent syllable-notes form melodic contours, but their particular timing also results in a particular speech-rate that can be related to the musical concept of tempo. One interesting *rhythmic* phenomenon in this regard is how a shared semi-regular pulse or tempo is usually adapted by speakers. The adjustment to a shared pulse also extends across turns in a conversation, with speakers often timing their responses to coincide with the pulse implied by the former speaker. How syllables express this pulse can be quite different though, and languages are generally classified as belonging to either of two categories of timing: In *stress-timed* languages (e.g. Germanic languages like English, German and Norwegian) the stressed syllables are placed at regular intervals approaching an even pulse while the unstressed syllables in between are sped up or slowed down in order to match this pulse. In *syllable-timed* languages (e.g. Roman languages like French and Spanish), all syllables are timed more or less according to the underlying pulse. Interestingly enough, this timing difference has even been demonstrated in music in a study on rhythmical differences between English and French classical music (Patel & Daniele, 2003).

These and other interesting prosodic phenomena have provided the background for identifying significant features in spoken language that also could be interesting to use as a foundation for exploring speech musically. That includes the choice of the syllable as basic rhythmic and melodic unit, the use of both stressed and high-pitched accents for creating derivative rhythmic structures, and the ability to play with the inferred tempo by gradual rhythmical quantization to a grid derived from the underlying pulse. This prosodic background has consequently influenced design choices and is directly reflected in the particular functions of the software instrument system used in these explorations, as described in the chapter below detailing the instrument system. It has thus formed the foundation for the musical explorations undertaken during performance.

Speech genres

A useful way to think about the prosodic traits of speech and their wider possible musical meanings and implications has been provided in this project by the concept of *speech genres.* Other perspectives could have been chosen.

For instance, one theme often raised when discussing the musical character of speech melodies is the obvious differences in intonation between local dialects, or between typical cultural stereotypes associated with different spoken languages and cultures. Linguists on the other hand, will perhaps note how the different phonetic structures of languages make them sound completely different on the articulatory level. In addition, one can observe a wide range of wildly different speaking styles caused by all kinds of personal idiosyncrasies and physical conditions relating to age, gender, health, and other individual traits in general, like stuttering or a hoarse voice etc.

However, when listening to recorded speech from many different languages and settings, I have been struck by how *similar* different languages and people actually sound in comparable situations. The word *situation* is a clue here, as the style or genre used in a given situation conveys the social context and purpose of communication. It would probably sound strange in any language or culture to speak in a very formal tone to an infant (with the obvious exception of ritual situations like baptism). One explanation for this similarity can perhaps be found if one considers the function of spoken utterances the same way as musical gestures in improvised music – as physical *gestures* in a social situation, and following that – how vocal gestures can overlap and extend the kind of meaning conveyed by body language and physical interaction. *"Sound is touch at a distance"*, the psychologist Anne Fernald noted, observing how parents of different cultures all started to talk to their babies in a comforting tone after putting them down, to kind of keep staying in *touch* with them (Radiolab, 2006).

I think there is something fundamental about how vocal utterances can be perceived this way as *touch* – as physical sensations comprehended through the wider cognitive apparatus, which includes emotions. Sound is after all physical vibration, and the physiological foundation for this kind of sensation-based cognition means that it extends across cultures and languages as something much more universal than cultural and stylistic codes, and even across species, as domesticated animals like dogs seem to have few problems interpreting intentions from speech gestures.

Returning to the concept of speech genres, these expressive styles can be viewed as formalized expressions of both such gestural sensations and social conventions. For musical purposes, this is interesting as it points to a deeper level on which music also might function as a kind of social, universal language. This becomes even more interesting when the focus is on musical improvisation from the perspective of social interaction. Based on these ideas, the motivation to use speech genres as the main perspective from which to

explore possible social and musical meanings of prosodic gestures has first of all influenced what kind of speech I have used as subjects of study, but it has also defined the methods I have used for generating musical structures from this material. These methods are described in the following chapter.

Methods of abstraction

The improvisational focus on prosodic gestures as primary musical material described in the former section, led me to intuitively adapt a general methodology of abstraction when working musically with speech. That approach is also related to the difficulty of listening to speech without focusing on the semantic content and the narrative implied by the words – the persons speaking, the setting, the time and place, and the story unfolding. In the early days of electroacoustic music, Pierre Schaeffer proposed a mode of *"reduced listening"* to focus listening on sounds as *"objets sonores"*, sound objects completely removed from any links to their cause or source (Schaeffer, 1966). While this was a productive way of listening to sounds from a new perspective, it seems to be a particularly hard exercise when listening to speech sounds, and as I was concerned with the more abstract gestural shapes of spoken utterances, I felt it was necessary to introduce some kind of filter or *veil* to make the words less intelligible and direct the focus towards the musical features.

This resulted in methods for abstracting and stylizing conventional musical parameters like rhythm, melody, harmony etc., in addition to extracting gestural features like phrase durations, pauses, pulse etc. The structural identity of the sound as utterance can still be quite strong, and it is interesting how far speech can be abstracted this way and still be recognizable as intentional communicative gestures.

Sound example: abstracted speech gesture
http://orchestraofspeech.com/wp-content/uploads/abstract-gesture-example2.mp3

The limit of recognition seems to be when approaching static time, to the point when gestures are no longer recognized as such, and fade away perceptually as background static or ambient texture. However, I found that even with the gestural proportions intact, if the resulting music became too abstract for longer periods it would soon lose the perceived connection to speech altogether, and appear just like any other kind of abstract-sound music. The topic of speech needed to be present in the music, and I found that the most interesting things happened when I managed to strike a balance between recognition and abstraction so that the focus of perception was right on the edge between the semantic and the aesthetic. I started developing additional methods for using *unprocessed* speech recordings in such ways that their original sources, contexts and semantic contents become so fragmented and relativized that the result is perceived as abstracted or poetic sound structures, or *sound objects* in Schaeffer's parlance. In that way, the formal aspects can be kept in focus while the topic of speech is never lost.

Sound example: collage
http://orchestraofspeech.com/wp-content/uploads/collage-example.mp3

So, on the one hand it seems like speech gestures can be recognizable as communicative utterances even after radical transformation and abstraction, but also that segments of perfectly intelligible speech can be organized in such a way that they are perceived as abstract sound collages while still being recognized as speech sounds.

A number of different methods and techniques have thus been developed and used in various ways for abstraction in this project:

> *Filtering and smoothing of the changes and contours of frequencies, amplitudes, spectrum, spectral resolution, etc.*

> *Stylization, extension and ornamentation of melodic phrases and gestures: arpeggio, overlaps, pointillistic clouds/swarms, voice shadowing, choral doubling, counter-voices, rhythmical diminutions and augmentations, etc.*

> *Abstraction by spatial distribution into different layers: foreground, background, frequency register, instrumentation/orchestration, etc.*

> *Fragmentation and repetition of segments organized as collages. Like in poetry, alternative arrangement of words dissolves the narrative and places emphasis on sound qualities and formal associations.*

> *Juxtaposition of different conversations in the same speech genre shifts focus to the common features of the genre, rather than each individual story.*

> *Juxtaposition of different languages (same as above, but even more generalized)*

> *Selection by function, using only specific parts of conversation (greeting, back-channeling, laughter etc.): focus on particular types of interaction.*

Early in the project I tried to conceptualize the musical possibilities resulting from these methods of abstraction in terms of *continua* between *opposites*. For instance, between the fluid continuous quality of speech and the discrete and clearly defined pitches and attacks of the piano, or between acoustic sound and speaker-mediated sound, between vocal and instrumental etc. As a result, musical ideas often took shape of *transformations* envisioned in a multidimensional space between such opposites.

Dimension space of opposites
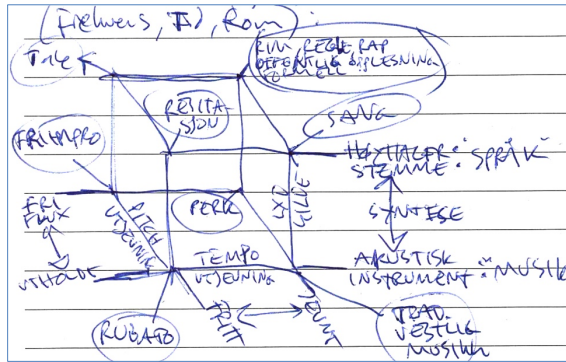
However, I found that not all ideas are best thought of in terms of such opposites, and also that this very formalistic approach had its limitations incorporating the social, interactional dimensions of the speech material. The concept of speech genres however, proved an interesting way of channeling the musical ideas generated through the methods abstraction into an overall concept of speech genre portraits. Not necessarily a faithful rendering of all the typical traits characteristic of a speech genre, but developing ideas based on certain prominent features that could serve as starting points for exploring musical implications of that genre's characteristics.

How these methods were used in practice is elaborated further in the chapter about performance methods, but before that it is necessary to look at the particular instrument system developed for this purpose, to understand how these methods were implemented in the actual tools used for performance.

Software instrument system

This section presents an overview of the software instrument system that has been developed as a tool in this project, constituting the main method for the performative exploration of musical features of speech through improvisation. Through the affordances of this instrument, much about the music can also be explained, and implicit in this technical description is therefore also an account of the musical choices, methods and possibilities arrived at in this project. The instrument system is a solution to the particular musical challenges posed by this project, and provides the practical foundation and means for the musical exploration. Its final design is the result of a long chain of aesthetical choices, as well as changes made in direct response to shortcomings identified through musical explorations.

Digital musical instrument context

The system can be classified as what is generally known as a *Digital Musical Instrument* (DMI). In the context of such digital musical instruments, several existing instruments have been developed based on voice and speech, many of which have been concerned with direct voice synthesis in a traditional gesture-to-sound instrument paradigm. Recent contributions include systems that control various kinds of voice and speech synthesis techniques by means of hand gestures, either on a modified accordion (Cook & Lieder, 2000), with gloves (S.S. Fels & Hinton, 1998), stylus (Delalez & Alessandro, 2017), guitar (Astrinaki, D'Alessandro, Reboursière, Moinet, & Dutoit, 2013), microtonal keyboards (Feugère, D'Alessandro, Doval, & Perrotin, 2017), or using motion capture to track hand gestures in space (Beller, 2014; Beller & Aperghis, 2011). Other systems have taken the opposite approach, using voice as input to control other kinds of synthesizers (Fasciani, 2014; Janer, 2008). The present system also use speech as input, and has also the ability to produce synthesized speech output, but unlike the traditional instrument paradigm where expressive performer gestures are used directly to control sound production, this system could rather be seen as a kind of real-time compositional tool that can be used to analyze and extract, transform and arrange several layers of rhythmic, melodic, harmonic and other musical features from a given speech source. As a concept, this is not unlike transcribing and scoring speech melodies on paper or electronically processing speech and composing an electroacoustic piece, but to do this through the dialogical process of improvisation it needed to be a real-time interactive *instrument-like* system.

Functional requirements

Few specific functional requirements for this instrument were clear from the start. Following the decision to focus on real-life conversations and speech genres rather than spoken performances on stage, the system had to be based on *recordings* of speech and not primarily live speech input. In addition, and as described above, to be used for *improvisation* it needed to work in real time and be playable like an instrument. It was going to be based on analyses of speech and have some means for generating sound based on these analyses,

but other than that I could not really foresee exactly what features would turn out to be interesting and useful beforehand. In that sense, the development of this digital musical instrument was different from building a system where the functional requirements are known in advance. The instrument development process became an integral part of the musical exploration of different ideas and approaches to speech as source material for music making. This changing and evolving nature is common for many such complex digital performance instruments where new ideas are developed and tested continuously (Trifonova, Brandtsegg, & Jaccheri, 2008).

Platform/Software Environment
Regarding the actual programming, the software instrument has been developed for use on conventional personal computers in the popular Max graphical programming environment, relying on some very useful external Max libraries of analysis and processing tools developed at the *Institut de Recherche et de Coordination Acoustique/Musique* (IRCAM): at first the FTM *("Faster Than Music")* library (Schnell, Borghesi, & Schwarz, 2005) and later also the MuBu *("multi-buffer")* external library (Schnell, Röbel, Schwarz, Peeters, & Borghesi, 2009).
For more information on the various signal processing techniques used, see the technical background in the appendix, and for detailed descriptions of the actual implementation of these techniques in Max, please refer to the comments in the actual Max patches available from the *"downloads"* page.

A fairly detailed technical description of the software instrument system and its different parts is presented below. Again, the purpose of this is not only to show how it works from a technical perspective, but as much to show the affordances and musical choices that has been developed through several rounds of design, practice and performance, and which in the end constitutes the practical foundation for the artistic outcomes of this project.

System overview
A digital *composed instrument* such as this can be viewed as composed of many processes that can be classified in terms of their function, such as *analysis, transformation,* and *synthesis* (Schnell & Battier, 2002). This is a useful division that has served as a guideline for my modular approach to building a playable digital instrument. In this instrument, *speech recordings* fulfil the function of *performer gestures* usually used input for an instrument. The performer in turn controls both the selection and playback of the recorded speech segments, as well as how these recordings are translated into musical structures. A helpful concept in this regard is the metaphor of *"orchestration"*, describing the process of arranging and distributing the musical structures extracted from the speech material.

System demonstration #1, system overview:

Video: https://www.youtube.com/embed/MLizm3Di7u4

In the diagram below, a schematic overview of the whole system shows the different parts and their function within the system.
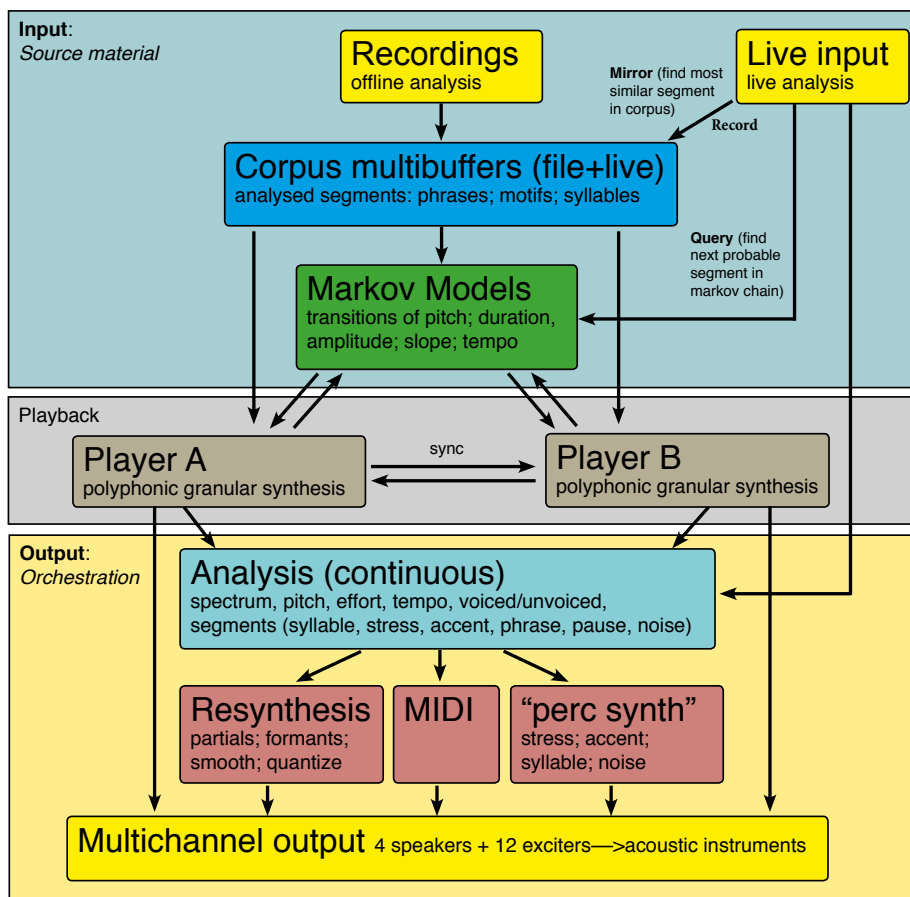
System diagram

Input: speech recordings

At the core of the system is a set of buffers for loading or recording collections *(corpora)* of speech recordings. The recordings can be analysed and automatically segmented into *syllables* (or more precisely: *vowels*), stressed syllable *motifs* and breath length *phrases.* For all these segments, five basic prosodic/musical descriptors are calculated: *mean pitch, segment duration, amplitude, pitch slope* and *tempo*. Collections of analysed recordings can be saved as *corpora* that can be recalled on the fly during performance, eliminating the need to perform such segmentation every time a collection of recordings is needed.

Machine Learning

Organizing recordings as collections or *corpora* also allows the possibility of using statistical analyses in order to "learn" the characteristic patterns in a corpus, known as machine learning. The system includes a simple technique for such machine learning known as *Markov chains*. In short, this technique analyses the likelihood of any transitions between different stages in a chain of events, and based on that model one can generate new chains that are statistically as probable, and therefore display some of the same overall structural characteristics without using the exact same order of individual segments. In this system, when a corpus is loaded, its segment's descriptors are used to generate such Markov models that describe the likelihood of transitions between any states in the corpus: between different pitches, durations, amplitudes, pitch slopes and tempi. How these statistical models then can be used to generate alternative sequences is described further below.

Playback

The recordings segments can be played back by two identical *player* devices, which can either be operated separately, or synchronized to create unison or shadowing effects. Playback is technically handled as *pitch synchronous granular synthesis* (Roads 1996, p.174), which means that it reproduces *"sound grains"* of one wave period at a time, at the frequency defined as the fundamental pitch. As reference for this fundamental frequency it uses a pitch track generated and stored in the buffer as part of the initial offline analysis and segmentation processing. The pitch-synchronous granular synthesis technique allows decoupling of playback speed and pitch, as the grain *rate* can be changed regardless of the original fundamental frequency indicated by this pitch track. This enables playback *speed* transformations like *time stretching* and *compressing,* independently from *frequency* transformations like pitch and spectrum transpositions. It also allows independent changes to the timing of segment onsets, making changes to the overall *speech rate* or *tempo* of syllables possible.

System demonstration #2, playback device

Video: https://www.youtube.com/embed/y7MKZHhicv0
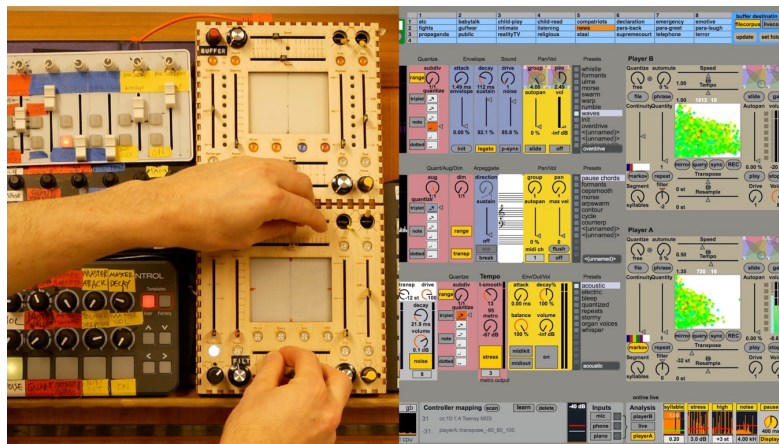
Playback control

As a result of the database organisation which helps keeping track of all the segment and their boundaries and descriptors, segments can be sorted and navigated in a graphical user interface based on their sonic features, displayed in a scatter plot as a function of mean pitch (y) tempo/duration (x) and amplitude (colour). Number of *voices* available for polyphonic playback can be set from one to eight, and a *quantity* measure allows up to 10 of the most similar segments to be triggered at the same time as a cluster or as a sequence (depending on the number of available voices), optionally *repeating* in a random order. If, however, the *Markov chain mode* is enabled, each triggered segment will query the Markov model for probable transitions for each of the segment's descriptors. These probable transitions are in turn used to find and trigger the closest matching segment in the corpus (this approximation enables output even when no exact matching segment is found, thus avoiding the dead ends typically a problem when generating new Markov chains from a limited dataset). In effect, this generates Markov chains of alternative but statistically probable sequences of segments. The degree of freedom for these sequences can further be limited with an imposed measure of *continuity*, influencing which descriptors are prioritized when searching for the closest matching segment. Ranging from *0% continuity* where all musical descriptors are weighted equally, through *50%* where only pitch and duration are weighted, towards a state at *75%* where the file index is weighted as most important thus limiting segment choice to only those of the same recording file. When the *continuity* measure is set fully to 100%, only the *file* and *segment* indices are weighted when searching for the next segment, forcing a perfectly continuous playback of segments in their original order. This way, just by changing the continuity measure, one can move gradually from a linear sequence of segments to a probabilistic sequence and vice versa, as demonstrated from 1:17 in the video demonstration above.

Continuous analysis and synthesis devices

The sound output from the playback devices can be fed into an online analysis stage, performing real-time segmentation and analyses of spectrum, fundamental pitch, vocal effort and tempo. Due to the modular layout, the resulting stream of continuous analysis data can then be used collectively by a number of synthesis devices, making it possible to create many different and simultaneous layers of orchestration based on the same speech input. In the present setup that includes an additive synthesizer, a MIDI-note output device and a hybrid audio/midi percussive device.

The modular layout also allows these devices to have separate but similar controls for the transformation, synthesis and output of the same incoming continuous analysis stream, as shown in the diagram and video demonstrations below.

Analysis (continuous)
Spectrum (fft & lpc), pitch (f0), vocal effort (harmonic slope), tempo (speech rate).
Segmentation: syllables, stressed syllables, high pitch accents, phrases, pauses, noisy consonants

Resynthesis	MIDI	Percussive
Segment selection	Segment selection	Simultaneous: stress, accents, syllables and consonants
Transformation - Spectrum (FFT or LPC) - Freq/amp smoothing - Cepstral smoothing - Partials selection - Transposition/scaling	**Transformation** - Spectrum (FFT or LPC) - Freq/amp smoothing - Cepstral smoothing - Partials selection - Transposition/scaling	**Transformation** - Pitch smoothing - Transposition - Tone/noise mix - Filter
- Rhythmic quantisation - Rhythmic subdivisions	- Rhythmic quantisation - Diminutions/Augmentations	- Rhythmic quantisation - Rhythmic subdivisions - Tempo smoothing
Synthesis - Spectral env additive synthesis - Fixed or pitch-synchronous OLA - Mix partials/noise spectrum - Envelope controls - Vocal effort driven overdrive effect	**MIDI note generation** - Arpeggiate/cycle through partials - Sustain notes	**Synthesis & MIDI output** - Simple subtractive synthesis

MIDI instruments
Backend of MIDI synthesizers

Multichannel output 4 speakers + 12 exciters/acoustic instruments

System demonstration #3, synthesis device:
Video: https://www.youtube.com/embed/0ZwaGnOvXxY

System demonstration #4, MIDI device:
Video: https://www.youtube.com/embed/R_eXMJY9mvU

System demonstration #5, percussive device:
Video: https://www.youtube.com/embed/g0z0PBiZfXw

Sound output

The sound of all playback and synthesis devices can either be mixed to a stereo output or routed to multiple output channels. In the current setup, 16 channels are used, connected to a set of stereo loudspeakers, two small low fidelity radios, and 12 transducers attached to acoustic instruments resulting in a hybrid *"electric/acoustic"* sound. This unorthodox multichannel setup is reflected in a *panning* interface where the sound position can be controlled, not as an exact *"panorama"* position in the room, but sent to the different speakers and instruments in a kind of direct *orchestration* of the sound output. In line with the general metaphor of orchestration, this *"orchestra"* of loudspeakers and acoustic instrument-speakers is organised into *sections* of four different instrument classes: *speakers, drums, strings* and *cymbals*, with four instruments/speakers in each group. The physical interface consists of one joystick for panning *between* the instrument sections, and another for panning *within* the groups. Thus, with these two controllers, one can move the sound gradually between any of the 16 outputs of the system (shown at 2:54 in the video demonstration of the playback device above).

In addition to these controllers, an *"autopan"* function can change the pan position automatically at segment onsets, either as a range within the same instrument group or across all outputs.

MIDI channel routing is handled in a similar way, sending MIDI notes to a backend of 16 software instruments (for convenience hosted in the popular music production software Ableton Live), but the sound output from these software instruments is routed to the same multichannel outputs of transducers and speakers, resulting in a coherent and intuitive layout for controlling multichannel pan position for all audio and MIDI devices in the system.

Live sound input

In addition to the signal flow described above, it is also possible to feed live audio *input* into the system. This input can be used in three very different ways. Sound can be *recorded* in one of the buffers and instantly be available for playback. It can be analysed and used to *trigger playback* of segments that are already in the buffers, either by finding the most similar segment (*mirror* mode), or by triggering a *response* by querying the Markov model for a statistically probable continuation of the input segment (*query* mode). Finally, audio input can also be routed directly into the online analysis stage for instant *live orchestration* by any of the synthesis devices.

Control and automation

The whole system can be controlled with a conventional graphical user interface (GUI) with dials, sliders and switches. Physical MIDI controllers can be connected and mapped to these control parameters by an automatic learning function. The same centralized parameter subsystem is used for storing and recalling *presets* both globally and for each "device" in the system.

Through accessing this subsystem, it has also been possible to enable complete automation of any parameter in the system by scripting cue files, a feature that was implemented in order to use the system in a complete self-playing mode for use as a sound installation.

Technical affordances

How the particular features of this instrument system have been used in music-making is described further in a chapter below about performance methods. As stated above, the possibilities of the system are based on a combination of linguistic phenomena, musical ideas and instrumental needs identified through testing and performance. But in addition to this, the technical development of an instrument has in itself contributed to the musical ideas, generated from the encounter with certain technical possibilities provided by the system. The technical development process often involved the creative use (and abuse) of different techniques just to explore how they would sound and if something interesting would happen. One example of this can be provided by the result of heavy *cepstral smoothing*. Reducing the frequency spectrum to just a few wide frequency bands, it can result in an apparently formalistic play between high and low registers, quite abstract but nevertheless conveying some dynamic traces of the original speech gestures.

Sound example: cepstral smoothing:
http://orchestraofspeech.com/wp-content/uploads/cepsmooth.mp4

Another example of such technically derived ideas is the possibility to express vowel formants as musical *chords*. Formants are the characteristic spectral peaks that define vowels, and are not usually expressed or perceived as individual pitches, but act more like filter frequencies that shape the spectrum of a source sound. But if synthesizing the formant frequencies directly, it is fascinating how just a few formant frequencies are enough to render speech intelligible, even with no fundamental pitch present. Using the Linear Predictive technique to track formants and output the resulting frequencies as chords can result in an interesting abstraction of this "almost intelligible" spectral shape of speech.
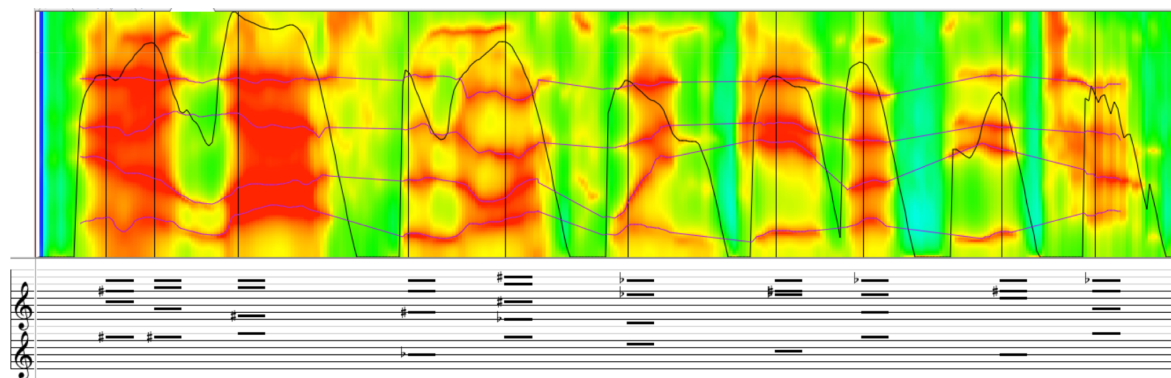
A mel-frequency spectrum showing the four lowest formants and the corresponding musical chords transcribed from the sound example below.

Sound example of iterative transformations of formants into abstracted chord sequences:

http://orchestraofspeech.com/wp-content/uploads/formant-transformation-example.mp3

In addition to these technologically inspired ideas, purely musical ideas have of course also driven development of the instrument system and the music it can produce. The motivation was always to make *music* and not didactic demonstrations, so at some point it is absolutely necessary that intuitive associations and connotations take over in the further process of making music from the speech sources. This can for example be the creation of complex polyrhythmic layers, not really a part of speech at all, but inspired by speech tempo variations or based on some strong rhythmical feature typical of a speech genre and therefore still related to this material. Or it can be stretching out vowels and layering several voices to create dense microtonal choral textures, only using the vowel timbres from speech to pursue an otherwise abstract musical idea:

http://orchestraofspeech.com/wp-content/uploads/freezechoir-example.mp3

In this way, both prosodic features, speech genres, technical possibilities and original musical ideas have contributed in shaped the affordances of the instrument system and served as starting points for music making, presenting interesting phenomena that generate new sonic ideas, explored further through musical discourses with the speech material.

A more thorough review of the system's possibilities and limitations is provided below in the chapter entitled "Reflections on the design, development and performance of a new instrument". But before that, we must turn to how this software instrument evolved into a coherent performance setup through developing a concept of *sound* and *orchestration.*

Sound and Orchestration

As it has been established in earlier chapters, it is not the actual sound of the *voice* that has served as the primary musical material and focus of this project, but rather the underlying abstract gestural structures of speech prosody. But music has to *sound,* and initially I did not have any particular ideas about *how* my music based on these abstract gestures should actually sound. This part describes how the digital instrument system evolved into a coherent performance concept that eventually became the *Orchestra of Speech.*

What I experienced during early trials was that the choice of instruments and sound production tools can introduce very strong stylistic connotations, each with a very particular history and certain genre conventions. The use of commercial synthesizers for example, quickly frames the music as related to popular music. Improvising on acoustic instruments can easily be associated with genres like jazz, while certain synthesis techniques typical of computer music sounds very much like, well, typical computer music. I was searching for a way to avoid such clear stylistic markers. I was also looking for a way to relate the use of software tools and speech recordings to my practice as a keyboard performer, something that could point to the underlying theme of speech and music as well. The solution I found was to avoid the pure loudspeaker format normally used with computer-made electronic music, and instead bring the electronic recorded and synthesized sounds into the acoustic sound realm by using small contact loudspeakers (known as transducers or exciters), playing back plain or transformed speech sounds through physical objects. Such transducers are really just loudspeakers without the usual paper cone membrane attached, instead with its voice coil fixed directly onto another surface which then acts as the transducer's membrane or resonator. When attached to an acoustic instrument, which is usually designed to have good resonance for its typical range of frequencies, the result is a distinctively acoustic sound quality with associations to instruments and music but still with enough fidelity to reveal the speech source if needed.

Transducer testing video: https://www.youtube.com/embed/Hjq7xvDTCjI

This blend of electronic and acoustic sound worked so well for my purpose that I developed the idea further into a complete performance concept that includes a whole array of such instrument-mounted transducers. The instruments that worked best were instruments with large resonating bodies or chambers that also in ordinary use depend on clear excitation signals to operate, such as drums, cymbals and stringed instruments with a music box (e.g. guitar) or sound board (piano). Wind instruments need a stable air pressure to sound, and I found that the resonance of the tube and bell was not enough to amplify and colour the sound from an exciter when mounted on the body or reed of a wind instrument. Alternatively, other resonant objects like scrap metal or buckets or other household objects could also have been used, but part of the idea was to create a musical framing for the recorded and synthesized speech sounds, so an ensemble of acoustic instruments was assembled and connected to a multichannel setup of exciters to form an *"orchestra"* of electro-acoustic loudspeaker-instruments.

In line with the compositional approach of analysing and arranging different features of speech into layers and using the overall metaphor of *orchestration*, I called this performance concept the ***Orchestra of Speech***.

Hybrid electroacoustic orchestra sound check video:
https://www.youtube.com/embed/hx0MrzyLWZc

By bringing the digital instrument into the sound realm of acoustic instruments this way, a natural connection was established between these seemingly separate sound spaces, and it became possible to explore this whole extended sonic world ranging from recorded reality and complex synthesized digital and electronic sounds to the delicate details and physicality of real acoustic instruments, all within the format of a single integrated solo performance concept.

The use of acoustic instrument also provided a natural way to include the piano in performances. This was something that I from the outset did not even consider, but as this hybrid sound world was established I found that I could actually approach the music/speech continuum from both ends, playing speech-like phrases on the piano and creating music-like structures with the software instrument system at the same time, and even engage in meta-dialogues between speech and music.

In its current form, the hybrid electric-acoustic *Orchestra of Speech* setup have been organised as four sections of different instrument categories: strings, drums, cymbals as well as conventional loudspeakers, with four members in each group. The use of instrument groups makes it easy to orchestrate musical ideas using one instrument group at a time or combining different instrument groups. In combination with the conventional loudspeakers, this physical electroacoustic orchestra blurs the line between the electric and acoustic, between voice and instrument, and between virtual and real soundscapes. In this way, *sound* emerged as another important musical theme this project: how sound source and sound quality affects the framing, perception and meaning of sound. This is discussed further below in the chapter with reflections on perception. Before that, an account is given on how this instrument and performance concept was put to use, developing strategies for playing, improvising and making music with the ideas, approaches and materials presented above.

Performance methods

One initial aim of this project was that it should result in a *repertoire for improvisation,* as a foundation for music making. Not as a set of fixed clichés that can readily be produced when needed, but rather as a vocabulary of a language or coordinates in a musical space of possibilities to be explored. *Repertoire* in this context refers to a whole range of internalized know-how about how to use these speech sources to make music.

For the improviser, the process of developing and internalizing such knowledge includes the aspect of *performance,* both practically mastering the instrument, but also musically knowing one's ways around the material at hand at a moment's notice. Usually one can rely on an accumulated reserve of experience as a performer and improviser, but I found that for this project I was literally starting from scratch. Not only was I developing new approaches to making improvised music based on speech analyses, I was also building a new instrument and learning to play it at the same time. In retrospect, it is not so strange then, that it took much more time than I had anticipated reaching the point where I could improvise and perform on a sufficient level. To get there I went through several rounds of exploring, learning, developing, rehearsing, performing, and revising. On a small scale these rounds typically took the form of a repeating creative loop:

||: Idea -> testing -> evaluation -> revision :||

On a larger scale, over the course of the whole project, the process of musical development through performance can be summarised with these stages:

Studies

During early instrument development, I made a series of simple *studies* to explore basic musical possibilities, usually focusing on only one element at a time. Such limited studies have been suggested as a useful method for developing and discussing isolated elements in artistic research (Michael Schwab & Borgdorff, 2014). This early into the project it was not so much about discussing, but more about just identifying basic material. The results are not terribly interesting to listen to as music, but making these studies was an important first step in the process of defining and exploring material for later use.

Sound example:
Organ study, exploring ways to use different musical features from the same speech source to arrange multiple polyphonic layers, in this case three parts for different registers of a MIDI-controlled pipe organ, arranged and improvised in real time:

http://orchestraofspeech.com/wp-content/uploads/2015/02/Organ-Chatter.mp3

Improvisation as method for discovery

The next step was to play more freely with different elements in open-ended improvisations, which I recorded and evaluated. These improvisations were often of long durations with little attention paid to form or development, but served as a productive method for discovering interesting musical ideas as well as identifying general shortcomings of the instrument system.

Sound example: improvisation
http://orchestraofspeech.com/wp-content/uploads/improvisation23.mp3

Through systematically working through such open-ended improvisations, I was exploring new sonic and musical possibilities as well as different formal ideas, thereby gradually building and internalizing a vocabulary for improvising with this material and instrumental setup.

Repertoire development

The accumulated results from these improvised explorations gradually took the form of a repertoire of *musical possibilities*, like amassing building blocks or mapping out elements of a formal language. To better understand what this repertoire consisted of and what the relations between the different musical ideas were, I worked on different ways of how to categorize these ideas conceptually. Early on I had tried to define the musical ideas in this project in terms of continua between opposites, constituting dimensions in a space that could be explored and navigated. Although this was appropriate for some ideas, not all was best thought of as spatial dimensions. Working on other ways to organise the repertoire conceptually, I found that one productive approach was to view ideas in relation to a set of different categories of significance relating to the material, instrumentation or musical properties, serving as a framework for further musical exploration with these ideas:

Speech genre

*Perception of time (narrative, interactive, static, cyclic)*

Sound source / orchestration (voice, loudspeaker, instrument, acoustic, synthetic)

Formal organization: structured thoughts (speech) or structured sounds (music)

Traditional musical parameters: layers (polyphony), instrumentation, density, register, tempo, dynamics, continuity, etc.

Composing for improvisation

Starting to get a grip on what my materials were, I was still not able to maintain the control or overview needed to improvise freely with this repertoire. As a step towards greater improvisational control I composed sketches with graphic notation. This way I could work on formal issues on how to combine these elements and notions into coherent forms. The many ideas I had developed earlier through improvisation were then organised within the framework of these main categories of significance, and from that I could compose the broader form of a piece of music while still leaving the details to be improvised.
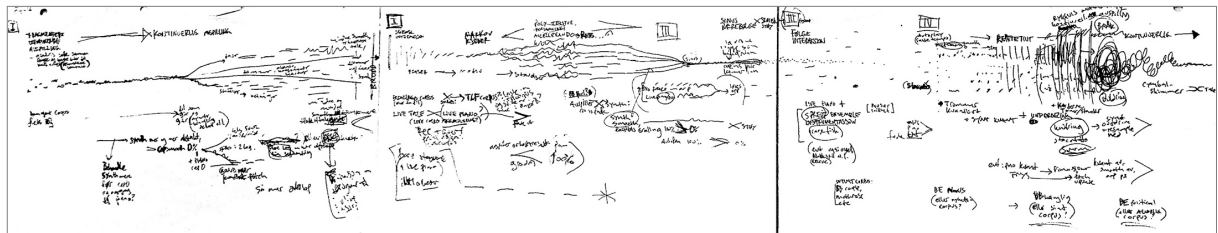
Graphic composition

The composition sketched above consists of five main parts which explores musical associations of different speech genres, different aspects of time perception, as well as different musical ideas, orchestrations and soundscapes:

Part 1: LINES

Time concept: forward motion through a continuous stream of speech.

Source material / speech genre: social interaction in reality-TV.

Soundscape: Speaker mediated speech is abstracted and moved to acoustic instruments, different lines of rhythmical layers (syllables, accents, stress, phrases) orchestrated out to different instruments, more and more stylized like musical figures (smoothing tempo, pitch and adding diminutions and augmentations), before returning to the free flow of continuous speech.

Part 2: DIALOGUE

Time concept: accentuated present, interaction between performer and machine, music and speech.

Source material / speech genre: paralanguage, telephone conversations.

Soundscape: Live acoustic piano against speaker mediated speech. Develops from interactive dialogue through musical association to polyphonic texture: a crowd where each individual voice disappears. The sound of the crowd is then drawn out and abstracted into a static time chordal texture.

Part 3: SPACE

Time concept: still, static, stationary, open space rather than measured time.

Source material / speech genre: intimate thoughtful conversations.

Soundscape: Sparsely orchestrated parts across the room, blending both speech recordings, synthesized sound and acoustic instrument in a wide-open soundscape.

Part 4: RHYTHM

Time concept: cyclic, repetitive, ever present.

Source material / speech genre: representative, non-personal radio broadcasts.

Soundscape: acoustic drums. Isolated similar syllables make a rhythmic web accompanied by a metronome measuring the mean pulse. Switching between repetitions and narration and gradually adding more and more subdivisions it develops and culminates into a dense distorted electroacoustic soundscape, a raw mass of sound that stripped of any resemblance with speech still conveys some overall phrasal gestural shapes.

Part 5: PAUSES

Time concept: the spaces and gaps between utterances, between lines.

Source material / speech genres: dignified official speech from a head of state to its citizens, characterized by the downwards melodic finality of each statement and the long pauses between them.

Soundscape: spatial arpeggii in bells and acoustic cymbals fills in and sets music to the pauses, articulating the gaps between utterances and their imprint on the aural memory. The sound of a gong and the dark drone background adds to the ritual character of this speech genre.

Performance logistics

Preparing for improvisation by way of composition also made it much easier to rehearse the logistics of *performing,* and really concentrate on the practical aspects of controlling the new instrument. This is of great importance, as it is easy to underestimate the amount of motor memory training and neural rewiring necessary to properly internalise control over an instrument. The first public performances were of this kind with an overall composed form.

Solo performance at Harpefoss Poetry Festival September 2016:

Video: https://www.youtube.com/embed/y4astF6RCjo

Improvisation and interaction

It was not until after these developments that I began to reach a level of control and internalisation where I could actually keep up with the ideas and opportunities of the moment and really start to improvise freely with this material and instrument setup. The main performance concept was still the solo improvisation setup, but now I could also try out experiments by bringing other musicians into the mix, exploring how this concept could work in interaction with other performers. One such experiment was a completely improvised duo set with vocal and electronics performer Tone Åse.

Duo improvisation with Tone Åse at Dokkhuset February 2017:

Video: https://www.youtube.com/embed/vcxHy5Qh0_w

Another such experiment was an ensemble performance with a semi-composed piece for strings and percussion, where one issue was how to combine the improvised nature of my approach with the need for composing a piece and writing parts for the musicians. One option was to transcribe a piece in detail based on my current material and methods, arranged for that particular group of musicians. But then the reactive and interactive nature of the improvised interplay would disappear. Instead I wrote a score with general directions for improvisation within set limits. This is also how I usually conceptualize musical ideas when working with improvised music, also drawing on the tradition of indeterminate notation developed by composers like John Cage, Christian Wolff, Earle Brown, Pauline Olivieros a.o.

The piece has five parts, each about 3 minutes long each notated on one page:

Conversation game   *Fight alarm*   *Intimate space*   *News rhythms*   *Chers compatriotes*

Performance of piece for strings and percussion, by musicians from the *Trondheim Sinfonietta*, performed at Kunsthall Trondheim, October 2017:

Video: https://www.youtube.com/embed/Ek3nyEl_Gec

Despite these experiments with different ways of approaching ensemble performances, it is the solo performance setup that has remained the main performance format throughout this long process of developing both a new way of making music, a new instrument and a new performance concept at the same time. Nevertheless, for the final presentation of artistic results, I also chose to include two alternative versions in addition to the solo format – the instrument setup as a self-playing interactive sound installation, and employing the instrument in a free improvised ensemble performance with six additional musicians. These experiments provide alternative uses of the system and point forward to possible future explorations of the ideas developed in this project. These performances are documented in the following chapter on final artistic results.

Final Artistic Results

The final presentation of artistic results from this project took place in Verkstedhallen in Trondheim on November 15, 2017. This event featured the *"Orchestra of Speech"* performance concept in three different versions, each of which invites different modes of listening and presents different approaches to form, time, sound and interaction. I arranged the room with a grand piano in the centre, loudspeakers in the corners, and all the transducer-instruments hanging from the ceiling above head height, so that the audience could move around during the installation and later take a seat wherever comfortable during the performances. During the ensemble performance, the additional musicians were placed along the walls, surrounding the audience with acoustic sound and in effect making them part of the event as a social situation as well.

Sound Installation

The first part of this concert featured the *"orchestra of speech"*-setup as a self-playing installation. The solo performance has really been the main format throughout this project, but as an experiment I wanted to try out a sound installation version of this concept in order to shift the focus away from the expectation of a certain narrative and dramaturgy that comes into play whenever a performer enters the stage. An installation allows listeners to move around and structure their own experience, both in *time*, as well as in *space* in relation to the orchestra as an entity in the room.

The musical content and behavior of the installation is controlled by scripting *cue-lists*, handling all aspects of playback and orchestration otherwise performed live. The micro-compositions presented by these scripts revolve around one musical idea at a time, usually for about five to eight minutes before changing.

Like the solo performance setup, the installation also features a way to interact with the speech-music material. An analogue telephone set connected to the system rings from time to time, inviting members of the audience to pick up the phone and interact with the orchestra by speaking to it. The phone conversation as a social phenomenon is special in that it is the only commonplace experience where we communicate and interpret solely by voice. As such it seemed interesting to be able to interact with the otherwise disembodied recorded voices through the familiar interface of an old telephone, provoking nonsensical but somehow musically probable responses from the system.

Even though the installation format invites other readings than the conventional concert situation, I still think of this primarily as *music* and not as conceptual sound art. The use of a telephone in particular might seem like a way of staging a very theatrical situation. My intention however, was primarily to facilitate the curious experience of interacting with this speech/music-machine through one's voice, with speech recordings played back into the telephone receiver and simultaneously orchestrated as musical phrases on the speaker-instrument orchestra. This experience can also be related to our future prospects (and fears) of having to communicate with machines that apparently understand the non-verbal underlying emotional communicative layers of speech as well.

The use of a telephone set also creates an interesting situation for the remaining spectators who are unable to hear the voice in the telephone receiver. From their point of view, the exchange looks almost like a dialogue between the speaker and the orchestra, as a kind of meta-dialogue between speech and music with the telephone acting as the literal line of communication between these phenomena.

Sound installation video (1 hour reduced to 8 min):
https://www.youtube.com/embed/PQRNeF1JbbM

Solo performance

The next part of the performance presented the *"orchestra of speech"* system as a solo instrument. This has been the main working performance format throughout the project development, and the one that most clearly reflects the ideas of the project. In the solo setup, the software system is complemented with a piano, used both as a musical counterpart in the overall orchestral texture but also providing musical input for the system. Used this way, the piano can trigger responses and act as a dialogical partner for the artificial orchestra, with which I in turn can explore the whole continuum between the discrete stylized musical structures of the piano and the continuous flowing gestures of speech.

https://player.vimeo.com/video/250236499
Video: Alice Winnberg

With this performance, I feel that I have reached the level of control where I succeed in integrating my practice as a performer and improviser with the new ideas developed throughout this project. I think the performance sums up the project well and manages to convey the topics that have been central to the project: the connection between conversation and improvisation; the musical character of speech genres; soundscape and sound sources as meaningful; and the idea of interrelations between speech and music in general.

Ensemble improvisation

The last part of this concert was an ensemble improvisation. Like the installation, this version was mainly included as an experiment, exploring how this performance concept developed as a solo format might work in a musical discourse together with other improvising musicians. In this experiment, I wanted to bring the speech material back into a dialogical setting that could perhaps also draw attention to musical ensemble improvisation as a fundamentally *social* situation. The added ensemble consisted of three vocalists, all with optional microphones and one with additional electronics, and three acoustic instruments with no amplification – acoustic drums, cello and contrabass. This instrumentation was chosen to mirror the project's existing themes of acoustic/electronic, vocal/instrumental and speech/music.

https://player.vimeo.com/video/250361107

Video: *Alice Winnberg.* Performers: *Daniel Formo (piano and electronics); Heidi Skjerve (vocals); Tone Åse (vocals and electronics); Sissel Vera Pettersen (vocals); Marianne Baudouin Lie (cello); Michael Francis Duch (bass); Ola Djupvik (drums).*

Even though the solo format has been the main performance concept throughout this project, I think it was interesting to see what happened when expanded into a larger ensemble. Both conceptually, as a dialogical parallel to the interplay going on in the recorded conversations, but also musically how this changes the experience and the resulting music.

In this improvised ensemble performance, the focus clearly shifted towards a more diverse improvisational discourse featuring more musical subjects and less overall unity. Some of the musicians reported that it was very interesting to engage in new kinds of interplay, not only with the other performers but also with the voices and sounds embodied in the speaker-instrument hybrids of the orchestra system. Another interesting change was the increased effect of blending and merging of sounds and soundscapes, blurring the line between electric, acoustic, virtual and physical sound, to the point where it sometimes was hard to discern where sounds were coming from. Most of these musicians had previously been involved in my project in one way or another, either in performances or small studio experiments, but this was the first time gathered as a larger ensemble. The resulting lack of a common reference is perhaps why this performance had an overall tentative, searching character, with less clear direction and initiative than the solo performance. That is only natural for a first performance, as any encounter in unchartered territory. Nevertheless, I think the experiment clearly shows that it is possible to use the performance concept developed in this project outside its original conception as a solo format, and that it opens up a great field of opportunities to be explored further in the future.

REFLECTIONS

The following chapters presents a series of wider reflections on topics related to the work and results of this project, ranging from the concrete process of developing and performing with a new instrument and how to relate to conceptual implications of using speech recordings, to more philosophical thoughts on perception, form and language, and concluding with some thoughts on the musical results. Some of these reflections are rather essayistic in form and based on what Borgdorff describes as the kind of *unfinished thinking* that is a part of artistic practice – a kind of reflection that perhaps has more in common with an *un-academic* speculative philosophy than with academic rigour (Borgdorff, 2012, p. 71). I nevertheless think it is important to include these subjects as they form an integral backdrop of thoughts and ideas for understanding and appreciating the creative work carried out in this project.

Design and Development

This chapter offers some reflections on issues relating to the design, development and performance of the *Orchestra of Speech* software instrument system and performance concept developed in this project, as well as a discussion of its possibilities and limitations.

The system works as a playable musical instrument, but can also be thought of as an instrument in a wider sense of the word, like a prism or microscope that can zoom in and show different musical structures that is part of speech. It is not meant to represent the kind of ground-breaking technical innovation as, say, a new cutting-edge synthesis technique or instrument control paradigm, but must rather be viewed as a case of how to put together – how to *compose* – a selection of techniques in order to realise a particular artistic vision. The innovation lies in the personal combination of ideas and system, and in the musical outcomes made possible by this particular combination.

Design issues

As described earlier in the technical overview, the system was essentially developed from scratch with few initial specifications given. The features and the ideas about the instrument changed repeatedly in response to testing and performances, making any functional requirements a moving target. Nevertheless, over time it stabilised into a set of features that seems to fulfil the most important artistic needs identified during development. That includes functions to create several different polyphonic layers from the same speech source, the functions for indeterminacy and generating alternative combinations and sequences of segments, and the ability to use live sound input to provoke responses and engage in interplay. Though developed into a fully functional and stable performance system, a first design like this will still have the character of a prototype. In future work, this system might very well be developed further, or perhaps a new system can be built from scratch based on the experiences in this work.

Interface design and control intimacy

Early in the development of the software instrument, issues regarding performance was often of a technical or practical character – how to integrate the different functions into a manageable whole, and how to control the instrument through an appropriate interface. I knew that I needed a very intuitive way of controlling this complex instrument I was developing in order to achieve any meaningful performance results. The problem of control is a common issue when designing so-called digital musical instruments (DMI's), and dependent on what Moore first conceptualized as *control intimacy* (Moore, 1988). The conventional personal computer interfaces of screens, keyboards and pointer devices are far too narrow, slow to use and single-task oriented to be used for interacting with a complex digital musical system. Even when using external controllers with physical knobs and switches, I find that there can be an unsatisfactory spatial and cognitive split of attention between using the physical controllers for input but getting visual feedback

from a screen. In addition, one of the most important concerns with regard to control intimacy is to develop the most appropriate *mappings* between user gestures and instrument response. This is central to achieve the kind of *embodiment* of the instrument that is critical for performing successfully with any instrument (S. Fels, 2004). For many parts of the system this was achieved by assembling a setup of various general-purpose MIDI controllers and then gradually developing the most appropriate mapping strategies through several rounds of trials and revisions. But after experimenting with a lot of different ways to interact with the large collections of recorded speech segments, including using MIDI keyboards, tablet computers, touch screens etc., I found that to really be able to put the screen away and overcome the cognitive gap between hand and eye, I had to make some additional purpose-built physical controllers with just the right kind of layout and controls that I needed. When placing a combined setup of these controllers on the music stand of the piano, my experience was that I actually managed to integrate the act of performing with this new complex instrument into my existing musical relationship with keyboard instruments, and could seamlessly switch between playing the acoustic piano and the digital instrument even within the same musical gesture and the same musical line of thought. The additional way of interacting with the instrument through using actual musical utterances as sound input, either from voice or piano, made this cognitive integration even tighter. After an extended period of rehearsing, developing and internalizing a repertoire of musical possibilities, I was able to treat the system as an extension of my musicality.

This is not to say that this is a general musical instrument with an interface that will be intuitive and easy to use for any untrained person. Like any instrument it needs practice to master, and like most complex digital systems this instrument is based on certain preconceived notions about what kind of things will be interesting to do, and what kind of music will result from it, with aesthetic choices embedded in every step in the design and construction of the instrument. There is an obvious overlap of the way I already approach improvisation and think musically on the keyboard, and what kind of musical output is possible to create with the digital instrument system. The idea was never to design a general-purpose musical instrument, but to realize a personal artistic vision, an extension of musical ideas that I was unable to realize on keyboard instruments alone.

Performance issues
In addition to these technical aspects of control and instrument design, one overall concern relating to this instrument has been how to integrate the new approaches to *music making* presented by this project into a performance practice honed over many years in the role of an improvising keyboard player. During early system testing I was essentially just playing back individual recordings of speech from beginning to end, trying as best as I could to keep up and make some interesting musical arrangements or *'translation'* on the fly. This felt insufficient. I was anxious that this whole approach to live *"orchestration"* of recordings was flawed, and that it was going to be more like

superficial "remixing" than the creative process of musical exploration I associate with improvisation. One of the ways I addressed this was to bring the sound into the same sonic space as the piano. Using transducers attached to acoustic instrument as 'acoustic' loudspeakers, I found that the electronic computer instrument came closer to the sonic realm of physical acoustic instruments, and it became easier to relate to my role as a performer. During trials with transducers mounted on the sound board of the piano, I found that when digital instrument sounded through the piano and was controlled from sitting by the piano, I could actually draw on the close relation with the piano and somehow transfer this embodied way of thinking music into performing with the new instrument system.

Example: piano dialogue study, performing simultaneously with the software system and piano, one hand on each, in a kind of dialogical exploration of the musical possibilities suggested by the analyses coming out of the software:

http://orchestraofspeech.com/wp-content/uploads/2015/02/piano-duo-01.mp3

But the system was still quite slow and impractical to operate, and not intuitive enough to use to be able to pursue new musical ideas appearing in the moment. For instance, to change musical character, I had to know exactly which recording to use. To speed up changes, I started using pre-composed lists of selected recordings that had the kind of character I wanted, but though this made navigation quicker it also narrowed the options considerably during performances. A turn came when I started to organise recordings into *corpora,* databases with whole collections of analysed and segmented recordings describing the character of every segment in the corpus. Then it became possible to sort and find recordings based on musical criteria like tempo and register, and much easier to navigate large collections of many different recordings. It was also much easier to juxtapose several recordings and make collages of similar segments on the fly, opening for other ways of organizing recordings based on association rather than narrative. However, the most significant change with this corpus approach was the added ability of making statistical models based on the descriptors of the segments in the corpus. Such a model describes the likelihood of any transition in the corpus, and can be used to create alternative sequences and patterns that are statistically as *likely* based on the original sequences, and therefore share the same overall characteristics.

Initially I had reservations about fragmenting the speech sounds too much, as the particular *timing* of how events combine into phrases is one of the most important characteristics of the *speech gestures* that I was trying to hold on to and explore. I had after all wanted to investigate the musical structures implied by these original speech gestures, and not just dressing up any old habitual ideas with fragmented speech sounds. However, with such statistically probable alternative sequences, the overall characteristics of the

52

original timing and intonation patterns are actually preserved, even when the gesture itself is rearranged and made up by many shorter fragments. This is because no transition between any pitch, duration or other feature is used that does not already exist in the original sequences, and how often they occur is also determined by their statistical frequency in the original material. This way, typical patterns of successions of long and short syllables, of high and low pitches, phrase pauses etc., are actually carried over when new sequences are generated. If an alternative organisation less typical of speech is imposed, such as cycles of repeated segments or collages of similar sounds, then a gradual transition between such music-associated organisation and plain speech organisation is still possible. This way such formal distinctions can be dealt with musically and reflected upon in the music itself.

Another important feature made possible by using analysed corpora and statistical models, is the ability to trigger segments with *live sound input,* using speech, song or even a musical instrument as a kind of acoustic "controller". Live sound input can be used to query the statistical models, returning the most likely *continuation* based on the sequences already the corpus. In effect, this creates a kind of rudimentary, Dadaistic "speech recognition" system, listening and producing musically probable (while of course semantically nonsensical) utterances in response to live sound input. This way, the active interpretation and reaction to unexpected responses could also be placed right in the centre of the musical focus on improvisation.
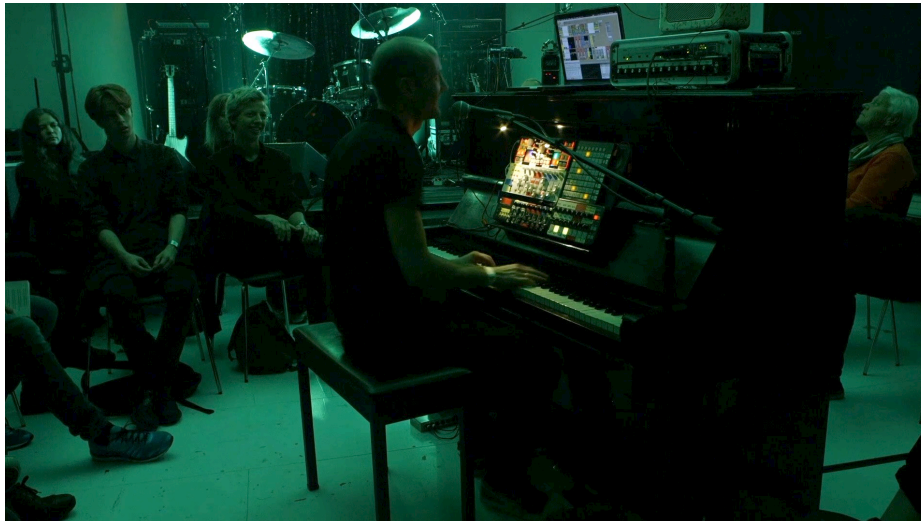
Improvised interaction with the Markov model. Video: https://www.youtube.com/embed/RLGgfLFo4Q4

With these features, the system became much more responsive and intuitive to relate to as a performance instrument, reducing the conceptual gap between controlling a computer program and interacting in the 'here and now' world of acoustic instruments. This meant that I could integrate both software instrument and piano into the same performance setup, and use them as extensions of each other.

53

Possibilities and limitations

The present instrument system is a result of a long chain of decisions based on both articulated and unarticulated aesthetic notions of what kind of music would be interesting to make, at every step imposing certain limitations and possibilities. It was designed to be quite flexible and allow for a wide range of musical ideas, but any instrument will have practical and aesthetic constraints that shape its affordances. This is an attempt at describing the musical possibilities and limitations of the system, both in terms of what kind of musical ideas it can and cannot realize, but also possibilities and limitations relating to musical form, interaction and instrumentation.

Possibilities

Though conceived as a specialized instrument for exploring the musical potential of speech gestures, the system has been designed to be able to realize a wide variety of musical ideas based on such gestures, from pure melodic passages, rhythmic imprints, dynamic envelopes, spectral shapes, collages of fragments, to sound textures and abstract timbre-based ideas very far removed from the gestural qualities of the speech sources used. The basic speech phrases can be radically transformed, either simplified into abstracted shapes, or ornamented into greatly more complex gestures by for instance inserting rhythmical subdivisions, arpeggiating through the partials of every syllable, assembling dense clouds of similar speech segments, played back at multiple times the original speed or paused in the middle of a vowel to dwell on the details of its spectral characteristics on the microscopic timescale. One of the core ideas is the notion of a continuum between speech on one hand and a conventional melodic music on the other. This is reflected in the ability to reproduce speech sources in a continuous field of transformation anywhere from the original free flowing and continuously changing speech sounds, to a highly abstracted rendering of the speech source in a stylized musical expression of conventional instruments with clear attacks, tempered pitches and quantized rhythms, not unlike the musically quite traditional accompaniments derived from spoken poems in the music of Paul Lansky.

Regarding form, it is also possible to pursue many different ways of *developing* musical ideas, from working with single soloistic ideas, supplementing a main motif with background and middle-ground accompaniments, to layering several voices and counter voices in multi-layered and multi-subject narratives, to churning round in cyclic repetitive structures, ending parts through gradual transitions or abrupt changes, or coming to a total halt with stretched-out static soundscapes, and any combination of the above.

Another aspect that also relates to form is the possibility of interaction and having the system react in an apparently responsive way to sound input. This allows for a completely different way of relating to the instrument and the speech sources, as it highlights the dynamics of dialogue and active interpretation as a topic in itself in the performance situation. Either in direct

call-and-response exchanges, or providing responses as shadowing accompaniment, as automatically generated textures or repeating background patterns, or accumulating live recordings and reusing the input sound allowing one to develop 'conversations with oneself'.

Finally, the system has possibilities for very different ways of producing sound through the ideas of *instrumentation* or *orchestration,* using a combination of both conventional loudspeakers, low fidelity radios and acoustic instrument/loudspeaker hybrids as sound sources. This allows for a thematic play with sound sources and sound realms, as well as complementing the electronic soundscape typical of digital musical instruments with the richness and spatial qualities of acoustic instrumental performances.

Limitations

What kind of musical ideas that are *not* possible to create on this instrument is harder for me to define, as the whole instrument is designed according to the kinds of musical ideas I am interested in working with in this project. It is nevertheless possible to identify some practical limitations of the instrument system, mostly resulting from aesthetic decisions, that will affect the kind of musical ideas that are possible to realize. For instance, even though the playback of recordings is handled as granular synthesis, few of the common and idiomatic granular synthesis techniques are implemented, such as backward playback and random or fluctuating variation of pitch transposition, grain rate, grain size, grain clouds etc. Segments are also not analysed with any spectral descriptors, meaning that they cannot be used to create *audio mosaics* based on spectral characteristics. The decision to use syllables (vowels) as the lowest musical unit also imposes limitations, since it means that the system cannot make musical textures based on individual phonetic units. Considering that the instrument is more like a real-time interactive compositional aid or arrangement tool than a conventional gesture-to-sound instrument, it is also clear that it is not particularly suitable for synthesizing voice and speech directly based on expressive performer gestures.

Another constraint that is related to form and musical development is the fairly limited way the system incorporates interaction. It is basically unable to act on its own, and it can really only react in a direct response to input. This is largely due to it being conceived primarily as a solo instrument made for performance and not as an automatic accompaniment system that can function like a semi-autonomous dialogue partner. One limitation this imposes on the music is perhaps that it will be perceived as mostly expressing *one* musical subject, that of the performer who controls the overall aspects of musical development, even when individual details and seemingly polyphonic layers might be generated automatically by the system.

If we look beyond the explicit aims of facilitating improvisation in this project, the strict real-time mode of operation adapted in the system is another limitation. Performing with the instrument, it is quite hard to repeat something exactly the same way every time and generally no way of recording the output symbolically as a score or other exact representation. This makes it hard to use for sketching ideas for composing written music, and also hard to use the instrument for performing detailed compositions with exact notation.

Finally, there is an obvious artistic potential on the semantic level of speech which is not touched at all in this project and therefore not part of the system. There is no segmentation or recognition on the semantic level of *words*. One could easily imagine functions also dealing with semantic content and wider poetic use of words, like for instance in the *Mask Mirror* performance system developed by Alessandro Bosetti (Bosetti, n.d.), where recordings of words are classified according to categories like adjectives, places, times, etc. and then used by the performer to dialogically construct and explore semi-random poetic narratives. So even if these kinds of features were omitted by choice, it is clear that this choice severely limits any ideas where semantic meaning is important.

Speech sources and concept

This chapter presents some thoughts on the process of choosing speech sources for this project, and the challenges resulting from having to relate to the contextual and conceptual implications of using sources taken from real-life situations.

This has to do with the relation between the musical reality of sound and the conceptual reality of words, and describes some of the reasons for the move away from the semantic content of speech in this project, as well as the actual kinds of speech sources I ended up using.

Speech sources

Early on in this project I considered basing the performance concept on live speech as musical sources during performances – readings, poetic improvisations, acting, or perhaps other kinds of staged speech. However, I quickly became aware that I was more interested in the interplay going on in everyday, real-life conversations, without the implications of performative public or theatrical speaking styles used when standing before an audience.

As described in the chapters above, I chose to use *speech genres* as the main perspective to approach the abstract gestural musical content of such conversations. To explore variations in such speech genres I needed to gather recordings from a broad range of different social situations, such as intimate conversations, formal and informal dialogues, quarrels, public encounters, interrogation, negotiation, confessions etc. After doing some tentative recordings of my own, I realized that I would not get myself into a lot of such situations with a microphone and recorder very easily, and started to search for other sources. During that search, I listened to a lot of different recorded speech, and I was often struck by how sensitive perception is to subtle nuances in character. For instance, at some point I thought that actors ought to be experts at speech genres, so I turned to recordings of acting on screen, on stage, from radio drama, improvised theatre etc., but when listening to these recordings I experienced that these acted situations also constitute their own speech genres, differing slightly from what we would expect if they were for real and not acted. We expect acting to conform to these acted styles, which is part of the message that it is just fiction after all, even though it often portrays reality. Perhaps this ability to discern authenticity and sincerity from a staged or acted performance is linked to our extremely social nature and almost obsessive preoccupation with what other people are *really* thinking, feeling and planning, which is not always obvious from what they actually say. We are experts interpreting speech intonation and figuring out if that laughter was authentic or acted, if that excuse was really heartfelt or just courtesy.[1]

---

[1] As a side note to my decision to focus on *real* conversations as material, it must be said that I later also found it interesting to use recordings of highly emotive speech by actors making dramatized stereotypes of different emotions. While not sounding convincing at all, they nevertheless had a kind of almost abstract poetic quality. Perhaps in the way they sounded stylized, they actually came closer to poetry and music.

Conceptual framing: a search for the non-specific situation

The decision to use authentic conversations as sources was the reason that I had to rely on *recordings* of speech instead of spoken word on stage. Those recordings have to come from somewhere, and as speech can potentially be sourced from anywhere in the sphere of human activity, this *somewhere* can potentially bring all sorts of new content into the mix.

My first aim was to gather a wide range of speech genres that could be starting points for musical explorations. When working on this I felt the need for an overall concept regarding which recordings to use. An idea about from where and when these conversations took place that could emerge as a theme throughout the project.

That set me off in many different directions. One of the first useable sources I found was a series of reality television used for linguistic research at my university. Though from a narrow demographic selection, this material included a wide range of different speech genres both formal and informal, happy, angry, sad, personal, public, leisurely etc. The participants were recorded day and night for a long period of time, and a bit into the season they seemed to act quite natural even in this unnatural setting. This is also probably why this material have been used for linguistic research on natural speech as well. However, the somewhat ironic conceptual framing of a TV show with all its references to popular culture was not something I was looking for.

Another concept I looked into was the politics of power, searching for, and listening to, covert and private recordings of conversations by people in powerful positions. But away from the public spotlight, these people tend to sound very commonplace, very much the opposite of what we get from acted portrayals of such people on screen. Perhaps this can be viewed as a parallel to what Hannah Arendt famously described as the *banality of evil*, in the sense that even though people in such powerful positions can potentially affect the lives of millions of others by the very words they utter and the decisions they make, the sound of them speaking can actually be as trivial as any other everyday conversation.

When looking at possible sources I also constantly ran into the typical anthropological problem of how to observe and record authentic situations without affecting them by the very presence of a microphone. This led me to consider surveillance recordings as a possible source of authentic situations. Apart from the obvious ethical considerations I thought this approach could also become an interesting topic in itself, providing a conceptual frame for the musical explorations and also a relevant comment to the trend of more surveillance and interception of mass communications that we see in society today.

I contacted the Stasi Archive in Berlin and got permission to obtain and use actual surveillance recordings that agents at the Ministerium für Staatssicherheit recorded in East Germany during the 1980s. But when trying to use this material, it became clear that the strong connotations and the depressing context made it far too powerful to use the way I had intended. The context simply took over and made it all about living in a totalitarian surveillance state while the music became secondary. While this material was clearly very interesting, it was an entirely different project than the one I was working on.

To continue with the musical exploration of speech I instead sought to gather a wide range of less historically specific recordings. From different public sources like radio and television, anonymised linguistic databases of recorded natural speech, as well as making recordings of my own.

When working with this diverse material over time, it became evident that while for instance the reality-TV recordings were very rich in expression, what worked best in practice was actually using the least recognisable sources. Especially when I was improvising with another performer, such easily identifiable sources with a common popular cultural reference from mass media really seemed to stick out and conflict with the on-going musical discourse. I had an exchange about this with my supervisor Øyvind Brandtsegg that perhaps can shed some light on these issues:

> **Daniel**: *It seems to be a problem when bringing too much contextualized content into an otherwise non-contextualized musical situation, especially when improvising with another musician. It takes the focus away from what is happening musically.*
> *On the other hand, I also feel a need to have a consistent selection of material constituting a conceptual framing of the project.*
>
> **Øyvind**: *This is an Interesting issue that is somehow central to your project.*
>
> **Daniel:** *Yes, it is an important question for the whole project. I have been contemplating this from the very start, but up until now it has been more pressing to get things to work technically and musically.*
>
> **Øyvind:** *Are you perhaps looking for some way of staging of the music? A context that provides a direction but at the same time freedom?*

Daniel: Yes, some sort of anchor point, framing or approach that clearly defines a field within which I can operate freely.

Øyvind: has this to do with the literary content, situation, mood, setting, or also that the dialogue is bound in time?

Daniel: I think it probably has most to do with the connections to the concrete reality of any recognisable persons, and the context that this imposes on the music.

Øyvind: Is this comparable to the content of lyrics in vocal music? Or the stories that are told in spoken word music? (Laurie Anderson, Golden Palominos med Nicole Blackman, etc)

Daniel: Yes, me and (fellow performer) Tone Åse talked about that. She felt the same about using written texts in an improvised setting, and for this reason would often choose the most abstract texts. I think this is more of a problem in improvised music where the music is already about the interaction of musical ideas, than in a composition that already relates to the textual content.

Øyvind: Do you need one solution for this or could you do different things for different musical results?

Daniel: No, not necessarily one final solution. Different pieces and musical settings could use different solutions.

Øyvind: What is it about ensemble improvisation that makes this problem more precarious?

Daniel: Perhaps improvised musical dialogues seeks to establish a common formal language that creates its own abstract context, and when bringing in material with an external context pointing outside the time and space of here and now, this creates a conflict of what we are dealing with.

So, what I have been looking for, I think, is something that can represent the idea of generalised social situations, without being so specific that one gets involved in each individual story. At the same time, I felt that it has to be a sort of consequence, an overall idea or concept that makes clear why I am using this or that source, but without taking over as the sole content of what the music is about.

Using the Stasi surveillance recordings introduced too much context, to the extent where it completely took over as content. To record conversations myself could have been preferable conceptually, but that would result in a very narrow range selection of languages and situations. And what would the subject be? A portrait of me and my close environment? At this point, that was not something I was interested in.

One possible solution was to mix a large number of less recognisable sources. There is still no overall theme or conceptual content other than the diversity of language and expression, but if the main subject is the generality of language and music then this is perhaps how it has to be. The diversity of human experience would be the overall concept. Another way to look at this question of conceptuality is that the focus on *speech genres* already is a concept that dictates what kind of material I need to use, and that additional conceptual framings would ultimately conflict with this concept of generality. This is above all a *formal* concept, and one that fits well with my fundamentally formalistic approach to music. This is also why I think of any outcome of this project primarily as *music* and not as *sound art.*

As these thoughts shows, it became necessary – at least for the time being – to somehow do away with the semantics of speech to be fully able to focus on the nonverbal content and the connections to improvisation and musical communication.
To reflect a bit further on this choice of not focusing on the semantic content and its implicated narratives, contexts and historically situated references: The point was to move beyond the conceptual reality of words and focus on the part of language we learn before the words – the communicative vocal gestures, and to explore interaction in conversation in relation to the kind of physical/sonic meaning-making going on in improvised music. This, I think, is on a philosophical level also related to the meaning and function of music as a social phenomenon: playing music primarily as a meaningful way of being (and thinking) together, and not just making a product for consumption – echoing the emphasis by composer Cornelius Cardew on music as a community activity, and the view that the audience (i.e. concert tickets) was a capitalist invention. This line of thought is also linked to the ideas of Roland Barthes that there are two *musics* – the music one listens to and the music one plays (Barthes, 1977), describing how the music one plays is not just perceived as sound but that its meaning is actively created and comprehended through the gestures of the body (Barthes calls this *muscular music*). This physical, and in the case of shared rituals like for instance church song, also very *social* meaning of music, is what I have tried to show as a parallel to the likewise very social nature and physical togetherness that is a large part of the function of conversation.
Perhaps the approach in this project can be viewed as more of a search into the traces of musical meaning-making that can be found in speech, rather than dealing with all the other (and equally interesting) aspects of speech, such as words, poetry, semantics, voice, identity, personality, community,

stories, history, society, ideas, concepts and so forth. In other words – the (physical) act of speaking together rather than the spoken word as a concept seems to be the core of what I have been trying to use as the focal point in this project.

Speech recordings

Following the thoughts above about conceptual framing and semantic content, I developed an approach where I tried to use as many different sources of recorded speech as possible, deliberately avoiding any clear contextual markers. In contrast, in some speech-based music the process of collecting and recording material can sometimes be the very starting point and *raison d'être* for a musical piece, making up a large part of the actual work of making the music. For the piece *"Encounters in the Republic of Heaven"*, Trevor Wishart reportedly spent a year just establishing contacts and recording local sources, and then an additional 18 months editing and cataloguing the recordings (Wishart, 2012, p.136). Compared to this, I did not know what kind of material I needed for this project until I was far into the process. Investigating what could work as material and exploring the particular effects of using this or that kind of source became integral parts of the research process in this project.

The reflections above give an account of some of this process, and how I arrived at the conclusion that in practice, what turned out to work best was to use a mix of many different sources of recordings: my own recordings, shared community recordings, clips from reality-TV, many different linguistic corpora, surveillance recordings, court recordings, telephone recordings, radio broadcasts and documentaries. These are only some of the kinds of sources I have explored:

> *Santa Barbara Corpus (linguistic corpus of natural speech)*
> *Big Brother television series (TV Norge 2001)*
> *CallFriend and CallHome (linguistic corpora of telephone conversations)*
> [*archive.org*](archive.org) *(community recordings, field recordings, religious sermons)*
> [*talkbank.org*](talkbank.org) *(collection of many different linguistic speech corpora)*
> *BBC podcast - "the listening project" (recorded community dialogues)*
> *News broadcasts*
> *ATCOSIM (Air Traffic Controllers simulation speech corpus)*
> *Emotive speech corpora (Emovo, EmoDB, Ryersons)*
> *Supreme Court of the US archive*
> *Najonalbiblioteket (National Library of Norway – historical recordings)*
> *Stasi surveillance recordings*
> *Watergate covert recordings from the Nixon Library*
> *Emergency telephone recordings (linguistic corpora of emergency calls)*

Some of the most interesting sources to use were *telephone conversations*. A phone call usually has a clear closed form with a definitive beginning, middle and end, reminiscent of (western) narrative musical forms. It is also the main

reference for interacting by means of the voice alone. As the voice through the receiver is the sole medium of communication, and no body language or other visual cues are available, the prosody, intonation and speech genre become very important aspects of the interpretation. This is also why I chose to use a real telephone for the sound installation in the final presentation of artistic results. Perhaps this particular mode of communication has the potential to be made into an overall conceptual framing in future work.

Perception of speech and music

One of the key questions articulated in this project is how the use of speech gestures as material in improvised music affects the perception of both music and speech. In this chapter I will try to present some of my observations and reflections in this regard, relating to the different *modes* of perceiving speech and music, how different *sound sources* influence the frame of perception, and different perceptions of *time* in both speech and music.

Semantic and aesthetic *modes* of perception

The most fundamental difference between speech and music seems to me to be between perceiving speech as semantic content, and music as aesthetic form.

In his famous 1917 essay *Art as Technique*, the Russian writer and formalist literary critic Viktor Shklovsky describes how perception of everyday phenomena is usually highly automated through familiarisation, to the point where they are not actually *seen* any more but only recognised as symbols. Through the technique of *de-familiarisation,* Shklovsky notes, art can turn attention back to the sensation of their features and forms as they are experienced for the first time (Shklovsky, 1965). Similarly, from a cognitive perspective, Per Aage Brandt proposes that the mode of perception is determined by the context – if the experience is strongly *framed* or not (Brandt, 2006). In the trivial and unbounded stream of everyday experience, perception is *pragmatic* and oriented towards content and action. In a strongly framed *aesthetic* context, the mode of perception is *extraordinary,* an intense mode of *form*-oriented hyper-perception. For Brandt, the concept of *form* in art is key to this framing of reality that enables aesthetic perception.

This means that everyday speech, by its very *everydayness,* by default is perceived through this pragmatic, content oriented mode of perception. Even when formal operations like abstraction and fragmentation introduce an aesthetic framing, there still seems to me to be a cognitive split between still recognizing and focusing on the everyday speech sources through this pragmatic mode and perceiving music as abstract aesthetic forms. In my experience, it is very hard to focus on both the semantic and musical aspects of ordinary speech at the same time. Instead, perception seems to flip between the two modes, with one always being in the foreground depending on if I know the language, how intelligible the words are, and if there is a clear narrative or story that draws attention.

This is not necessarily just about aesthetic framing. Perhaps it also has to do with the way music and language seems to be processed in different parts of the brain. The fascinating "*split brain*" research by Roger Sperry and Michael Gazzaniga, carried out on severe epilepsy patients that as a last resort had the connections between their brain halves cut, has provided some interesting insights into how processing in the brain is divided (Gazzaniga, 1967). According to this research, the left hemisphere seems to be in charge of intellectual processes like storytelling and creating narratives and semantic

meaning, while the right brain hemisphere deals with sensations, and spatial, emotional and musical cognition. One way to look at this, is that we are equipped with two parallel perceptual systems in effect creating two parallel experiences competing for attention, and that might explain the difficulty of retaining focus on the musical forms of speech when the spoken narrative grabs our attention.

Sound sources as framing

This tendency for semantic content to capture the main perceptive focus is the background for my whole methodology of abstraction in this project, reflected in the reformulated aims to explore the musical and communicative potential of speech primarily as vocal prosodic *gestures*. In addition to the abstraction methods offered by signal processing, this also extends to the actual sound production setup of different physical *sound sources*. Contrary to Schaeffer's ideas of an acousmatic "reduced" listening, I believe that the sources and inferred causes of sounds will always play some part of our perception of sound. So rather than to supress this tendency, the active play with sound sources is embraced as a central feature in this project. The use of a whole range of different sounding loudspeakers and resonating bodies creates a layer of formal differentiation that shifts the attention to their sonic characteristics and the sources as entities in the room. The use of these hybrid acoustic instrument-loudspeakers can render speech perfectly intelligible while at the same time colouring the sound just enough to create an aesthetic framing. This framing also creates a contrast to the conventional loudspeaker sources and enhances the perception of the qualities and connotations of them as well. The sound of an acoustic instrument somehow *means* music, and creates a frame of reference that invites musical listening, while speech mediated through a low-fidelity radio have connotations to broadcast and public address. In contrast to these physical objects, the invisible sound wall produced by stereo loudspeakers creates imagined, virtual sonic spaces. New and interesting perspectives for listening appear when these sonic realms start to blend in an orchestration of different sound qualities and physicalities.

Time perception

Another interesting aspect of perception contemplated in this project is different experiences of *time*. One common kind of time perception relates to the *narrative,* creating an expectation of a linear story to be told, unfolded from start to end in a forward motion through time. This is very different from the kind of time experienced in the moment of unplanned, responsive interplay, like in a spoken conversation or in an improvised musical exchange. Even with a shared experience of what has happened so far and what might happen next, the present is somehow *accentuated* by the very fact that any response is unknowable before it happens, reflecting the direct etymological meaning of the improvised as something *unforeseen*. The dynamics and the risks and rewards of that moment can result in a feeling of an *expanded* present, a *continual now.* Indeed, according to merited improviser Derek Baily,

improvisation can even be considered a *celebration of the moment* (Bailey, 2004). This is an experience of time that is qualitatively different from the linear narrative, and one that is fundamental in improvised music.

In addition to these ideas, some additional concepts have been adopted for thinking about time in this project: the standstill of a static soundscape or a sparse background, and the circular time experienced with fragmentation into cyclic repetition. Together, these four perspectives on time perception have emerged in this project as a result of – and played a part in – the exploration of differences and similarities in speech and music:

Forward linear motion

Narratives are natural in both speech and music, as a story with an implicit dramaturgical development. That can be a monologue, the presentation of documented dialogues (where the act of presenting this dialogue becomes a narrative) or any kind of public address like the very situation of a musical performance where conventions create an expectation that performers will walk on stage and produce sound intentionally for an appropriate period of time before the performance will end and the audience will applaud.

Attentive present

The alertness of the moment is also a feature of speech and music alike, appearing in every unexpected response and unforeseen twist in everyday conversation and improvised music, perhaps best exemplified with the concept of question?

Cyclic time

Repetition on the other hand, is a phenomenon that is more related to the physical experience of body movements, of dance or manual labour, or the repetitive or cyclical structure of movement in the physical world in general (raindrops, heartbeats, respiration, walking, machinery, weather, seasons, planets). Repetition for its own sake is not often encountered in everyday speech except to emphasize something, but is very much present in the aesthetic domains of music, poetry and ritual. Indeed, repetition is a defining feature of so much music from all cultures around the world that it is stressed by some cognitive researchers as *"a fundamental characteristic of what we experience as music"* (Margulis, 2014, p. 5). The use of cyclic structures with speech will quickly draw attention to its formal structure in what is known as *semantic satiation*, framing it within an aesthetic mode of perception and shift the focus away from the semantic content, and for most purposes in everyday spoken communication this defies the very purpose of communicating. But for the purpose of making music, allowing varying degrees of repetition with the speech material makes it possible to play with this clear marker between language and music and direct the mode of perception towards the one or the other.

Static time

Stillness complements these modes of time perceptions as the opposite of intentional communicative gestures, as the indifferent (but perhaps expectant) background without the sound or movement of any acting agents. It offers other perspectives on sound as a phenomenon in space, which was one of the reasons for making a sound installation version of my performance concept – a sound installation without performers or linear narrative, but with the possibility to walk around and structure the experience spatially.

These are some of the ways the use of recorded speech in improvised musical settings has highlighted the shifting modes of perception of both speech and music in this project. That has also played a part in developing thoughts about form in this project, something that will be presented in the next chapter.

Form, meaning and language

This project is not meant to be a theoretical dissertation into matters of philosophy and aesthetics, but I think it nevertheless is interesting to touch upon the subjects of *language, form* and *meaning* in relation to the music developed in this project. Especially so since I am kind of using the *form* of spoken language as the meaningful *content* for my music. In this case the *form* of speech refers to the prosodic structures that have been identified as significant in linguistics and conversation analysis: melodic intonation contours that express the course of utterances or cue turn-taking, change of key to signal change of subject, accents and stress to mark new or important information, convergence of tempo etc. These underlying musical structures of speech have been starting points for the musical explorations. This exploration of the musical potential of speech is interesting because it can also say something about the semiotic potential of music and point at ways in which music might make sense.

In relation to the significance of prosodic gestures, I think it is very interesting to see how infants seem to develop this kind of underlying melodic framework even before they learn a single word. *Baby-talk* between infants and parents during the first months after birth seems to work as a fundamental coordination of vocal cues with actions, reactions, facial expressions, emotions, and intentions. This first step of language acquisition is apparently common in all cultures and involves not just learning simple vocal gestures, but combining them into complete melodic narrative structures that later provide the formal foundations for interaction and constructing utterances of speech (S. N. Malloch, 1999; S. Malloch & Trevarthen, 2009b; Miall & Dissanayake, 2003; Snow & Balog, 2002). Such *melodic messages* in baby-talk carry direct meaning by conveying the basic communicative intent of an utterance – for example if its function is attention, prohibition, approval, comfort or play. This is particularly apparent in what linguists calls *infant directed speech,* where prosodic contours tend to be exaggerated, but is also clearly identifiable in interaction between adults (Fernald, 1989). This is interesting as it shows how this musical foundation for language that we learn as infants constitutes a kind of underlying melodic vocabulary that we use and understand intuitively in speech, and that most likely provides a background for our perception and appreciation of music as well. This is also the reason I chose to focus on these pre-semantic aspects of speech rather than the concepts and ideas of words.

Language and meaning

One of the key features that are conveyed through prosodic gestures is intention. The significance of intention has also been stressed in several theories about language, like in the *speech act theory* developed by John L. Austin (Austin, 1962) and later extended by John Searle (Searle, 1969). This theory seeks to understand the meaning in speech not primarily from the semantic content of words, but from the *performative* function and intention of the utterance as an *act* – of what it is trying to achieve. This perspective on

meaning-making is also related to Wittgenstein's ideas about *language-games* where the meaning in language derives from its actual use, and that philosophy cannot deduce some kind of essential meaning in language separate from that (Wittgenstein, 1953). According to Searle, the very act of speaking presupposes an intention, defined in his concepts of different *illocutionary forces,* such as *declaring, demanding, ordering, warning, promising, inquiring, exclaiming, asserting* etc. Seen in relation to the kind of intention conveyed by speech melodies in infant-directed *baby-talk*, these forces can be viewed as a kind of *abstracted intentional meaning* that it is reasonable to believe could play a part of musical experience as well, even without the particular propositional content of spoken utterances. This inherently social function of speech as *action* and *interaction* can also be related to the ideas of Mikhail Bakhtin. In his view, the form of utterances is not only shaped by the intention of the sender, but by a dialogical relationship between sender, receiver and the social circumstances. This is the process whereby meaning is created, and why *speech genres* make up an important part of the meaning of utterances (Bakhtin, 1986). Since speech genres involve the use of certain prosodic patterns that can be described in terms of musical characteristics, this is also why I have found speech genres to be an interesting approach to exploring the musical content of speech. Interesting because this layer of meaning forms a *common* reference – a kind of shared social musical language that can have as much precision as the specific words used. The features of some typical genres are often well defined and easy to tell apart, to the degree that their acoustic characteristics can even be reliably identified and classified by automatic computer analyses (Obin, Dellwo, Lacheret, & Rodet, 2010; Obin, Lacheret-Dujour, Veaux, Rodet, & Simon, 2008).

The significance of speech genres and thus a hint of some kind of conventional musical meaning in speech does not however mean that it is possible to point at *one* essential meaning of a musical utterance. Music does not represent the kind of formal communication system that defines languages. The regularity, and the seemingly orderly harmonic, melodic and rhythmic *"rules"* observed in certain styles of music have nevertheless tempted many scholars to approach music as a formal language complete with rules for grammar and syntax, from the highly developed rhetorical figures of the baroque music, Rousseau's view of music as *"impassioned speech"* (Rousseau, 1781), Leonard Bernstein's Harvard lectures on musical semantics (Bernstein, 1976), Jackendoff and Lerdahl's musical adaption of Noam Chomsky's generative grammar (Lerdahl & Jackendoff, 1983), and various approaches to music as semiotic systems of signs derived from the tradition of Roland Barthes etc. Raymon Monelle gives a thorough account of how such linguistic theories have been applied in musicology (Monelle, 1992). While many interesting insights can be gained from such perspectives, there is an underlying assumption that all music can be explained with one scientific method. In the search for universality, such approaches overlook the multitude of ways music (and speech) makes sense at the same time. And while languages are well-defined, pragmatic communication systems that can

be fairly easily described by rules of grammar and syntax, music is an open-ended poetic mode of expression in the aesthetic domain that produces meaning also by challenging such rules and conventions.

Regarding linguistic ideas, I have found the creative semiotic approach of the musician, filmmaker and semiotician Theo van Leeuwen to be more fruitful: Instead of the descriptive *"what is"* of scientific explanation, this approach offers a creative *"what if"*, treating sounds as untapped semiotic *resources*, structures with many layers of potential meaning (Leeuwen, 1999). Such potential includes for example how sounds with similar proportions to human breath periods (duration, dynamics etc.) easily can be perceived as intentional and communicative utterances. In fact, our perception seems so overly hardwired for interpreting meaningful outlines of such patterns that we end up seeing faces in clouds and hear whispers in the wind as well, something that in gestalt psychology is known as *illusions*. Further along this focus on *communicative intent*: how the characters of utterances, actions, movements and gestures are intuitively interpreted and empathically mirrored as signs for inner states, thoughts, emotions, intentions etc. This needs of course not be interpreted literally, as our infinite ability to create metaphors can easily make these signs into poetic pictures of something else.

From the perspective of music psychology, John Sloboda has proposed that in the broadest sense, the perceptual background for experiencing the dynamic processes presented by music is our experience of the physical world in motion, and in that world – particularly the moving, living organism (Sloboda, 2012, p. 170). Not as mimicry, but on a deeper level how motions are initiated, experienced and mediated by a human agent. This could even be true in a more general sense for abstract thought as well, like in the way we tend to use spatial metaphors when talking about ideas *behind* concepts, thinking something *through,* being *on top* of things, looking *further,* view from another *angle*, against a *background* etc.

Form and meaning
Regarding the relationship between content, context and concept, it seems clear that my approach to music and speech is quite formalistic. But not relating to the semantics and concepts of words does not mean that I am not concerned with meaning. There is a popular myth that music without lyrics do not express anything specific beyond general emotions, but my experience as an improviser is that music certainly can *make sense* – and in very particular and nuanced ways as well. With Bakhtin, we could counter that *words* do not mean anything specific either, only how they are used in actual utterances. Which in turn can create meaning on several different planes, both in speech and music.

In linguistics, it seems easy to make a clear distinction between the form and content of language. Such a division has also been common in thoughts about art, for example in the view by art critic Clement Greenberg, that *form is the*

handle allowing content to be grasped (Kim-Cohen, 2009). That might sound easy enough, but according to philosopher Lars-Olof Ahlberg, this simplistic metaphor of form versus content can be very misleading, as it is far from obvious what actually separates *form* from *content* in different art forms, or even within the same art form (Åhlberg, 2014).

To relate this to my work in this project, I first experienced that it quickly became very monotonous to listen to speech when the semantic meaning of the words – the *content* – was filtered out. At the small scale of phrases, such abstract speech sounds were still musically interesting, but at a larger timescale the uniform rate of events and the overall lack of consequence made it too monotonous to work as music. Dealing with storytelling in theatre improvisation, Keith Johnstone has described the need to *reincorporate* elements to create coherence, otherwise it just becomes a meaningless sequence of events that can start and end anywhere (Johnstone, 1981). This, I believe, is similar in musical improvisation. So, it seemed that what was lacking was some kind of musical *consequence, distinction* or intended *differentiation* of the musical features. This is perhaps the kind of musical *formal* consistence I felt I needed to provide, to replace the semantic content that had been removed.

This view of form *as* content can be related to the role of *theme and variation* in art, as discussed by Nelson Goodman in *"Languages of Art"* (Goodman, 1976). According to Goodman, the modification, elaboration, differentiation and transformation of motifs and patterns are processes of *constructive search*, and such *progressive variation* is a typical way of advancing knowledge. This seems to be especially true of how artworks explore the world, constructing their own *formal languages* through such processes of differentiation, and thereby also providing a *content* expressed by these formal languages.

In a more general sense, progressive variation is perhaps how knowledge is expanded in other fields as well, including in science. It can also describe how a topic might be explored in conversations or in writing, but then these processes relate to the exploration of concepts, of *thought ideas,* while music is an organised exploration of *sound ideas.* Following this line of thought, improvised interplay can be viewed as a *dialogical construction and collective exploration of such a formal language*. This, I think, is at the core of what I have been trying to do in this project and that is articulated as one of the aims: to investigate the musical potential of the sonic gestures of everyday speech, through formal differentiation and improvised exploration of its features as *sound ideas*.

If I should attempt some kind of conclusion to these thoughts on meaning, it must be that music and musical utterances have no single meaning, but can convey countless meanings at the same time. This is however also the case with language, which is certainly not as precise as we like to think, and where meaning is just as often inferred from the context and intonation (the utterance *"–apple?"* might for example be an opening line, an inquiry about hunger, a nutrition advice, a reference to the laws of gravity, or to a computer company, or it can imply that you are a princess, or that I am a witch, or serpent, or none or all of these at the same time). Not to mention that some words like *swallow* or *hide* refer to completely different concepts when used in different contexts. While some have seen this ambiguity of language as a flaw in an otherwise near-perfect communication system, it has been viewed in cognitive science as an absolutely necessary feature. Language would be far too cumbersome to use if one had to specify everything exactly and unambiguously all the time, like one has to do in computer programming languages. The same ambiguity also makes it possible to convey several things at once, like in puns or poetry. Like all utterances, gestures and actions performed by living (and imagined) beings, it is possible to interpret speech with a whole range of possible intended and unintended meanings. This, I think, must also be the case with music – it has a multitude of potential meanings, all at once.

As a final remark, it must be stated that the focus on speech in this project has not been an attempt to reduce the content or meaning of music to identify *one* underlying universal "explanation" of music. Many theoretical approaches to meaning in music have often taken the form of universalistic generalisations, like Leonard Meyer's *"Emotion and Meaning in Music"*, where meaning is viewed essentially as arising from tension and release relating to the expectations formed by learned styles (Meyer, 1956). The idea in this project was rather to broaden the experience of both music and speech, by shedding light on some interesting connections between these two universal phenomena of human nature. One artistic aim was to see if it was possible to make music that could show both the *musicality* of speech as well as the *language-like* logic of music. I hope that this work can also show how music can *make sense* as a way of thinking, and, like speech, make sense as a way of being together.

Reflections on the results

This chapter presents some reflections on the results and possible contributions of this artistic research project, in relation to the context of relevant fields and practices.

But before discussing context, it is necessary to define more precisely what the results of this project are.

In line with the originally formulated aim of developing an improvisational foundation based on the musicality of speech, the most important result of this project is the accumulated improvisational repertoire of musical possibilities coming out of this work. This is the combined know-how of how to go about turning this particular speech material into music, the intimate direct musical knowledge that slowly has been mapped out and internalized through repeated trials, rehearsals and performances, embodied as a reservoir of musical and instrumental possibilities and forming a foundation for exploring this area through improvisation.

As with other improvisational skills this embodied repertoire is partly tacit knowledge, but is observable and shareable through the documented process of its development and through the resulting improvised performances. The chapters on speech gestures as musical material, methods of abstraction, and performance methods in particular, present important parts of this development process. In addition, the reflection chapters also offer perspectives on the kind of possible meanings and insights that form a part of this tacit knowledge. Of the performances, it is particularly the solo performance from the final presentation of artistic results that best represents the aims and ideas contained in this project, and where this repertoire and improvisational foundation is best demonstrated. The sound installation and ensemble improvisation parts must be viewed as artistic experiments in this regard, as attempts at alternative uses of the resources mainly developed through the solo improvisation format, and not as final end goals of the project by and in themselves. However, what *should* be regarded as a part of the results is the particular performance concept and instrument system used for making and performing this music, as this work is tightly connected with the realization of the improvisational foundation and therefore an integral part of the particular way of making music developed in this project.

The results, then, consist of the improvisational repertoire of musical possibilities that has been developed, the performance concept and instrument used to realize these possibilities, and their realization into actual music as in the solo performance of the final presentation of artistic results in particular.

In epistemological terms, the actual knowledge produced by this artistic research would, in addition to all this practical artistic know-how, be the insights into the relation between spoken and musical interaction that this particular way of making music makes possible.

Readings and contexts

How do these results relate to a wider musical context?

As shown in the historical account sketched out in the background chapter, speech has been used in music in a great number of different ways and explored from many different perspectives. The *universal* character of both music and spoken language also means that the topics of this project can be associated with almost anything. The results can likewise be viewed from many different perspectives, each with its own context, from electroacoustic composition, to sound art, technological art, docu-music, sound poetry, free improvisation, musique concrete, interactive art, etc. Through the concept of different *readings,* I will try to build further on the context already established in the background chapter and show how the musical results might be related to several different fields and contexts.

Composition

In the context of composition based on speech, there are certain similarities in my project with the way many composers have approached speech as raw material, as sound structures which can be analysed, extracted and processed into musical forms. For instance, this includes the speech based music of some electroacoustic composers like Trevor Wishart and Paul Lansky, but also the more direct instrumental approach of Peter Ablinger in his voice transcriptions for piano as well as his 'talking' mechanical piano pieces.

Wishart has actually a wider approach not only limited to speech but working with all aspects of the human voice, from the very *granularity* of the voice as raw sound material, to the concept of voice as *identity* for individuals or communities, to far beyond speech in the extreme possible expressive uses of the voice in his vocal pieces. He has also focused on phonetic characteristics of languages, as well as more traditional musical parameters like the melodic, harmonic and rhythmic features of speech, but still mainly within the very electroacoustic tradition of the raw sound material as sound objects for constructing a modernistic electroacoustic sound realm from scratch.

Lansky on the other hand has much more of a lyrical approach, often basing his speech-derived music on expressive readings of poems transposed onto formalized grids of well-tempered pitches, fixed tempi and quantized meters reminiscent of the recited prose-poetry popular (art)music of Laurie Anderson or Robert Ashley.

Ablinger also uses recordings directly as a musical source material, but instead of poetry performances he transcribes historical speeches and interviews with well-known artists and other celebrities into *musical portraits* for piano, choosing different styles of musical arrangement according to the historical and personal associations of the person portrayed. In addition to this particular focus on the personality of cultural icons and their stylistic associations, Ablinger has also stated that his work is about perception of reality, with the piano presenting a 'phonorealistic', analytical representation of reality (Ablinger, n.d.).

In contrast to these compositional ways of using speech as musical material, I have approached speech not as raw sound material, expressive poetry or personal voice portraits, but as forming *conversations,* as utterances within specific social contexts where the particular phrasings and musical features are not just personal attributes or independent acoustic shapes but point to the gestural meaning of sound in a particular social situation, in the same way that musical gestures aquire certain meanings and communicative functions when used in improvised interplay. My choice of using *speech genres* as a focal point for the exploration of this gestural meaning has at least one parallel in Luciano Berio's piece *A-ronne*, where the performance of a poem is subjected to the changing meanings caused by various *"vocal situations and different expressive characters"* (Berio, 1974). But as a composition for vocal performers, his piece is a performative staging of these situations and characters, composed as stereotypical expressions that end up in a very theatrical performance. In my view, such theatrical speech genres have additional layers of very different meanings from the real-life genres they attempt to represent, and this is exactly the kind of staged theatrical result I have tried to avoid in this project, and the reason for basing the whole project on recordings of real-life conversations rather than enacted dialogues.

Instrument development

There is also a relevant but completely different *technological* reading of the results of this project, experiencing the music primarily in terms of instrument development and themes like technological art and human/machine interaction. One relevant context in this regard is the field known and community of modern day luthiers, engaged in creating new interesting musical instruments and interfaces, often in relation to new technologies. Relevant references in this field are digital musical instrument systems for improvising with captured sound material, such as the ImproSculpt instrument and the later cross-adaptive live processing techniques developed by Øyvind Brandtsegg (Brandtsegg, n.d., 2007; Brandtsegg, Saue, & Lazzarini, 2018), or the corpus based concatenative synthesis instrument developed by Diemo Schwarz (Schwarz, Beller, Verbrugghe, & Britton, 2006), to name but a few. More specifically related to *speech* are the many kinds of voice- or speech-based instrument systems mentioned in the background chapter, such as the spatial-gestural play with voice samples in the instruments developed through Grégory Beller's Synekine project (Beller, 2014), the voice-accordion of Perry Cook (Cook & Lieder, 2000), Sidney Fels' work on digital voice modeling and gestural voice synthesis control (S Sidney Fels, Pritchard, & Lenters, 2009), amongst others. With my use of the acoustic piano as an integral part and focal point of the instrument system, this luthier context extends to a category of so-called *augmented, expanded* or *hybrid instruments* – usually traditional acoustic instruments that are extended with various technological means to broaden musical possibilities and playing techniques. Examples on such instruments include the electronically enhanced HYPER(sonal) piano project of Morten

Qvenild, the Electrumpet of Hans Leeuw, the magnetic resonator piano of Andrew McPherson, as well as the *Strophonion* of vocal performer Alex Nowitz, to name a few. Further, this project's use of transducers and orchestration of physical electroacoustic objects (or *subjects* rather) forms a parallel to pieces like Alvin Lucier's *Music for Solo Performer*, where he used sonification of his brain waves played through loudspeakers placed on acoustic percussion instruments. The move beyond conventional loudspeakers in this project is perhaps also influenced by the general current *'hacker'* and *'maker'* trends of building physical musical contraptions, made possible by the increasing availability and ease of use of cheap and lightweight microcontrollers, advanced electronics and emerging technologies like robotics and machine learning etc.

In relation to this technological context, it is clear that this project is also about developing new means of musical expression, but in addition to this technological side it has the explicit aim of exploring a particular musical material coming from outside the realm of the instrument itself, in a way that goes beyond the mere *instrument-inventor* approach. Even though the instrument part is crucial in the project as the main method and tool for actually realizing the music, the project is not just about the possibilities of this particular instrument, but what it can reveal about the speech/music relationship.

Musique concrete: 'reality' as music

With this project's use of real-life recordings and the emphasis on authentic situations, there is also a possible reading in connection to a genre of music called *Musique Concrete.* This genre emerged as the antithesis to the highly abstract electronic music created in early years of electroacoustic music studios. In contrast to these artificially synthesised sounds, this *concrete* music was instead based on recordings of reality, from ambient soundscapes, machine sounds and traffic, and human activity like speech and conversation. One relevant reference in this regard is the music of Luc Ferrari, which often plays with gradual transitions between realistic documentaristic field recordings and subtle musical treatment and transformation of these soundscapes based on the musical associations and possibilities they offer.

This play with reality, and the musical perception and analysis of that reality, is something that also plays a part in this project, and which also can be linked to the focus on perception of reality in the music of Ablinger as described above. But the intention and content of this reality is very different – the music of Ferrari often revolves around very site-specific field recordings, creating narratives of particular places and times, or in the case of the piece Far West News, taking the almost journalistic form of a radio travelogue from a trip across the American south-west. In contrast to this, I have tried to erase all explicit references to specific times and places, and instead sought the common features and the generality of expressions linked to certain types of

social relations and interaction, to see how this relates to the wordless non-referential but presence-based interaction of improvised music. Which brings us to the context of improvised music.

Improvisation as dialogue

A context that might not be the most obvious judged just from the appearance of the performance concept itself, but which is actually most in line with the performative approach in this project, is that of *free improvisation,* where the music is not based to any particular genre but created and shaped through improvised musical discourse. *Free improvisation* refers here to improvisation as a method and not a particular genre, in the tradition that George Lewis has categorised as the *eurological* perspective (Lewis, 1996), as opposed to the *afrological* perspective of the much more politically charged *free jazz* movement with its roots in the civil rights issues of 1960's America.

In his thesis *From Small Signs to Great Form*, improviser and saxophonist Njål Ølnes goes to great lengths to show how the dialogical communication in free improvisation actually works in practice, analysing in minute detail how musical signs or *gestures* are used in the collective development and shaping of larger forms, of musical *gestalts* (Ølnes, 2016). The aural approach that Ølnes emphasizes as a fundamental aspect of improvisation is also central in this project, approaching and interacting with speech gestures primarily by ear and with the immediacy that this approach provides, just like one does in spoken conversation.

From an overall perspective, the situation occurring when musicians improvise together can obviously result in a multitude of different musical outcomes, music that does not by any means have to resemble a conversation or dialogue. All the semiotic potential of any sound structure is available for the musicians to create music that can sound like anything and "be about" anything. Yet, the situation itself, improvising something together with others by means of sound in real time, is fundamentally the same as in any spontaneous conversation, even if cause and expression is totally different. As the work of Ølnes shows, improvising musicians make statements with sound, and one way or another have to relate to their own and other musician(s) utterances (and *not* relating is also an act of communication in that regard). The same apply to structural aspects of improvisation, including the continuous negotiation of what it is about and in which direction it should go, with past experiences influencing the expectations of *what* can happen and *how* it will develop. The difference of course being that with music there is an aesthetic purpose and public framing that dictates certain kinds of roles and modes of interaction. Nevertheless, it is still a *social* situation where musicians have to relate both to the audience and to their fellow musicians' sounding utterances to interpret intentions and ideas. Therein, I think, lies the similarity to spoken conversation and the possibility of

experiencing improvised music as dialogically meaningful, as the exchange and collective development of musical ideas.

This is one reason for why I think it is interesting – as set out in the aims described in the background chapter – to juxtapose musical imprints of spoken conversation with improvised musical interaction, highlighting these similarities and the implicit discourse apparent in improvisation.

Beyond this general dialogical aspect of improvisation, there are also more direct references in the work of several improvisers that also have introduced varying degrees of speech or speech-like sounds in their music, like for instance the invented poetic languages in Sidsel Endresen's vocal improvisations or the multitude of approaches to speech and song used in the performances of the composer, performer and improviser Maja Ratkje. A more direct link to sound poetry and Dadaism can also be found in the vocal improvisations of Jaap Blonk. Text and sound poetry is also present in the work of pianist Sten Sandell, who has explored speech as an integral part of instrumental improvisation in a performative exploration of the space between text, voice and music (Sandell, 2011, 2013).

The way the present project differs from these performative approaches, is perhaps most distinct in the explicit focus on *everyday conversations* as the actual musical material, shifting the focus away from the expressive performer on stage to the dynamics and expressivity of everyday spoken interaction. This introduces a very different framing, opening the closed abstract musical reality of the performance space to the trivial experiences of everyday life.

Sound art

With the presentation of a sound installation, in addition to the already latent conceptual implications of using documentaristic recordings, another possible reading could be to view this as *sound art*. There are different opinions of the boundaries between sound art and music, but sound art is here understood as the sonic version of conceptual art which historically is more related to the conceptual tradition of visual arts than with music.

This connection to sound art is perhaps the least relevant for this project. As described in the chapter reflecting on sources and concepts, an overall conceptual approach was considered but effectively abandoned in this project as it so clearly conflicted with the focus on the here-and-now of the improvised situation and the musical reality that is manifested through this situation. Other, more formal concepts like the use of speech genres are still present, but play a lesser role in the overall picture and does not make it into sound art in my opinion.

Contexts – conclusion

As this review of different possible readings of the musical results shows, the music might be experienced in very different terms depending on the particular focus of the listener, each perspective relating to different sets of relevant references, expectations and contexts. The multidisciplinary nature of the project makes all of these contexts relevant to some extent, so there will not be only one correct reading. Still, as this review shows, some are more relevant than others.

Critical aspects

If we step back from the contexts and related fields discussed above and take a critical look at some underlying assumptions in this project, one aspect that really should be problematized it is its use of *recordings*. The very performance concept itself is based on a paradox of wanting to explore the ephemeral *presence*-based prosodic phenomena of everyday conversation by way of improvisation, but in order to avoid acted or staged conversation, doing so using *recordings* of past conversations. This obvious paradox of using recordings to study real time interaction can seem like trying to square the circle. The most intuitive way to explore speech in improvised music would be like Sten Sandell, to seamlessly integrate the act of speaking into the action repertoire of the instrumentalist improviser. Likewise, the most intuitive way to make music based on recordings would be by way of electroacoustic composition, staying both in the same temporal domain as well as the same medium of recorded sound. But with my particular interest in the dynamics of everyday conversation and their obvious parallels to musical improvisation, I found that developing some way of improvising with this recorded source material was the only way I could explore this connection without at the same time ending up with theatrical drama performances, which would be the result if I had to rely on performers enacting these conversations. That is why exploring *how* to integrate the musical exploration of everyday speech in a performance concept based on improvisation became one of the central aims in this artistic research project.

Contribution

As is clear from the above discussion of contexts, this artistic research project is not situated within *one* homogenous field, and so its contributions cannot be defined in relation to one single context.

With regards to the context of instruments, the system demonstrates an alternative way of using speech than many existing instruments based on voice synthesis in a more traditional direct gesture-to-sound instrument paradigm, as well as the more general-purpose sampler-instruments frequently used with speech sounds by many vocal performers. Theoretically, this instrument could also be used with other sources such as animal sounds or other phenomena, but many of the analysis parameters (such as segmentation based on syllables etc.) and other functions of the system in general have been made specifically to deal with the particular features of speech, so the system would not seem like a good tool for working with other sound sources.

From the perspective of improvisation, I believe this project represents a novel way of introducing musical material that is sourced from recordings, in such a way that its parts and features become a main content of the improvised discourse.

Recordings have of course been used in improvised music before with samplers, turntables or computers, but in this project recorded speech have played a much more fundamental part for shaping the music and providing the raw material and structures for creating layers and arrangements of new musical structures that goes beyond the mere use of recordings as aural quotes.

On the other hand, the topic of composition has not been the main focus in this project, but to the degree that the approaches to making music from speech through improvisation in this project also can inform compositional practices, I believe that for instance some of the ways that machine learning has been used to reorganise speech segments in musical ways can represent an interesting example also for this field.

But if we return to the fundamentally cross-disciplinary nature of this project, with its implied connection to linguistics, improvisation, different technologies and compositional approaches, and contemplate the paradox described above – how even the performance concept's foundation represents an attempted merging of very different musical practices. Has this worked out in practice then, have I managed to find a way to overcome this paradoxical starting point?

It was not a given that this would work, but I believe that the results clearly demonstrate a convincing way of doing this, a way that has proved a productive and interesting new approach for making music. As described in the reflections on instrument development, this was achieved largely by two factors – by making an interactive instrument system for playing with speech recordings that also was responsive to sound input, and by developing an appropriate hybrid electric-acoustic performance concept that merged the different sonic realms of recorded speech and electronic and acoustic instrumental sound. This made it possible to place the piano at the centre of this exploration and yet base the whole concept on recorded speech, and the music created in this project is a direct result of these developments.

In terms of innovation, I believe that it is precisely these solutions to the paradoxical combination of employing electronic processing of speech recordings in an acoustic improvised setting that represents the *main contribution* of this project. These results show one model for how this can work in practice, a model that can be of value as a reference for further related work of artistic explorations into the multi-disciplinary field of mixing technology, linguistics, speech, music, improvisation and interaction.

APPENDIX

Artistic Research – challenges in a new academic field

This essay describes some of the conditions and circumstances leading to the recent emergence of *"artistic research"* as an international academic discipline and field in general, and the establishing of the doctoral level *Norwegian Artistic Research Programme* in 2003 in particular. The ideological trends and political changes preceding the reforms of education structures during the 1990s serves as a backdrop for discussing the challenges that the artists entering this new and still evolving field are facing. Both regarding questions of relevant content, methods and formats, but also to gain a better understanding of the underlying premises and foundation for this practice based research discipline.

Societal and political changes

Before dealing with the challenges of the emerging field of artistic research itself, it is necessary to take a closer look at the historical development of education policies and the political and societal changes that have taken place in Europe the last 70 years, and which can shed some light on the foundations and inherent challenges of this new field.

According to Guro Hansen Helskog's account of the historical changes ahead of the far-reaching reforms of education structures in Norway during the 1990s (Helskog, 2003), several factors that can be seen as influential on the political need for these reforms. Similar changes took place in the rest of Europe after World War II, so to some extent this can be seen as representative of a general trend in many countries. Factors include the huge rise in students completing higher education during the 50s and 60s, dramatically increasing government expenses for education. In addition to the prevailing ideas of equal access to education, personal growth and democratic development, this was perhaps also due to the overall emphasis on economic and welfare growth after the war, and the technological competition with the Soviet system for which the space race during this period can serve as an excellent example. In Norway, a number of new colleges were established to cater for the huge rise in student numbers, often built in small towns scattered around the country in accordance with the political ideal of decentralisation. After growth started slowing and industry started declining, economic uncertainty made itself a reality during the 70s and 80s and several things happened. One thing was the post-industrial introduction of the idea of the *knowledge society* that were to follow the previous industrialised one, and another was the now all too familiar talk of the need for *austerity policy* to keep expenses down in the new economic reality. This need to keep costs down while still giving priority to education meant that education structures had to be organized more efficiently. From the late 80s on and through the 1990s this led to a series of

extensive education reforms in Norway, inspired by the ideas of new public management spreading through all parts of state administration. As part of this still on-going process of increasing efficiency by mimicking corporations, most small colleges were merged into larger entities like universities and university colleges. Thus, former small specialized and independent knowledge cultivating colleges were turned into bigger centralized knowledge production facilities, resulting in standardisation of education structures, goals and organisation. The affordance of still growing student numbers were surely helping to democratize higher education, but in the new knowledge economy the stated political goal was now to serve the businesses' need for qualified workers. These changes involved a shift in the view of education and knowledge. From the late 19th century humanistic formation viewpoint – education as a right for individuals and better knowledge for the greater good of democracy, towards a view of higher education as a way of making sure the workforce's full potential could be exploited, at the same time turning knowledge into a commodity that could contribute to the overall production and economic growth. This is illustrated by the title of the Report to the Parliament (Stortingsmelding) no. 43 (1988/89): *"More Knowledge to More People",* by then minister of education Gudmund Hernes.

These events using knowledge as an economical driving force can also be related to the changes in academia and research communities described by Michael Gibbons and colleagues in "The new production of knowledge" (Gibbons et al., 1994). They identified a shift towards what they dubbed a "Mode 2" way of conducting research, differing from what they saw as a traditionally autonomous *basic research* kind of knowledge production, which they then classified as "Mode 1". According to this description, the traditional Mode 1 research is characterised by being carried out in an academic context (without concern for practical application), within individual disciplines and by homogenous groups of participants, autonomously and without social accountability, and assessed through traditional quality control with peer reviews. In contrast to this, the authors claimed that there was a major shift towards research that is highly interactive and socially distributed. Characteristic of this mode of knowledge production is that it is carried out in context of application, is often trans-disciplinary in scope, produced by a diverse and heterogeneous mix of organisations both inside and outside academia, reflexive and accountable in their relation to society, and assessed through novel forms of quality control not only limited to disciplinary peers.

This trend and its description has been widely debated, and though the concepts have been adjusted by the original authors and critiqued and reconsidered by others (Hessels & van Lente, 2008), it seems to be consensus that there is a general trend towards a more diverse mix of different ways, places and actors engaged in research. The debate was also somewhat affectional as there seemed to be a conflict of values involved, threatening the hegemony of "proper" scientific basic research. As Donald E. Stokes points out (Stokes, 1997) this view that the pure basic research is of a somehow higher standard and hence higher value, can be traced to the research policies laid

out in post war America. According to this account, the success owing to the scientific progress during World War II especially in nuclear physics inspired policymakers to keep and further increase funding for basic research during peacetime. This was both to encourage economic growth and to gain the upper hand in the ideological and technological competition with the soviets. Vannevar Bush, director of the wartime Office of Scientific Research and Development was the one tasked with working out the post war science policies by US president Roosevelt. He was the one who coined the term *basic research* and he postulated its superiority by insisting that the creativity of basic science would be lost if hindered by practical ends. These thoughts had an important influence on the scientific boom during the 50s and 60s and set the premises for later policies and ideas about science, where basic research is the supplier of the knowledge taught in universities. This applies still after the turn to more diverse and application oriented research and when knowledge is seen as rivalling industry as an economic factor – research is still seen as the knowledge basis for higher education.

Emergence of a New Field

Following the events described above, the new and larger education structures that resulted from the 1990s college mergers in Norway brought about a whole new institutional environment for higher education in the field of visual arts, music, dance, theatre, and design and crafts. This is also true for the development in many other European countries.

In Norway, many private run arts schools had been professionalized as colleges a few decades earlier during the 70s and 80s. In Trondheim for instance, Trøndelag Music Conservatory became a college in 1973 when the former privately owned Trondhjem Music School founded in 1911 was split into a municipal music school for children of school age and a state-run Music Conservatory for higher education. Similarly, the Trondheim Academy of Fine Arts established in 1951 became a state-run college in 1979. As the education reforms of the 90s attempted to find more efficient ways of administering the many colleges that had been created in Norway from the 60s on, arts colleges were included in the mergers into larger entities and because of that also affected by the law of 1995 on higher education, specifying that universities and university colleges should provide higher education *and research* on an international level. Thus in Trondheim, Trøndelag Music Conservatory and Trondheim Academy of Fine Arts became parts of the newly created Norwegian University of Science and Technology with the 1996 merger of the Norwegian Institute of Technology, the Norwegian College of General Sciences, the Museum of Natural History and Archaeology, and the Faculty of Medicine ("NTNUs historie," n.d.).

Part of the process of increasing efficiency was standardizing the different education structures of the former independent colleges, and so the former art colleges comparable to professional schools were transformed into

university departments subject to the university education system where research was now one of their duties.

This process of standardization of higher education was mirrored by a political drive within the European Union to include education in the free flow of a common inner market. With the economical aim of "facilitating mobility" of work forces, education ministries in Europe started a process of harmonizing the structures of higher education in the late 1990s, creating a framework for standardizing education credits, levels and grades. This process is known as the Bologna Process, named after the initial accord signed by the European education ministers in Bologna in 1999 ("BOLOGNA - an overview of the main elements," n.d.). Part of this standardization was the specification of a three-tier division of higher education into a first cycle (bachelor degree), second cycle (master's degree) and third cycle (doctoral degree). This meant that by definition, higher education was to include doctoral degrees based around research, and so the higher art education institutions had to come up with ways to implement this.

According to Torsten Kälvemark's account in his chapter on University politics in "The Routledge Companion to Research in the Arts" (Kälvmark, 2010), this was the direct cause behind the establishing of many doctoral programmes in artistic research in Europe. In the US, a system with "professional doctorates" had been practiced for decades, awarding doctoral degrees based on artistic achievements when hiring professors and treating professional work in the arts as a *research equivalent.*

In Europe and Australia, the move was instead to try to broaden the definition of research. This was however controversial both in academia and the field of arts education. For one thing, in the authoritative classification of science disciplines developed by UNESCO in 1979 to facilitate comparison of science statistics, the definition explicitly excludes *artistic research of any kind.* This classification became the de facto standard and lived on in the framework used by the Organisation for Economic Co-operation and Development (OECD) as laid down in the *Frascati Manual* that in its latest edition from 2002 still explicitly excludes artistic research *of any kind* (Borgdorff, 2012). Based on this it is no surprise that many academics were critical of artistic practice being classified (and even worse – funded) as research. On the other hand, many in the field of art education were likewise sceptical of having artistic practice adhere to academic requirements. From a broader perspective, the trend of treating artistic practice as research was perhaps not so foreign to artists, at least for the visual arts field that have been working conceptually for a long time. To explore new frontiers was an important part of modernism, and the term '*research*' was already being used by curators and other actors in the art world to describe what artists were doing. Still, as Kälvmark concludes, the establishment in Europe of an full-fledged artistic research field *within* academia was by and large, with the possible exception of Finland, primarily a

consequence of the reforms of education structures described above (Kälvmark, 2010).

Following these developments, doctoral and research programmes in the arts started to appear in many countries such as the UK, Sweden, Norway, Holland, Austria as well as the ones already established in Australia. All followed different paths in the challenging task of trying to protect the requirements of artistic practice while at the same time adhering to academic standards. In Norway, the Ministry of Education directed two studies during 1999 in order to work out how to establish a doctoral degree in the arts. Representatives from the Art academies considered the programmes in place in Sweden and the UK, and then proposed to establish a national fellowship programme for artistic development, common for all arts education institutions. Cautious that the standard academic PhD model would provide too strict a framework for doing research through art practice, they instead recommended to set up an independent programme, *on level* with the doctorate but not linked to any degree. This position at the outskirts of academia helped ease potential tension between the two spheres that were to be brought together. In 2003 this turned in to reality when the Norwegian Programme for Artistic Development was launched, admitting about 6 fellows a year drawn from all arts fields and giving associate professor competence though without awarding a formal PhD degree – at least for the time being (Malterud, 2012).

Artistic Research, *why, what* and *how*

As these accounts show, establishing *artistic research* as an academic field in Europe was a pragmatic result of the changing education policies and of a general professionalization of society giving more people higher education, as well as the inner market standardization of education structures within the European Union. So, without really having to address *why* artistic research should exist as a field, artists in academia now face the questions of *what* artistic research is supposed to be and *how* it should be carried out. This has been constantly debated for the past 15 years or so, among others by Henk Borgdorff who was involved in setting up art research programs in Holland and who has taken upon him to review the relation between artistic research and academia and map out a framework for the new field. Through his articles written between 2004-2011 and compiled in the book *"The Conflict of the Faculties"* (Borgdorff, 2012) he paints a thorough vision of what artistic research ought to be and, at least from a theoretical point of view, how it can be conducted in academic terms. In addition to argue for the inclusion of art education amongst the PhD-awarding "higher" faculties (paraphrasing Immanuel Kant's call for inclusion of natural sciences, humanities and philosophy amongst the then higher faculties of theology, law and medicine in 1798), Borgdorff goes to great lengths to define when art practice should qualify as research as compared to other artistic practice and traditional academic research. He bases his criteria on the existing definitions of

research used by funding bodies such as the UK Research Assessment Exercise and Arts and Humanities Research Council, which in short describe research as *'an original investigation undertaken to gain knowledge that is then effectively shared'.* This is similar to the ideas in other official definitions such as the one used by OECD and in the set of definitions set out in the Bologna accord's so-called *Dublin descriptors,* and which then again is reflected in the current Norwegian law on universities and higher education.

Following these undertakings of Borgdorff and other academics in the field during the last decades then, a coherent theoretical framework seems to be in place for this emerging field of practice based artistic research. There is no need to keep arguing for its legitimacy as its existence is now a fact. For artists and scholars involved though, having artists conform to academic standards without reducing the artistic validity still poses a series of challenges, and there is still much tension involved in both directing and describing such research.

From a practical point of view there is the lack of defining previous work usually to be found when a new field is established from within. With few preceding examples, there is also little common methodological knowledge to draw upon and artists are instead encouraged to invent their own methods, as each project is so different. Having to inventing everything from scratch leads to high risk when it comes to the success of the result, both as art and research, and the newness of the academic framing can still inflict much uncertainty as the focus of the masters and bachelor tiers of arts departments is still more akin to the professional schools from which they originated.

Theoretically and from a philosophy of science point of view, there is the problem of identifying with either of the two modes of knowledge production described by Gibbons et al. It is also a open question if the field can relate to the ethical norms originally formulated by Robert K. Merton and known under the acronym *CUDOS (communism, universalism, disinterestedness, originality, scepticism)* (Merton, 1973), which otherwise can be seen as hallmarks of the success of the scientific method. Even the term '*research'* is still difficult. The Norwegian programme avoids this science-laden term in its Norwegian name and instead uses the less controversial '*Development' ("Program for Kunstnerisk Utvikling"),* while native English speakers seem to prefer the descriptive *Practice Based Research* over the perhaps pretentious sounding *Artistic Research* used elsewhere in Europe. Also, practical questions about how such research is to be assessed is not completely clear. Borgdorff points to the academic model of peer review, but who are qualified peers in a completely new field? Finally, even Borgdorff concedes that even after the shift towards Mode 2 and more diverse ways of conducting research has made it possible for academia to include the non-conceptual insights articulated through artistic practice, there is still a fundamental unease in the relationship of art practice and academia that is not simply transitory. He relates this to the fact that art in being a reflective practice might have more in common with a more *un-academic speculative philosophy.* Art invites much

thinking but carries no definite thought expressible in language. This he calls the *radical contingency* of artistic research – the possibility of the un-thought, that which is unexpected, and this is what makes it so difficult of having it conform to traditional academic conventions and definitions.

Back to the *why* of artistic research

Even if this debate about what artistic research is and how it can be conducted has been going on for the last 15-20 years and the issues is by many seen as settled (interestingly enough mainly by academics like Borgdorff), I still want to dwell a bit more upon the fundamental basis for this new field. Perhaps that is just a necessary exercise for an actor in the middle of an artistic research project in order to navigate and prioritize in this new and unfamiliar academic landscape with few points of reference. But maybe it is also necessary to still keep an eye on the justifications of *why* we should have something called *artistic research* in the first place, even as the field establishes itself and becomes more mature.

This might be repeating some of the arguments heard in the aforementioned debate, but the intention is to once more reflect on the very foundations of the field.
First of all, something that is very foreign to the themes normally associated with artistic practice is the economic goals and market ideology present in the ideas about *production* and knowledge as a *commodity* in the efficiency reforms leading to the creation of *artistic research* as a discipline. But aside from that, the field is established on the academic premise that *research* is the basis for development of the knowledge and the teaching and training in a field. While this might be true for most other academic disciplines, is it really applicable for art practices? Do not innovative artists drive the development within their field regardless of framing their art as research or not? Can perhaps all art practice be seen as research then? There are already several mechanisms in place for funding artists in Norway and elsewhere that also encourages innovation and artistic development, and there is also the American concept of *research equivalence* to deal with this more institutionally. In that light, the introduction of artistic research into academia could be seen as having more to do with the political-bureaucratic need to measure and systematize, and the spread of the so-called *"Management by Objectives"* system originally developed by the Australian organisation consult Peter Drucker and now spreading to all spheres of society after being introduced as a management principle by the Norwegian government in 1990s (Helskog, 2003).

But if we accept the present way of organising society by defining artistic development as research and assigning it to academia, does artistic practice really have to adhere to the official definitions of research requiring that it must be an original investigation leading to new shareable insights in order to

qualify as the basis for teaching and training? Are not these characteristics already a part of all artistic practice, which to a certain extent always is an original expression of a content (*knowledge* if you will) that is shared with an audience? On the other hand, if these definitions are really needed to discern *artistic research* from other artistic practice, is there not a danger that this will lead to an artificial divide between *academic art* on one hand and something perceived as *real art* on the other, or that you could have the situation where a work is deemed good as research but not very interesting as art? One could also argue that if we are to take the premise of research as basis for knowledge and training really seriously, then artistic research should perhaps also be open to those leading artists who are unable, unwilling, or just not interested in adjusting to traditional academic conventions, such as the practices of reflecting and discoursing through words. Unlike other academics, artists often have less interest of articulating their insights through spoken or written language, as their artistic medium is perceived as much better at both dealing with and communicating this (for instance, painting is a much more nuanced tool than words when it comes to dealing with colours). It is also not, as some academics seem to think, just a matter of artists picking up some theory at the door when entering academia. Thinking, conceptualizing, and the academic craft of reading, reflecting, writing and discoursing critically is a *practice* of its own, and an altogether different one than the *artistic* practice. Professional artists that have spent years training and developing their work cannot simply slide into this language based tradition unnoticed, the same way that a highly trained academic cannot just pick up some musical scales and then walk on to a concert stage and perform.

For this reason, the decision to establish the Norwegian artistic research program outside of the traditional university system and PhD degree seems like a good idea, at least for a start, to try and define for itself what artistic research can and should be.
It might also be that the perceived differences are not that fundamental. Maybe it is just a case of striking the right balance between practice and reflection, and that this holds true even for art outside academia. As James Pritchett recounts in his excellent book *"The Music of John Cage"* (Pritchett, 1996), the American composer John Cage was at first dismissed by critics when departing from the likeable style of his early prepared piano pieces. But after the publication of his essays and lectures giving a direct account of his thinking and the rationale behind his music and development, the critics praised his efforts even if they personally did not like the music.

As these thoughts illustrate, there are so many similarities between artistic and academic practice yet they somehow seem to be very different. That makes it hard to say something definitive, and that is precisely why it is important to keep these things in mind when shaping the field from within through practical artistic research work.

Now, what the academic frame of artistic research *can* offer the otherwise competitive and sometimes overly individualistic world of art (in addition to some much-needed funding and long-term working conditions), is the institutionalized tradition of constructive discourses with fellow peers, of playing with open cards so to speak, sharing insights and being honest about inspirations and indebtedness, also raising awareness about both one's own and other disciplines. This can result in very different and broader presentation of artistic practice than traditional venues and channels, and as such it can constitute a welcome addition to the traditional art practices.

Science and research have always been changing, and this new field claiming its own methods of knowledge production and quality control, is perhaps just representative of the general professionalization and the shift towards new contexts of knowledge production in our society, as described in the concept of a "Mode 2". In fact, the creation of a new yet not fully defined discipline is seen as an advantage by some, as in the editorial statement of the "Journal of Artistic Research (JAR) in 2011: *"Not knowing what exactly artistic research is, however, is a good thing for a number of reasons"* (M. Schwab, 2011)*.

To conclude this essay into the challenges and opportunities of the new academic field of artistic research, only time will show if it can mature into a format consistent with academic traditions without breaking off from the rest of the art world. The field is anyway already well established, so now it is up to the artists involved to grapple with the tension and unease arising from the construct, and work out in practice *how* it can be conducted and *what* it can be.

Technical background

This appendix provides some additional technical background for the particular signal processing techniques implemented in the software instrument system, as described in the instrument overview.

Signal Processing

A wide range of available approaches for managing and processing sound in general and speech in particular is relevant in this regard. Primarily from the field of electronic music but also from the fields of speech signal processing used in telephony, speech recognition, speech modelling and linguistics. As much more detailed descriptions of various common techniques for signal processing can be found in other well written sources like *"The computer music tutorial"* by Curtis Roads, only brief descriptions will be given here.

Analysis/re-synthesis

Particularly useful when dealing with a defined source material like speech is the *analysis/re-synthesis* approach, which includes a range of different ways to analyse, process and then resynthesize sound from a given source (Roads, 1996).

Linear Predictive Coding

One such approach is the *source-filter model* of speech production, that has proved useful in speech signal processing (Fant, 1960). According to this model, speech is treated as a combination of a *source signal* (the vibrating vocal cords/glottis pulse train) and an *acoustic filter* (the vocal tract, including throat, mouth, lips and nasal cavity). The filter part has typically been approximated with the *linear predictive coding (LPC)* analysis technique that makes an estimation of the filter spectrum based on the difference between the (relatively) slow movements of the filter relative to the much faster pulses of the glottis (Atal & Hanauer, 1971). This simplified model of the speech organs allows analysis and treatment of *voicing* separate from the *formants* - the peaks in the spectrum that characterises different vowels. Because of its compact representation it has long been in use to encode and synthesise speech digitally with a low bitrate, for instance in voice over IP (VoIP) telephony. The LPC technique was also picked up early by composers working with speech and computers, such as Charles Dodge and Paul Lansky.

Fourier Transform

An analysis/re-synthesis technique more commonly used in musical applications is the short-time *spectrum*. Typically obtained through the *fast Fourier Transform (FFT),* which transforms short slices of *sound waves* (air pressure variations in the time domain) into frames of *frequencies and amplitudes* (amplitude differences in the frequency domain). This information can in turn be used for additive synthesis of individual partials allowing a wide range of processing techniques that affects both partial frequencies and spectral shape (Roads, 1996).

Cepstrum

Another way of approaching the source-filter division is the technique of *cepstral smoothing* (Smith III, 2014). In this technique, *another* Fourier transform is performed on a log-scale representation of the short time spectrum itself, resulting in a kind of a *spectrum of the spectrum.* This imaginary domain has been dubbed the *cepstrum,* which is just an anagram of the word 'spectrum' (Bogert, Healy, & Tukey, 1963). One can view the cepstrum as a description of the *shape* of the original spectrum, as if the spectrum was a signal frame in the time domain. Filtering out higher bins (called *quefrencies*) in this cepstrum, and inverse-Fourier transforming it back to the spectral domain, results in a *smoothed* spectrum (less jagged and with fewer peaks), which like the LPC spectrum can be used as a filter or for detecting formants. Another cepstrum-based technique that must be mentioned in relation to speech processing is that of *mel-frequency cepstrum coefficients,* known under the acronym MFCC (Mermelstein, 1976). A MFCC is the *cepstrum* of the *mel-spectrum,* which is a spectrum with an alternative frequency scale better suited to represent the formant regions most important for speech. It is a very robust and compact way to describe only those parts of the spectrum that is important for discerning phonemes, and therefore very common in automatic speech recognition applications. The MFCC technique is very powerful as a spectral descriptor, but in the analysis/synthesis approach adapted in this project it has been used only tentatively, and mostly explored in relation to syllable segmentation (see for instance the blog post http://orchestraofspeech.com/blog/syllable-detection/).

Re-synthesis

Regarding the synthesis stage of this approach, *resynthesizing* the processed results from such analyses back into sound can be done in several ways as well. In particular, the overlap-add (OLA) technique of resynthesizing signal slices, obtained through the inverse FFT of spectral frames, has proved an efficient way of synthesizing large numbers of partials and noise components at the same time (Rodet & Schwarz, 2007). This technique, in addition to its pitch-synchronous variant (PSOLA) (Moulines & Charpentier, 1990), allows for a wide range of possible transformations and abstractions of the same input/output-chain and is one of the main synthesis techniques used in this system.

Corpus approach and machine learning

In addition to such signal processing techniques, some overall approaches for organising recordings and data have also been influential in the development of this instrument. In the statistical approaches widely adopted in corpus linguistics and speech recognition applications, large numbers of recordings are organised as whole bodies – *corpora* – of analysed segments. By looking at the corpus as a whole, the *relationships* between its elements can more easily be explored. Such approaches have been applied successfully in digital musical instruments as well, as in the *audio mosaicking* and *concatenative synthesis techniques* developed in Diemo Schwarz' *"CataRT"* instrument (Schwarz et al., 2006).
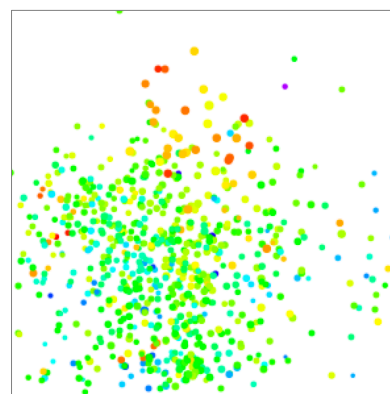
Segments as coordinates in a space of mean pitch (vertical) speech rate (horizontal) and vocal effort (colour)

This opens up for a much more musical way of using this material, with variable degree of fragmentedness and removedness from the original speech structures. One possibility is the exploration of fragments that occupies the same area in the prosodic space, creating sequences that make more sense musically than based on the lexical content and thus shifting listening focus to their musical structures. This can involve repetition and progressive variation of shorter or longer segments, more in line with a typical musical exploration of this material.

[time]	duration	pitch	loudness	effort	slope	rate	index	file
794.2456	32.35876	278.1527	18.17515	−5.543281	−7.528183	571.2523	0	1
899.278	69.94509	228.8727	18.24911	−21.83022	44.37697	380.3114	1	1
1057.043	45.62107	220.168	13.49872	−27.06207	35.22589	1023.868	2	1
1115.645	22.53215	444.0258	12.61825	−60.56793	−5.662277	212.428	3	1
1398.093	52.84219	151.421	20.23517	4.106823	0	1135.454	4	1
1450.936	101.151	148.2932	26.78622	16.03532	−21.46344	167.7052	5	1

Excerpt from database of analyzed segments

An extension of the database approach is the use of *machine learning* typically found in automatic speech recognition. Machine learning is a huge field by itself, also for pattern recognition and generation in interactive music systems. It was never meant to be the main focus of this project, it has proved a useful influence for introducing improvisational elements like *interactivity* and the *unknown response* into this project.

References

Ablinger, P. (n.d.). Voices and Piano program note. Retrieved December 1, 2017, from http://ablinger.mur.at/docs/voices_engl_dt.pdf

Astrinaki, M., D'Alessandro, N., Reboursière, L., Moinet, A., & Dutoit, T. (2013). MAGE 2.0: New Features and its Application in the Development of a Talking Guitar. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Daejeon, Republic of Korea: Graduate School of Culture Technology, KAIST. Retrieved from http://nime.org/proceedings/2013/nime2013_214.pdf

Atal, B. S., & Hanauer, S. L. (1971). Speech Analysis and Synthesis by Linear Prediction of the Speech Wave. *Journal of the Acoustical Society of America*, *50*, 637–655.

Austin, J. L. (1962). *How to do things with words*. Cambridge, Mass: Harvard University Press.

Bailey, D. (2004). Free Improvisation. In D. Warner & C. Cox (Eds.), *Audio Culture: Readings in Modern Music* (pp. 255–265). New York: Continuum.

Bakhtin, M. M. (1986). The Problem of Speech Genres. In *Speech Genres and Other Late Essays* (pp. 60–102). Austin: University of Texas Press.

Bartel, D. (1997). *Musica poetica : musical-rhetorical figures in German Baroque music*. Lincoln: University of Nebraska Press.

Barthes, R. (1977). Musica Practica. In *Image Music Text: Essays selected and translated by Stephen Heath*. London: Fontana Press.

Beller, G. (2009). *Analyse et Modèle Génératif de l ' Expressivité, Application à la Parole et à l'Interprétation Musicale*. (Doctoral thesis). Universite Paris VI, Paris.

Beller, G. (2014). The Synekine Project. In *ACM International Conference Proceeding Series*.

Beller, G., Schwarz, D., Hueber, T., & Rodet, X. (2005). Hybrid Concatenative Synthesis On The Intersection of Music and Speech. In *Journees d'Informatique Musicale* (pp. 41–45).

Beller, G., & Aperghis, G. (2011). Gestural Control of Real-Time Concatenative Synthesis in Luna Park. In *P3S, International Workshop on Performative Speech and Singing Synthesis*. Vancouver, Canada.

Berio, L. (1974). A-ronne (author's note). Retrieved from http://www.lucianoberio.org/node/1420?1747386730=1

Bernstein, L. (1976). *The Unanswered Question: Six Talks at Harvard*. Cambridge, Mass: Harvard University Press.

Bogert, B. P., Healy, M. J. R., & Tukey, J. W. (1963). The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In *Proceedings of the symposium on time series analysis* (Vol. 15, pp. 209–243).

BOLOGNA - an overview of the main elements. (n.d.). Retrieved April 8, 2015, from http://www.eua.be/eua-work-and-policy-area/building-the-european-higher-education-area/bologna-basics/Bologna-an-overview-of-the-main-elements.aspx

Bonds, M. E. (1991). *Wordless Rhetoric. Musical Form and the Metaphor of the Oration.* Cambridge, Mass: Harvard University Press.

Borgdorff, H. (2012). *The conflict of the faculties: perspectives on artistic research and academia*. Leiden: Leiden University press.

Bosetti, A. (n.d.). Mask Mirror. Retrieved June 12, 2018, from http://www.melgun.net/live-projects/mask-mirror/

Brandt, P. A. (2006). Form and Meaning in Art. In M. Turner (Ed.), *The Artful Mind* (pp. 171–186). Oxford: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195306361.003.0009

Brandtsegg, Ø. (n.d.). Cross Adaptive Processing as Musical Intervention; Exploring Radically New Modes of Musical Interaction in Live Performance. Retrieved June 13, 2018, from http://crossadaptive.hf.ntnu.no/

Brandtsegg, Ø. (2007). *"New creative possibilities through improvisational use of compositional techniques, - a new computer instrument for the performing musician."* Norwegian University of Science and Technology, Trondheim. Retrieved from http://oeyvind.teks.no/results/

Brandtsegg, Ø., Saue, S., & Lazzarini, V. (2018). Live Convolution with Time-Varying Filters. *Applied Sciences*, *8*(1). Retrieved from https://doaj.org/article/b0e2759b1d0d4954a89d7e489f357005

Brunson, W. (2009). Text-Sound Composition – The Second Generation. In *Proc. of EMS-09 Conference on Electronic Music Studies*.

Cook, P. R., & Lieder, C. N. (2000). SqueezeVox: A new controller for vocal synthesis models. In *ICMC*.

Delalez, S., & Alessandro, C. (2017). Vokinesis : syllabic control points for performative singing synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 198–203).

Fant, G. (1960). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. The Hague, Netherlands: Mouton.

Fasciani, S. (2014). *Voice-Controlled Interface for Digital Musical InstrumentS*. PhD Thesis, National University of Singapore.

Fels, S. (2004). Designing for Intimacy: Creating New Interfaces for Musical Expression. *Proceedings of the IEEE*, *92*(4), 672–685. http://doi.org/10.1109/JPROC.2004.825887

Fels, S. S., & Hinton, G. E. (1998). Glove-Talk II - a neural-network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, *9*(1), 205–212. http://doi.org/10.1109/72.655042

Fels, S. S., Pritchard, B., & Lenters, A. (2009). ForTouch: A Wearable Digital Ventriloquized Actor. In *NIME* (pp. 274–275).

Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, *60*(6), 1497–1510.

Feugère, L., D'Alessandro, C., Doval, B., & Perrotin, O. (2017). Cantor Digitalis: chironomic parametric synthesis of singing. *EURASIP Journal on Audio, Speech, and Music Processing*, *2017*(2). http://doi.org/10.1186/s13636-016-0098-5

Frayling, C. (1993). *Research in Art and Design*. *Royal College of Art Research Papers Series* (Vol. 1). London: Royal College of Art. Retrieved from http://www.opengrey.eu/handle/10068/492065

Gazzaniga, M. S. (1967). The Split Brain in Man. *Scientific American*, *217*(2), 24–29. http://doi.org/10.1038/scientificamerican0867-24

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge*. London: SAGE.

Godøy, R. I., & Leman, M. (Eds.). (2010). *Musical Gestures: Sound, Movement, and Meaning*. New York: Routledge. http://doi.org/10.4324/9780203863411

Goodman, N. (1976). *Languages of art: An approach to a theory of symbols*. Indianapolis: Hackett.

Harnoncourt, N. (1982). *Musik als Klangrede : Wege zu einem neuen Musikverständnis*. Salzburg: Residenz.

Helskog, G. H. (2003). Den humanistiske dannelsen og 1990-tallets utdanningsreformer. *Norsk Pedagogisk Tidsskrift*, *87*(01–02).

Hessels, L. K., & van Lente, H. (2008). Re-thinking new knowledge production: A literature review and a research agenda. *Research Policy*, *37*(4), 740–760.

Janer, J. (2008). *Singing-driven interfaces for sound synthesizers*. Universitat Pompeu Fabra.

Johnstone, K. (1981). *Impro : improvisation and the theatre*. London: Methuen.

Kim-Cohen, S. (Ed.). (2009). *In the blink of an ear : towards a non-cochlear sonic art*. New York: Continuum.

Kälvmark, T. (2010). University Politics and Practice-based research. In M. Karlsson, Henrik, Biggs (Ed.), *The Routledge Companion to Research in the Arts* (pp. 3–23). New York: Routledge.

Lane, C. (2006). Voices from the Past: compositional approaches to using recorded speech. *Organised Sound*, *11*(1), 3–11.

Lane, C. (Ed.). (2008). *Playing with Words*. London: CRiSAP.

Leeuwen, T. van. (1999). *Speech, Music, Sound*. London: Macmillan Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Mass: MIT Press.

Lewis, G. E. (1996). Improvised Music after 1950: Afrological and Eurological Perspectives. *Black Music Research Journal*, *16*(1), 91–122. http://doi.org/10.2307/779379

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae*, *3*(1_suppl), 29–57. http://doi.org/10.1177/10298649000030S104

Malloch, S., & Trevarthen, C. (Eds.). (2009a). *Communicative musicality: Exploring the basis of human companionship*. Oxford: Oxford University Press.

Malloch, S., & Trevarthen, C. (2009b). Musicality: Communicating the vitality and interests of life. In *Communicative musicality: Exploring the basis of human companionship* (pp. 1–11). Oxford: Oxford University Press.

Malterud, N. (2012). Artistic research – necessary and challenging. *INFormation, Nordic Journal of Art and Research*, (1).

Margulis, E. H. (2014). *On repeat: How music plays the mind*. New York: Oxford University Press.

Mattheson, J. (1739). *Der vollkommene Capellmeister*. Hamburg: Christian Herold.

Mermelstein, P. (1976). Distance measures for speech recognition, psychological and instrumental. *Pattern Recognition and Artificial Intelligence*.

Merton, R. K. (1973). The Normative Structure of Science. In *The sociology of science: theoretical and empirical investigations* (pp. 267–280).

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Miall, D. S., & Dissanayake, E. (2003). The poetics of babytalk. *Human Nature*, *14*(4), 337–364.

Monelle, R. (1992). *Linguistics and Semiotics in Music*. Harwood Academic.

Moore, F. R. (1988). The dysfunctions of MIDI. *Computer Music Journal*, *12*(1).

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5–6), 453–467.

NTNUs historie. (n.d.). Retrieved from http://www.ntnu.no/ntnus_historie

Obin, N., Dellwo, V., Lacheret, A., & Rodet, X. (2010). Expectations for Discourse Genre Identification: a Prosodic Study. In *Interspeech-2010* (pp. 3070–3073). Retrieved from http://www.isca-speech.org/archive/archive_papers/interspeech_2010/i10_3070.pdf

Obin, N., Lacheret-Dujour, A., Veaux, C., Rodet, X., & Simon, A. C. (2008). A method for automatic and dynamic estimation of discourse genre typology with prosodic features. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* (pp. 1204–1207).

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), pp.35-45.

Pritchett, J. (1996). *The Music of John Cage*. Cambridge University Press.

Radiolab. (2006). Musical Language [Audio Podcast]. New York: WNYC Radio. Retrieved from http://www.radiolab.org/story/91514-sound-as-touch/

Roads, C. (1996). *The computer music tutorial*. Cambridge, Mass.: MIT Press.

Rodet, X., & Schwarz, D. (2007). Spectral Envelopes and Additive + Residual Analysis/Synthesis. In *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 175–227).

Rousseau, J. J. (1781). Essai sur l'origine des langues, Ou il est parle de la Mélodie & de l'imitation Musicale. In *Œuvres posthumes de J.J. Rousseau* (le Pléiade, pp. 371–429). Genève.

Rubidge, S. (2005). Artists in the academy: reflections on artistic practice as research. In *Dance Rebooted: Initializing the Grid Conference Proceedings*. Retrieved from http://ausdance.org.au/articles/details/artists-in-the-academy-reflections-on-artistic-practice-as-research

Sandell, S. (2011). Music inside the Language [CD]. Steninge, Sweden: LJ Records.

Sandell, S. (2013). *På insidan av tystnaden : en undersökning*. (Doctoral thesis). Konstnärliga fakulteten, Göteborgs universitet, Göteborg.

Schaeffer, P. (1966). *Traité des objets musicaux*. Paris: Le Seuil.

Schnell, N., & Battier, M. (2002). Introducing Composed Instruments, Technical and Musicological Implications. In *Proceedings of the international conference on new interfaces for musical expression* (pp. 156–160). Dublin, Ireland.

Schnell, N., Borghesi, R., & Schwarz, D. (2005). FTM : Complex Data Structures For Max/Msp. In *Proceedings of International Computer Music Conference (ICMC)*.

Schnell, N., Röbel, A., Schwarz, D., Peeters, G., & Borghesi, R. (2009). MUBU & friends - Assembling tools for content based real-time interactive audio processing in MAX/MSP. In *Proceedings of International Computer Music Conference* (pp. 423–426).

Schwab, M. (2011). Editorial. *Journal for Artistic Research*, (0). Retrieved from http://www.jar-online.net/issue-0/

Schwab, M., & Borgdorff, H. (Eds.). (2014). *The exposition of artistic research: publishing art in academia*. Leiden: Leiden University Press.

Schwarz, D., Beller, G., Verbrugghe, B., & Britton, S. (2006). Real-time corpus-based concatenative synthesis with catart. In *9th Int. Conference on Digital Audio Effects (DAFx)* (pp. 279–282). Montreal, Canada.

Searle, J. R. (1969). *Speech acts : an essay in the philosophy of language*. Cambridge: Cambridge University Press.

Shklovsky, V. (1965). Art as Technique. In L. T. Lemon & M. J. Reis (Trans.), *Russian Formalist Criticism: Four Essays* (pp. 3–24). Lincoln: University of Nebraska Press.

Sloboda, J. (2012). *Exploring the Musical Mind: Cognition, emotion, ability, function*. Oxford: Oxford University Press.

Smith III, J. O. (2014). Cross Synthesis Using Cepstral Smoothing or Linear Prediction for Spectral Envelopes. Retrieved December 4, 2017, from http://ccrma.stanford.edu/~jos/SpecEnv/

Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, *112*(12), 1025–1058.

Stokes, D. E. (1997). *Pasteur's quadrant: basic science and technological innovation*. Washington, D.C.: Brookings Institution Press.

Szczepek Reed, B. (2011). *Analysing conversation : an introduction to prosody*. Basingstoke: Palgrave Macmillan.

Trifonova, A., Brandtsegg, Ø., & Jaccheri, L. (2008). Software engineering for and with artists: a case study. In *Proceedings of the 3rd ACM International Conference on Digital Interactive Media in Entertainment and Arts (DIMEA'08)* (Vol. 349, pp. 190–197). Athens, Greece.

Vincent, M. (2010). *Music & Language Interrelations*. (Doctoral thesis). University of Toronto.

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

Wishart, T. (1994). *Audible design : a plain and easy introduction to practical sound composition*. Orpheus the Pantomime.

Wishart, T. (1996). *On sonic art*. New York: Routledge.

Wishart, T. (2012). *Sound Composition*. Orpheus the Pantomime.

Wittgenstein, L. 2001. (1953). *Philosophical Investigations*. Blackwell Publishing.

Ølnes, N. (2016). *From Small Signs to Great Form - Analysis of the musical interplay in free improvisation, using the tools of Aural Sonology*. Norwegian Academy of Music, Oslo. Retrieved from https://brage.bibsys.no/xmlui/handle/11250/2381966

Åhlberg, L.-O. (2014). On Form and Content. In *Notions of the Aesthetic and of Aesthetics : Essays on Art, Aesthetics, and Culture* (pp. 123–141). Frankfurt: Peter Lang.

Resources / Downloads

The "Orchestra of Speech" instrument system software

The system needs a copy of *Max* to run, and requires IRCAM's *FTM* and *MuBu* external libraries to work:

Max: http://cycling74.com/products/max

FTM : http://ftm.ircam.fr

MuBu: http://forumnet.ircam.fr/product/mubu-en/

The *Orchestra of Speech* instrument system software:

http://orchestraofspeech.com/wp-content/uploads/orchestraofspeech.zip (~5MB)

Source code for custom MIDI controller

Based on the Arduino-compatible Teensy microcontroller board.

Source code:

http://orchestraofspeech.com/wp-content/uploads/controller_v2_2.ino_.zip

Conference Paper

Paper for the International Conference on New Interfaces for Musical Expression (NIME) 2018.

"The Orchestra of Speech – a speech based instrument system":

http://orchestraofspeech.com/wp-content/uploads/The-Orchestra-of-speech-a-speech-based-instrument-system.pdf

Research blog

http://orchestraofspeech.com/blog/

Speech Music Symposium videos

http://orchestraofspeech.com/blog/speech-music-symposium/