Estonian Academy of Music and Theatre

**Tze Yeung Ho**

# Phonetic Features as a Tool in Music Composition:

## From the Perspective of a Composer

This thesis is submitted in partial fulfilment of the requirements for the Doctor of Philosophy (Music) degree.

Supervisor: Prof. Jaan Ross

Tallinn 2021

## Abstract

My artistic research in music composition examines the two phonetic features of *quantity* and *tones* as tools for generating musical materials. This dissertation establishes a theoretical framework in the symbiosis of the documentation of my creative process and the phonetic analyses of the Cantonese, Norwegian and Estonian languages. My personal ties to these languages gives way for musical examples to be drawn from my own artistic output. In the scope of the research, seven subtopics are highlighted: 1.) 'Musical Interpretation of Articulative Properties in Speech', 2.) 'Quantity Distinction, Rhythm and Formal Construction', 3.) 'Pitch-Quantity Phenomenon', 4.) 'Relativity of Pitch in Tonal Languages and Its Impact on Song', 5.) 'Composing with Norwegian Pitch Accents versus Cantonese Tones', 6.) 'Pitch in Speech and Its Musical Representation' and 7.) 'the Composer and the Performance of Songs Set to Texts of Tonal Languages'. The interdisciplinary approach explores how knowledge in phonetic features of quantity and tones has contributed to my creative process in composing the song cycle *natt-öö-夜*, the electroacoustic tone poem *isjoriske forsteder,* the string quartet *vihik (a)*, the miniature mezzo-soprano concerto *vihik (c)*, the trilingual choral work *vihik (d),* the bilingual miniature for three voices *kj/ærlige ord*, the secular *messe norvégienne profane* and the staged works *hvorfor pusen?* and *minn(i)e*. The dissertation reflects on the compositional theories and practices developed in the course of the research in a broad discussion across disciplines such as general linguistics, poetics, score analysis, phenomenology and musicology.

**Table of Contents**

# 1. Introduction

The research subject embodies my personal experiences as a composer and a speaker of multiple languages. Raised in a Cantonese-speaking home between Toronto and Oslo, I was exposed to the nuances of various languages during my formative years. This exposure had been a vital source of ideas and inspiration for creative work from the time when I decided that I would pursue a career as a composer. My interests in languages further prompted me to pursue studies in Finno-Ugric languages during my undergraduate studies, adding both Estonian and Finnish. I was fascinated by the abstract differences in the songs sung in these different languages towards the end of my bachelor studies in Toronto. I began to formulate the various prototypes of the research questions explored in this thesis: beyond the texts themselves, why do Cantonese songs sound different from Estonian songs? What about Norwegian songs? What impact does speaking these languages have on writing music?

These questions developed over the years alongside my aspirations in composing. It eventually occurred to me that these open questions have seeped into my own musical creations. I began to ask: how I can incorporate these languages in a unified work to represent a part of my personal identity and musical voice? This initial inquisition became the prototype of a musical experiment tied to the resulting works of this research.

Already in my first years of composition studies, I recall that my then teacher in composition — Canadian composer Alexander 'Sasha' Weinstangel — spoke about the differences in orchestration between early 20th century French and German composers were results of their spoken tongues.[1] At around the same time this conversation occurred, a scientific inquiry on this very discussion was just published in psychologist Aniruddh Patel's ambitious text *Music, Language and the Brain* (2007). The second chapter of this book, titled *Sound Elements: Pitch and Timbre*, examines the subject of native speakers of specific languages and their mental framework in appreciating and understanding music related (or unrelated) to their languages. These ideas sparked my interest in pursuing a deeper study on *how* or *why* that is the case from a creative perspective.

[1] Taken from a personal conversation.

Exposure to more research, literature and conversations since my teenage years had led me to the linguistic domain of phonetics: the study of speech sounds. One could discuss language on social, cultural and personal levels. Language will always remain to be a vast concept. Language generates, embodies and embraces an array of discussions about communication through the means of speech, scripts, movements, metaphors and so forth. The Finnish cultural theorist Mikko Lehtonen suggested that "language is a material and practical activity" (Lehtonen 2000: 66). In this thought, I was able to limit the scope of defining language and focus on specific facets of the subject. The material of sound in language — or speech sounds — has been a fundamental element I use in my practice as a composer, as song is a practice utilising the same human voice apparatus. Phonetics has thus drawn my attention as a means to decipher the building blocks of how each language engage in its rules with sounds and music.

This dissertation should not be read as a scientific research on phonetics. This work is an artistic research with a focus on phenomenological approaches. The phonetic phenomena have been interpreted as theoretical tools for composers rather than close-ended conclusions of how languages and music should or do interact. My research has examined some possibilities to create or recreate speech sound materials in a musical context in the compositional processes of select works. Parameters — tone and quantity — where speech sounds meet in music and phonetic theories have been discussed. In the case of the languages I am engaged with (i.e. Scandinavian languages, Cantonese, Mandarin, Estonian and sometimes Finnish), there are two phonetic features where they often cross paths in my music: *quantity* and *tone*. The definition of these terminologies will be found in the corresponding chapters in this thesis.

My research aims to offer some insights in answering these two questions:

- How can understanding the phonetic features of *quantity* and *tones* help me, the composer, in generating musical materials, creative tools or ideas?
- How can I develop tools to understand speech sounds' influence on my creative process as a composer?

The interdisciplinary method of this research focused on musicological and phonetic practices, both fields trace their historical roots to phenomenology. The concept of 'research through practice' plays an important role within this research, as my personal compositional practice generates the core of the musical and phonetic inquiries shown in the seven thematic essays in the body chapters (**Chapters 3** and **4**) of this dissertation. Phenomenology is defined as the "study of structures of consciousness as experienced from the first-person point of view" (Smith 2018), and this very thought is embedded in the work on phonetics solidified by linguist Ferdinand de Saussure in the late 19th century. The research, however, deviates from general linguistics in its literature and context as experts within the branch of phonetics such as Peter Ladefoged and John Laver in the twentieth century have emerged to take on a quantitative approach to the study of speech sounds. The emergence of phonetics as an individual branch within linguistics has led to cerebral studies on certain speech sound subjects such as pitch and quantity. These two parameters are quantifiable acoustic phenomena. The laboratory experiments on speech sounds have though been argued as unnatural, as participants of speech experiments tend to speak in an artificial manner within such environments. Nonetheless, such studies have been fruitful for specific languages in different regards, particularly in understanding less mutable elements, such as quantity distinction within Finnish and Estonian and pitch variance within Chinese languages.

While compositional tools and ideas are not only developed from analysis of quantitative data, observing the trends concluded in these scientific studies can be enlightening in finding another perspective in discussing and understanding music, particularly those involving the use of speech sounds. In the phonetic discussions, the research draws from literature such as *The Temporal Structure of Estonian Runic Songs* (2001) by Jaan Ross and Ilse Lehiste, and *The Realization of Tone in Singing Cantonese and Mandarin* (2013) by Murray Schellenberg and *Principles of Phonetics* (1994) by John Laver in order to discuss the language-specific concepts within phonetics. The theoretical framework of phonetics will be in dialogue with musical concepts, as the two domains share various acoustic phenomena (e.g. pitch, volume, duration) as comparable parameters.

In the creative interpretation, the study of poetics is taken into account during my compositional process. I have placed special focus on texts such as the *Analysis of the Poetic Text* (1976) by Yury Lotman in my works with Estonian texts. The partial turn to literary theories generates not only a discourse between cerebral and creative understandings of speech sounds but serves also as a reminder that the composer — i.e. a first-person perspective in the creative process

— may interpret knowledge from different sources and ontological focuses. The same thinking can be applied to literature from musicology and score analysis, as perspectives on interpreting other composers' works versus one's own works may sometimes be antithetical to the scientific approach of phonetics.

The first-person perspective of the research with myself as the composer is in constant dialogue with phonetic theories and concepts surrounding the phenomena of the two phonetic features, *quantity* and *tone*. Observations, ideas and theories are put into practice through experimentation in my works. The resulting performance and experience offers feedback to evaluate the experiments, which calls for further experimentations and new attempts. Contrary to scientific experiments, the reproducibility of the phenomenon in a performance may not be a factor for evaluating the experiment's success. Rather, if the resulting experiences of the listener and performers are stimulating new ideas for discussion, it would be sufficient to generate a discourse in aesthetics (e.g. why the choice of these speech sounds), techniques (e.g. how to work with these speech sounds) or style (e.g. how the choice of these speech sounds compare to other composers) for working with speech sounds as musical materials. This meeting point between speculations and phenomena is where my artistic work generates artistic results. The uncertainties, feedback and opinions of the listener motivate musical creations in different directions. The terminology of 'experiment' used in the context of my composition is not one that seeks to prove or disprove phonetic theories — it aims to explore more creative possibilities.

The experimental approach taken in this research (i.e. composing in dialogue with the phonetic theories) is a common practice in artistic research projects conducted in the recent years. To see it one way, music composition is the design of sound in time. "[T]he possibility of design research being done on the basis of design practice or through practice, i.e. artistically/creatively making objects, interventions, processes etc. in order to gain knowledge" (Bang, Krogh, Ludvigsen & Markussen 2012) is one way to describe approach of this hands-on method of my research. My compositional experiments are tied new musical, notational, practical and phonetic questions. Some processes may aim to enrich my compositional understanding of interaction between the two phonetic features and composition and others may affect the resulting performances and the subsequent rehearsal techniques. The processes of my research can be visualised as the following diagram:

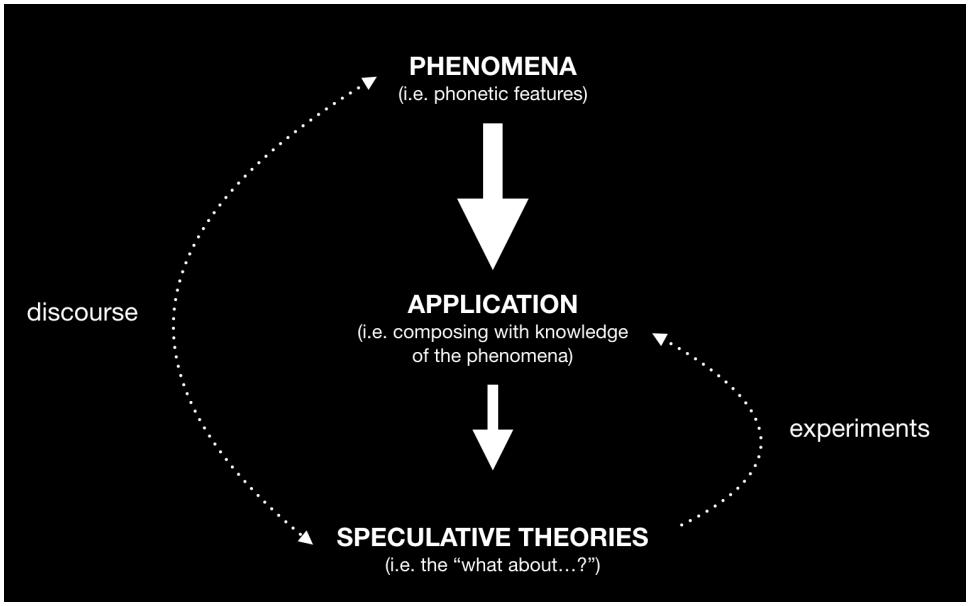



**Figure 1.0** — A visualisation of the different actions in this research.

The practice (i.e. compositional process) and phonetics theory discussed in this research are synthesised by concept of *play* derived from Hans-Georg Gadamer's theories in aesthetics. In the Stanford Encyclopaedia of Philosophy, this concept of play bears the following description:

"Spectatorial participation (like much art research) demands immersion in that which cannot be fully anticipated or controlled by individual consciousness. The game and the artwork are both forms of self-movement which require that the spectator play along with what they bring into being" (quoted after Gadamer 1986: 23). Gadamer asserts the "primacy of the play" over consciousness: "the players are merely the way the play comes into presentation" (quoted after Gadamer 1989: 92, 98). Participation takes the individual players out of themselves. The individual subject is that upon which success, satisfaction or loss is imposed from within the game. By analogy, the work of art is also "the playing of it". An autonomous event comes into being, something comes to stand in its own right which "changes all that stand before it" (quoted after Gadamer 1986: 25). Like the ancient *theoros*, the spectator not only participates in the event which is the artwork, but is potentially transformed by it (quoted after Gadamer 1986: 24)."

In my understanding, the artist approaches this concept of play through the 'working of the material' over a period of time. Texts and the phonetic potentials in my dissertation are interpreted as the 'material' in my compositional process. The material is to be 'transformed' by both the works' creator(s) and the spectator (listener). This can be interpreted by the free compositional

process with regards to the knowledge of phonetics. Music serves as the agent of the text and the transformational process. This means that the feedback from each performance will generate more potential materials for further inquiries and creative impulses.

It is important for the reader to note here that spectrographic analysis[2] is not utilised in the discussion of the resulting compositions of this research. I have chosen to omit methods of spectral analysis in discussing tone and quantity in my own music throughout my studies because the research focus is on the compositional *process* (i.e. using phonetic concepts as a means to create musical materials) and not on compositional results (i.e the reproduction of specific sounds or musical elements in more than one work and/or performance). In considering Gadamer's concept of *play*, phonetic concepts and my creative choices are taken part in the 'game' which eventually result in a composition. However, the scientific phonetic concepts may deteriorate in my deliberate misinterpretation, metaphor-making and re-imagination. This is very well a part of my artistic exploration. Perhaps it may seem irrelevant to phoneticians and linguists why scientific subjects are discussed or understood in this way but my compositions are undeniably born from the borrowed discussions on tones and quantity. This dissertation serves as an explanation for how the works are materialised, even if the phonetic concepts are not proven scientifically. I reiterate that this dissertation focuses on developing my compositional tools from phonetic features rather than proving any sort of scientifically correct universals. Concepts such as narrative, semantics and analogous topics which are most often byproducts of the work will be covered briefly in the limited scope of this thesis.

—

This dissertation is divided into several sections, starting with this current chapter, an Introduction (**Chapter 1**), which has now outlined the scope, research questions and methodology. To give the reader a historical moment from which this topic emerges and a rudimentary understanding of the relationship between phonetics and music, "Speech and Music: Starting with the *Sprechstimme* Dilemma" (**Chapter 2**) will problematise speech in a well-known compositional context, drawing particularly from the example of Schoenberg's *Sprechstimme* technique in his piece *Pierrot Lunaire*.

[2] Though this subject will be briefly touched upon in **Chapter 3** on the subject of quantity, as the temporal representation of Estonian syllabic durations is best demonstrated visually.

The body chapters of the dissertation will focus on the two phonetic features *quantity* and *tones*. To maintain focus in two overarching subjects, I have chosen to highlight seven relevant topics which have been salient in my creative practice as a composer. *Quantity* (**Chapter 3**) is divided into three subsections, with an introductory chapter offering a foundational understanding of *quantity* for the reader and provides examples of the theoretical concept. The three following subsections (**Chapter 3.1 - 3.3**) discuss topical issues related to composition and quantity; the issues are described at the end of the chapter introduction. The second body chapter (**Chapter 4**) discusses four topics on the phonetic concept of *tone* in the perspective of a composer (**Chapter 4.1 - 4.4**). Each subsection of the body chapters concludes with a point-form summary of the ideas discussed.

The conclusion chapter (**Chapter 5**) evaluates the ideas, tools and discussions on how my own composition process has been affected or changed over the course of this research. More importantly, removing my personal connection to Norwegian, Estonian and Cantonese languages, I have discussed how other composers or artists may incorporate concepts from phonetics into their own works with other sets of languages. Next steps of continuing the research of topical studies will be broadly discussed, with suggestions of how such studies or compositional practices in this method can be further pursued, reflecting on my experience in this research.

As the collection of compositions deriving from this research has grown significantly, this dissertation is unable to provide each work with a thorough description or analysis under each subsection listed above. I felt that lengthy explanations of each work may detract the reader from comprehending the topical discussions. I have instead chosen to provide some pages on the background of the works discussed in these subsections (**Appendix 1**). The larger works may especially require more background information for better understanding of its compositional context with the phonetic features. To offer the reader greater insight into the context of the example musical excerpts, I have also included larger sections of each work in **Appendix 2**.

At last, I would like to extend my sincere thanks to the Norwegian Society of Composers, Norwegian Culture Fund, *Komponistenes vederlagsfond*, *Eesti kultuurkapital*, the Royal Norwegian Embassy in Tallinn, Nordic Culture Fund, Nordic Culture Point, Tuuli Tuisk (University of Tartu), Elisabeth Hetherington, Hans-Gunter Lock, Professors Jaan Ross (EAMT), Toivo Tulev (EAMT), Kristel Pappel (EAMT), Helena Tulve (EAMT), Martti Vainio (University of Helsinki) and my

family. Only through their generous support, wisdom, musicality and patience could this research and its affiliated artistic projects be made possible.

## 2. Speech and Music: Starting with the *Sprechstimme* Dilemma

"Music aspires to be a language without intention. But the demarcation line between itself and the language of intentions is not absolute; we are not confronted by two wholly separate realms. There is a dialectic at work."

— Theodor Adorno, *Quasi Una Fantasia: Essays on Modern Music*

Phonetics is the study of speech sounds, which is an intrinsic part of understanding spoken language. Sung texts are frequently termed *lyrics* in music, while speech in a music performance remains to be conceived as speech. Regardless of the configuration, spoken language — written down, sung, implied or purely spoken — is a vital agent for bringing fully og a part of musical experiences to listeners. Singing and speaking both utilise the same human voice apparatus. Spoken words in communicating information could very well be sung. But how can we truly distinguish the two functions? Echoing the Adorno quote which opened this chapter: where does speech end and music begin? This chapter aims to challenge the reader to consider the human vocal apparatus as a single entity for sound making, which encompasses both speech and song. In my opinion, the segregation of the functions (i.e. speech versus song) may hinder the creativity of a composer, as all sounds can be considered materials of expression in music. Not only as simply sound materials, I would suggest that all utterances generated through our voices imply abstract structures like musical forms, since we could understand them in varying parameters like duration, pitch and stress.

"Given the prime communicative function of spoken language, one discipline that lies at the heart of any adequate study of speech is linguistics. But speech, […] is a carrier of more information than solely the meaningful patterns of individual utterances of spoken language" (Laver 1994: 2). Responding to this notion of the extra-semantic function of speech, the composer has regularly exploited this noted extra information for musical purposes. For example, word painting in madrigals of Renaissance composers, flow in rap music, prosodic meters' influence on rhythm in *Lieder* and analogous topics are some discussions among musicologists, artists, listeners, performers and composers engaging with the human voice. To find a common language in these discussion, we have to zoom in to the deeper realm of phonetics, where "[a] primary task of phonetics is […] to provide an objective description of speech" (Laver 1994: 4). For the composer, these descriptions are clues to understanding how speech sounds are imagined and created by its utterer and perceived by a listener. The composer who understands phonetics may become inventive

in their approach with words, playing with pitch, stress, rhythm and other parameters shared between music and speech. If these many parameters are common between the two fields, then what truly separates speech from song?

This question was — and for many still *is* — infamously brought to the forefront for composers and performers of Western art music alike in the early twentieth century. One culprit was Arnold Schoenberg (1874 - 1951) and the deed was the enigmatic *Sprechstimme* vocal technique in his work *Pierrot lunaire* (1912). The classic example of Schoenberg and *Pierrot lunaire* serves as a springboard to my own considerations of speech sounds in music, as the reader will discover that the phonetic features of tone and quantity are inevitable discussions in this paradox. Schoenberg's work asks of the singer to sing the passage in a speech-like manner, abandoning the distinction between speech and song. But what is this *speech-like* manner of singing? Taking into consideration that both song and speech share the common human voice apparatus, what were the parameters Schoenberg considered when making this distinction in composing the work?

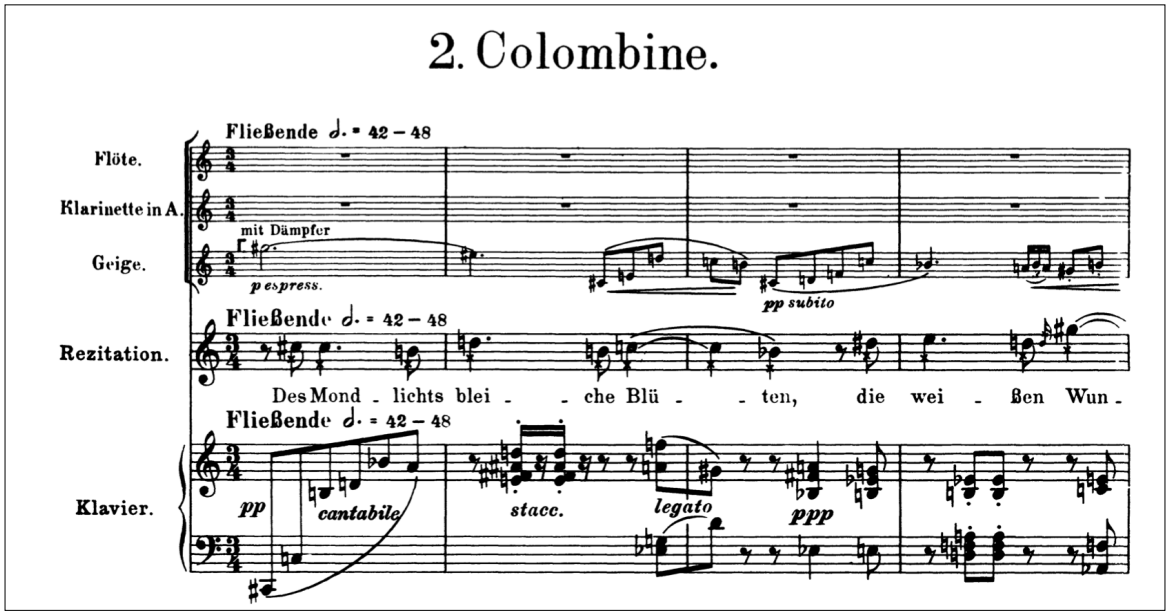



**Figure 2.0** — Excerpt from Arnold Schoenberg's "*Pierrot lunaire*" (1912) with the indication of *Sprechstimme* in the vocal line as 'x' on the stems of the notes.

In Avior Byron's article *The Test Pressings of Schoenberg Conducting Pierrot Lunaire: Sprechstimme Reconsidered*, it has been said that "both Darius Milhaud and Pierre Boulez, who conducted the piece, described it as creating 'insoluble problems'" (Byron: 2006). The article compares historic performances and recordings of the work in order to seek out Schoenberg's

intentions for this style of performance in *Pierrot lunaire*. While the goal differs from my research, it has nonetheless provided descriptions of how various performers, conductors and the composer himself interpreted the grey area between speech and song. Byron's article is full of metaphors and imageries by composers, conductors and performers alike and has provided some insights of what this speech versus song distinction may mean for the different personalities.

Byron considered the descriptions of Schoenberg's preface from the published score in 1914, which gives some illuminating ideas in the perspective of phonetics:

"The melody indicated by notes in the part of the speaker (with certain specially indicated exceptions) is *not* intended to be sung. The performer has the task of transforming it into a *speech melody* by taking well into consideration the indicated pitches. He can do this by

I. keeping to the rhythm just as precisely as he would when singing, i.e., with no more freedom than he would take in a sung melody;

II. being quite conscious of the difference between a *sung tone* and a *spoken tone*: the sung tone maintains its pitch without change, the spoken tone touches upon it but then leaves it immediately by descending or ascending. The performer must always be on guard against falling into a "singing" manner of speech. That is absolutely not intended. But neither should he aim for a realistic-natural speech. Quite the opposite, there should always be a clear difference between customary speech and speech that contributes to a musical effect. But this should never remind one of song." (Simms 2000, *from Byron's article*).

The important fragment from this description to be noted is, "[b]ut neither should he aim for a realistic-natural speech." Schoenberg stated clearly that *Sprechstimme* is *not* about recreating speech: this is a compositional — artistic — choice of a composer. Similarly to the disclaimer I have made in this thesis's introduction, neither Schoenberg nor myself has aimed to incorporate a comprehensive, scientific understanding of speech in music. In other words, the music takes precedence over the phonetic concepts. Instead of the emulation of speech, it is a question of how speech sounds lend their qualities to music. Here is my further attempt in deciphering the two parameters Schoenberg provided:

I would first like to unpack the second point in this passage on the matter of tone. The word, *tone*, has already lent itself to something directly related to my research. In Schoenberg's explanation, he has made the distinction of a *sung tone* and a *spoken tone*, with the former implying something stable and the latter being perturbed. Schoenberg has understood and shown that tone in

speech is something which fluctuates in the parameter of pitch (i.e. "the spoken tone touches upon it but then leaves it immediately"). This is true, as intonation in speech — dependent on the language at hand and German in *Pierrot*'s case — changes meaning or expression. In linguistic terms, intonation "serves to identify linguistic entities[3] at levels higher than the word, at the phrase and sentence level" (Laver 1994: 462). Furthermore, even if one is called on for repetition (e.g. reading the same text aloud), the speaker is unlikely to make the utterance exactly the same every time. "There is no recurrent biomechanical activity whose repetition is perfectly regular in its timing. The fundamental frequency of even a prolonged vocoid pronounced on a seemingly level monotone level in fact shows very small deviations […] from the the local moving average frequency […] running through the train of cyclic pulses of the vocal folds in voiced vibration" (Laver 1994: 453). Perhaps the *Sprechstimme* technique is striving for quality of pitch instability found in speech.

On the matter of tone, Pierre Boulez once commented that "[*Pierrot lunaire*] is [thus] both too high and too low," (Boulez 1986: 330). This remains to be a dubious claim, as the discussion of 'high' and 'low' lacks a context (e.g. high-low in terms of the singer's *tessitura*? or a range?) in which it functions. Furthermore, this is not the most phonetically comprehensive claim, as "the melody of a speaker's voice on any given occasion is not a matter of the absolute values of pitch displayed by the voice form syllable to syllable" (Laver 1994: 456). However, Boulez's following comment — "[…] when one speaks, the duration of the sound is usually short" (Byron 2006) — bears slightly more insight into the first point of Schoenberg's instruction: "keeping to the rhythm". In this instruction, the composer demanded the performer to respect the notated rhythms in the score as written. Contrasting with Schoenberg's expectation on speech tones versus sung tones, this request is not conducive to making the musical passages especially speech-like. It would be reasonable to believe that no German speaker would intuitively recite the text, without the sheet music of *Colombine* from *Pierrot lunaire*, to the tempo of 42 - 48 dotted half-notes per minute in the exact rhythms written by Schoenberg.

"Utterances are generally perceived as being spoken with a certain rhythm. The perception of rhythm in speech is predicated on the listener's recognition of a quasi-periodic recurrence in time of a given type of speech unit, such as syllables carrying peaks of prominence (achieved through either our both syllable stress or syllable weight), or syllables themselves" (Laver 1994: 512).

[3] Linguistic entities here is to be interpreted as the the vast possibilities of subjects in language such as semiotics, pragmatics, semantics, and so forth.

Without a doubt, Schoenberg's notation of *Pierrot* is not one of speech rhythm. It is a musical interpretation of the text's rhythm. But what is it that makes it *not* speech-like? Let us examine a larger unit of prosodic structures: stress. Returning to the excerpt form **Figure 2.0**, the syllabic segments *'Mond-', 'blei-', 'Blü-'* in the clause of '*Des Mondlichts bleiche Blühten,'* all bear heavier syllabic weight (i.e. more intensity and/or volume), which corresponds to how it would typically be spoken. From this short example, I suggest that we can assume the prosodic stresses notated in the score are natural with Schoenberg being a native German speaker and having a strong sense for its prosody. I must also note that some composers work against the 'natural'[4] flow of spoken language. An example of this anti-natural setting would be Igor Stravinsky and his treatment of the text in the opera *The Rake's Progress* (1951).[5] Stress and larger prosodic considerations are unlikely to be the causes of why *Pierrot* is not speech-like.

Beyond the syllable, the rhythmic problem of *Sprechstimme* in its relationship to natural speech may lie in the subject of quantity, or "phonological length" (Laver 1994: 436). When one speaks or reads a text passage aloud, the durational differences in the segments of words (e.g. syllables and even phonemes) seldom correlate to their sung form. In other words, singing the word *Mondlicht* in the tempo indicated by Schoenberg is not similar to how one would speak it. The absolute durational values of each syllabic segment would anyway be different in every repeated utterance. This can lead to a major point of contention among interpreters: While Schoenberg retained many speech-like qualities, such as allowing for pitch perturbations and maintaining stresses of the text, he did not do the same for durational values of the speech segments (i.e. syllables of the words). After all, Schoenberg was composing music and not writing poetry to be recited. This task of composition required a degree of artistic freedom. *Pierrot lunaire* is not about emulating speech, as Schoenberg stated, and this play with quantity in the in-between space of speech and song makes his compositional goals — for Boulez and many others — difficult to grasp. Boulez's comment of "[…] when one speaks, the duration of the sound is usually short" (Byron 2006) was in many ways a generally correct comment on why *Sprechstimme* is too abstract of a description in performance. Phonetic studies are though able to offer a more comprehensive

---

[4] As a human construct, language is not a natural phenomenon. *Natural* in this case denotes the commonly agreed upon practice, i.e. phonological standards of a language.

[5] Chandler Carter defends Stravinsky's text setting in his article "*The Rake's Progress* and Stravinsky's Return: The Composer's Evolving Approach to Setting Text" (2010). However, it was a common trope in my composition studies to understand it as 'unnatural' and 'impedes on the listeners' comprehension of the text'.

explanation on what constitutes the quality of speech, particularly through the understanding of durations of sound and quantity of speech segments.

Schoenberg's engagement with a speech-like form of singing, or *Sprechgesang*, is an artefact that is crucial to my own compositional practice. I have selected the *Sprechstimme* dilemma in *Pierrot* as an example for initiating the subject of phonetic features as a compositional tool because of the impact and response Schoenberg received since the work was premiered. The feedback all pointed towards how phonetic tone and quantity play a major role in vocal music. While it is not the most precise instructions written by a composer, the short preface to *Pierrot* has served as an important reminder that the parameters of speech sounds can be taken into compositional consideration. This far-reaching resonance of *Pierrot* has proven that the discussion of phonetics in conjunction with composing vocal music is not only historical but definitely one which is relevant to practising composers today. And not only for composers, "*Sprechstimme* in *Pierrot lunaire* engages the performer in action which connects both composer and performer in a fresh, mutual act of creation" (Byron 2006). Through bypassing the common, and sometimes stereotypical, distinctions of speech and song, composers, interpreters and their collaborators are situated in an experimental space where new techniques and expressions through the human voice may emerge.

Chapter Summary:

- Song and speech are tied together by the human voice. Although they utilise the same apparatus, song and speech have always been distinguished in discussions of music.
- Both music and speech sounds can be broken down into various comparable and measurable parameters, such as pitch and duration.
- Arnold Schoenberg's *Pierrot lunaire* is exemplified in this chapter because the *Sprechstimme* technique (to sing in a speech-like manner) was historically met with interpreters' different views on what constitutes 'speech-like qualities'. This thesis suggests that these qualities can be discussed through the phonetic features of tone (pitch) and quantity (duration of segments), which are often implied in this discussion of Schoenberg's composition.

## 3. Compositional Topics on Quantity

This chapter will focus on the phonetic concept of *quantity* and how it relates to varying compositional processes. This section will first introduce the theoretical concept through the lens of phonetics with broader references to music. Following subsections will explore three different topics on quantity which I have come across in the course of my artistic practice. As Estonian is one of the three focus languages in this dissertation and an integral part of my creative work, its quantity features will be exemplified, as "[t]he unifying principle of Estonian prosody is the contrastive use of the time dimension" (Ross and Lehiste 2001: 38). This makes Estonian spoken language a subject of interest for composers working in the parameters of time and rhythm. Though, this should not limit the reader in thinking that these exact concepts do not pertain to other languages as well.

---

*Quantity* in phonetics is defined as "phonological length" (Laver 1994: 436). As the syllabic segments of our speech may be measured in duration (milliseconds), the concepts of long and short can hence be derived. This differentiation of sounds in length measurements is a familiar idea to musicians, as rhythm in music is comprised of durational values. All speech sounds can be measured in absolute duration but this information tends to be purely statistical. Rather, "[c]ontrastive length is utilised by many languages as a phonological property of either or both consonants and vowels" (Laver 1994: 436). This means that, despite sharing similar phonemic structures, different meanings can be perceived, if the durations of one part or several parts of the utterance are lengthened or shortened. Some languages contrast only consonants, while others only vowels. Meanwhile, Estonian and Finnish "use both consonant- and vowel-length distinctions" (Laver 1994: 436). The implication here, for composers, is that if Estonian or Finnish texts are to be set to music, the listener's perception and comprehension of a text may be affected by the durational values. Readers should be aware of the terminologies in phonetics discussed here: *duration* are measured and absolute, while *length* is relative.[6]

[6] "For clarity, a distinction will be maintained here between **duration** as a phonetic feature, and **length** as a phonological feature manifested by relative phonetic duration. Another term often used for phonological length is **quantity**" (Laver 1994: 436). The reader should be reminded that *phonology* is the study of speech sounds pertaining to a specific language, whereas *phonetics* is the overarching field which examines speech sounds as a whole and encompasses phonological discussions as well.

"Estonian still is one of the few languages that differentiate between more than two degrees of length," (Lippus 2011: 18). In fact, the Estonian language distinguishes three quantities, *short* (commonly abbreviated as Q1, quantity 1), *long* (Q2) and *overlong* (Q3). Here is an example on varying the first [*a*] sound in the combination of 's-a-d-a':

| | **Short (Q1)** | **Long (Q2)** | **Overlong (Q3)** |
|---|---|---|---|
| **Word in Estonian Orthography** | sada | saada | saada |
| **Realisation of Word in International Phonetic Alphabet** | [sɑ.tɑˑ] | [sɑː.tɑ] | [sɑ::.tă] |
| **English Translation** | hundred | Send! | to get (infintive form) |

**Table 1.0** Comparison of the three Estonian quantities and the contrasts of semantics. [7]

To visualise the difference in absolute durational measurements, the software Praat [8]demonstrates in **Figure 3.0** the three quantity lengths of the same example as above. The three words are spoken by native Estonian speaker[9] and the durations in milliseconds are marked below each word. Q3 is marked as 'saaada' to distinguish itself from *saada* in Q2.

[7] This example is adapted from Pärtel Lippus's *The Acoustic Features and Perception of the Estonian Quantity System*, pp. 18.

[8] As Praat is a software created partly for spectrographic analysis, I would like to reiterate at this point that many of my compositional examples could have been discussed in this method. However, the disadvantage of such a method is that the discussion would then turn towards developing performance techniques (e.g. creating notation or instructions which help performers recreate sounds). This dissertation focuses rather on the more personal and abstract creative process of how phonetic concepts influence my writing process (i.e. interval choices, phrase durations, etc…).

[9] Courtesy of Katri Hiovain-Asikainen (University of Helsinki) for the *Praat* examples.

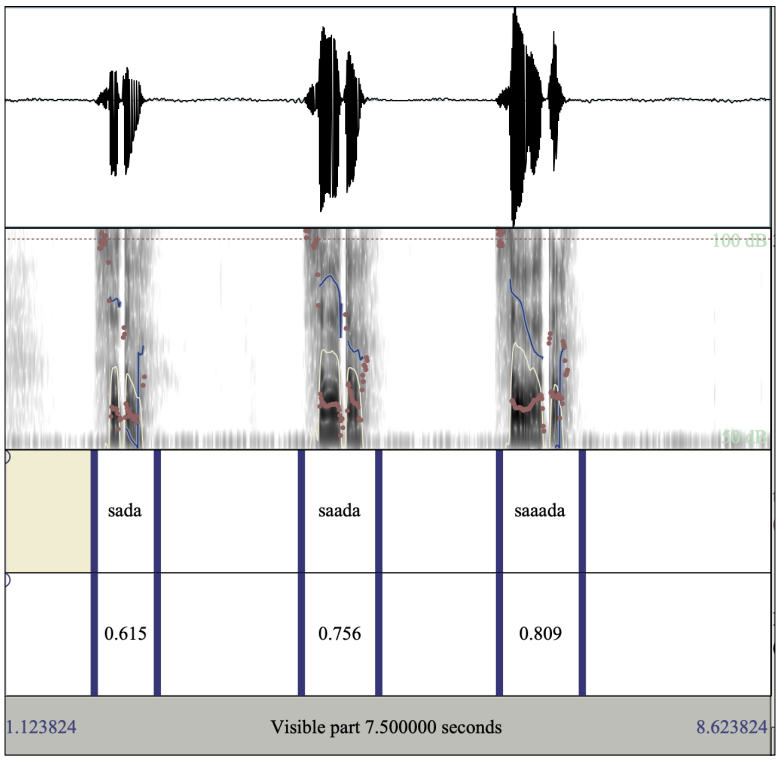



**Figure 3.0** A waveform graph (upper third), sonagraph (middle) and measurements (lower third) from the software *Praat* contrasting the three quantities from the words of Table 1.0: *sada* (Q1), *saada* (Q2) and *saada* (Q3). Note that the the waveform graph provides a visualisation of the frequencies, while sonagraph provides details such as pitch (in blue), formants (as red dots) and intensity (in yellow). The durational limits (i.e. where the word begins and ends) are shown in the number figures.

"Estonian, according to the traditional rhythmic dichotomy, is classified as a syllable timed language but is also said to be characterised by foot isochrony," (Asu and Nolan 2006). A language is deemed isochronous "with given units of speech recurring on a regular basis" (Laver 1994: 156). These units can be described as syllable-based, stress-based or mora-based (i.e. syllable weight). It has been suggested that the prosodic foot level of Estonian may be isochronous, this means that an average rhythm — or tempo on a larger level — could be derived from measuring a variety of speakers' utterance of syllable of the same vowel-consonant configuration. This is a consideration in the discussion of working in music with spoken Estonian texts, as isochrony is arguably something that is not speaker-specific and a universal feature of the language. Another consideration is that the stress is almost always placed on the first syllable,[10] this hinders the composer from using certain rhythmic and metric figures in speech-like musical setting of the texts. For example, anacruses are seldom if not impossible in certain types of texts.

[10] "Main stress in Estonian words is fixed on the first syllable which, together with the following unstressed syllable constitutes the domain of the three-way quantity contrast for which Estonian is famous" (Asu and Nolan 2006).

Ross and Lehiste in the book *The Temporal Structure of Estonian Runic Songs* has outlined a hierarchy of phonological levels, "segment, syllable, prosodic foot, phonological phrase and higher-level units equivalent to sentences and paragraphs" (Ross and Lehiste 2001: 37), in which variations of quantity occur within different contexts. Turning to spoken Norwegian, it has been said that the majority of its dialects require the stress on larger phonological units such as words and sentences to be 'immutably heavy.'[11] For example:

| **Vowel (V) - Consonant (C)** | **V > C** | **V < C** |
|---|---|---|
| ***Word in Norwegian Orthography*** | hvile | ville |
| ***Realisation of Word in International Phonetic Alphabet*** | ['ʋi:lə] | ['ʋil:ə] |
| ***English Translation*** | to rest | would *or* wild (pl. adj.) |

**Table 2.0** Examples in Norwegian of contrastive syllabic stress through phonemic lengths in the first syllable of two similar words.

The two words *hvile* and *ville* both share similar phonemic structures, as demonstrated in the International Phonetic Alphabet. The only visible distinction in this transcription is the position of where the weight — via quantity — is placed. In *hvile* the vowel [i] holds both the syllabic weight and therefore longer duration value. In contrast, the [l] segment bears the weight and duration in *ville.* In a context where these two words are sung on a longer note, this would imply that the singer would hold the pitch to [i] in *hvile* and [l] in *ville* in order maintain a clear semantic distinction. This further translates to a question of quantity, as [ɪ] is thus shorter in *ville,* for example.[12] Both words, having their syllabic weight on varying phonemes of the first syllable, follow the phonological rule where the vowel preceding double consonants (i.e. the [ɪ] in vi*ll*e) should be short and light. Meanwhile, vowels preceding single consonants should be longer (as in the [i] in *hvile*). This characteristic, which holds true to many Norwegian words, has also been described in spoken Swedish: "*Det har vært vanlig å regne konsonant som står etter en kort vokal i trykkstavelse for lang i norsk (og svensk), mens konsonanter som står etter lang vokal regnes som*

[11] "*I de aller fleste norske dialekter er kravet om at en trykksterk stavelse må være tung, ufravikelig*" (Kristoffersen 2015: 56). Author's translation: "In most of all Norwegian dialects, a stressed syllable is required to be immutably heavy."

[12] There is also a practical and musical consideration in this case, as airflow is blocked by the tongue in a longer /l/ sound in *ville*. Nonetheless, both words are not impossible to be sung and held over a pitch. It must also be said that in quicker tempi these considerations will likely be perturbed.

*korte"* (Kristoffersen 2015: 57).[13] One further trait which these two words share in common is that the second syllable — the [ə] — is light, as it compensates the preceding heavy syllable.[14] In this contrast of light versus heavy— or quantified as short versus long — spoken Norwegian differs from Estonian in the topic of quantity by distinguishing only two levels.

Quantity and prosodic stress in Norwegian have a complex relationship in the phonological level of the word. While one could make a similar claim about such relationship in the case of Estonian, the major distinction between the two languages is that Estonian almost always stress the *first* syllable[15] while stress in Norwegian words vary greatly, dependent on the type of words and as well as the context of dialect. Norwegian in its written form has two varieties: *bokmål* and *nynorsk*. However, these written forms do not always reflect the colloquial speech used everyday. As a spoken language, there are many dialectical variants with considerable lexical and phonological variations. Lexical stress, a unifying element across Norwegian's many spoken varieties, is a stronger determinant in the discussion of quantity compared to Estonian.[16] In a musical context, this means that in the phonological level of word, Estonian bears rhythms which almost always begins on an *ictus* or downbeat, while this is not always the case with Norwegian. Setting music to texts of these separate languages, these dissimilar features have strong musical consequences on rhythm and meter. In my experience of setting texts of *both* languages to the same piece of music, the composer must be even more wary of the fluctuating phonological contexts.

A moment where this became a compositional challenge for me is in the composing of *vihik (c)* for mezzo-soprano and orchestra, where an Estonian poem by Marie Heiberg (1890 - 1942) was juxtaposed with the Norwegian poetry of Hanne Bramness (b. 1959). As the Norwegian text has

---

[13] Author's translation: "It has been typical to assume that consonants are long after a short vowel in Norwegian (and Swedish), while consonants coming after long vowels are rendered short."

[14] It is said that the second syllable will be lighter or 'open' in a word in C-V-C-V, or cosonant-vowel-consonant-vowel, configuration. "*I innlyd, dvs. når det kommer en vokal etter konsonanten, vil denne naturligvis danne opptakt i andre stavelse, slik at trykkstavelsen blir åpen*" (Kristoffersen 2015: 56). Author's translation: "In the nucleus of a syllable — that is to say when a vowel comes after a consonant — this will naturally create an anacrusis for the second syllable, such that the stress syllable becomes open."

[15] Notable exceptions being loanwords such as *bensiin* ('gas') and the word *aitäh* ('thanks').

[16] "*En tung stavelse vil [som nevnt ovenfor] under ellers like forhold for tilhøreren framstå som mer prominent enn en lett stavelse fordi den på grunn av flere segmenter normalt vil realiseres med større durasjon enn en lett stavelse. Denne prominensforskjellen gjør at tunge stavelser lett kan bli definert som mer velegnet som trykkstavelser enn lette"* (Kristoffersen 2015: 55). Author's translation: "A heavy syllable within otherwise similar situations will [as mentioned above] be heard by the listener as more prominent as a light syllable because it is realised with a longer duration than a light syllable. This difference in prominence can be easily defined as stress than its light counterpart."

varying segments of stress from word to word, musical phrases do not require tempo changes to maintain a floating quality. Meanwhile, Heiberg's Estonian text is not syllabically regular and embodies an element of rhythmic uniformity (partly from the 'first syllable stress' rule in Estonian) which beckons for tempo changes to be notated in the score.

The manipulation of the contrasting quantities pertains to the control of the musical parameters of rhythm as well as metric stress. The composer's application of these phonetic concepts can affect the artistic consideration of speech-likeness. While "the material of speech is [hence] never completely predictable" (Laver 1994: 526), it is nonetheless a possible task for the composer to dissect what speech segments *could* be so that temporal musical systems (e.g. meter) and materials (e.g. rhythms) can be established within a score. "The perception of rhythm is not only analytic and predictive […] It can be constructive as well. The human cognitive system seems unable to resist the temptation to impose a constructed rhythm on suitable sensory material in the time domain" (Laver 1994: 524). Contrary to this 'temptation of constructing sensory materials through time', as seen in the Stravinsky-Schoenberg contrast mentioned in the previous chapter, the composer may also intentionally turn away from speech-likeness. However, this choice between basing an artistic judgment based on intuition or based on understanding is not always clear. As a non-native Estonian speaker and as a composer wishing to have better understanding of my musical decisions, the knowledge of what constitutes the qualities of regular speech in a language has been an asset in considering musical possibilities in my compositional process.

---

The next subsection of this chapter (**Chapter 3.1**) titled *Musical Imitation of Articulative Properties in Speech* will discuss mainly consonants and other small phonemic structures in composition. As some consonants cannot be sustained in duration (e.g. [p], [k] or [t]), they cannot be musically altered. Other consonants (e.g. [f] or [s]) may be sustained and semantic interpretation will not be changed. This subsection will discuss my experiences of working with these phonemic elements. The following subsection (**Chapter 3.2**) *Quantity Distinction, Rhythm and Formal Construction* follows my experience in working with building rhythmic structures through the prosodic reading of a text in the theory of Yury Lotman, as well as discussion phonetic problems revolving around such interpretation. *On the Pitch-Quantity Phenomenon* (**Chapter 3.3**) examines the phonetic feature where pitch and quantity meet in spoken language. This is an interesting

phenomenon for a composer as quantity in speech can be associated with pitch elements (e.g. Q2 and Q3 in Estonian).[17] In other words, it is a discussion of correlations between certain rhythm and pitch phenomena. As this is a phonetic phenomenon encompasses both quantity *and* tone, this subsection serves as a transition into the following chapter on compositional questions in tone.

Section Summary:

- *Quantity* denotes the lengths of a speech structure (e.g. a long or a short syllable). The duration of these segments may be measured in absolute numbers but this is purely statistical. However, phoneticians may collect quantity data by recording speech and measuring such durations for comparing habits of a collection of speakers.
- The Estonian language contrasts three quantity levels: short (Q1), long (Q2) and overlong (Q3). The Norwegian language contrasts only two: short and long.
- The element of quantity is important in the discussion of prosodic stress, as a stressed syllable in a word may partly or fully rely on lengthening the duration. Varying this may alter the lexical content and listener's understanding, albeit an artistic choice.
- The Estonian language places stress on the first syllable of single words while stress may vary from word to word in the Norwegian language.
- In my experience of setting music to Estonian and Norwegian texts, the discussion of quantity and prosodic stress plays a vital role in considering tempo, rhythm and meter. Norwegian texts are less attached to the meter and relied less on tempo variations to create variety. Meanwhile, the opposite can be said about my music set to Estonian texts.

[17] Lippus 2011: 20.

## 3.1 Musical Interpretation of Articulative Properties in Speech

This subsection discusses the musical interpretation of articulative properties in speech. The smallest unit of phonetic studies is known as a *phoneme.* This concept, described by 20th century linguist Roman Jakobson from a "purely sensualist[,] point of view, [is] the direct observation of the motor and acoustic facts and the instrumental analysis of them would show us that the observable characteristics of this phenomenon are essentially the same in various languages" (Jakobson 1990: 219). In more contemporary scientific terms — and a more fully encompassing one — that are discussed in phonetics studies today, it is "[t]he minimal unit in the sound system of a language, according to traditional phonological theories. The original motivation for the concept stemmed from the concern to establish patterns of organisation within the indefinitely large range of sounds heard in languages" (Crystal 1991: 258).

This scrutinisation of phonemes falls under the category of quantity because the "'manner of articulation' and 'place of articulation'" (Laver 1994: 130) of a segment "will often have some intrinsic duration which it must exhibit before it can be perceived as a segment or feature of that type" (1994: 432). For example, the unvoiced dental-alveolar consonant (i.e. [t] in the word *tie*) bears a sequence of physiological actions (i.e. tongue placed on the dental-alveolar region of the mouth and then released by the outward motion of the airstream) which require a certain duration to produce. It is physiologically difficult to alter this duration. Different segments have different actions and properties which must be fulfilled by the utterer in order for the listener to distinguish meanings.[18] The durational differences can thus affect the perception of quantity, and consequently the semantics of a text.

As texts are comprised of various segment combinations which can be read aloud, this is an inevitable part of song and the pedagogics of diction. Composers have frequently assumed that the performers bear at least some background knowledge of the language of the texts, in addition to the common practices of the genre. If the performers do not possess such knowledge, then it is presumed that the performer is responsible for learning how to produce these speech sounds according to its phonology through other sources than the score. This assumption, however, is not always conducive to the composer's creativity, as the the interpretation of speech sounds is limited

[18] "The phonetic specifications of the sounds (or *phones*) heard in speech, it was realised, contain far more detail than is needed to identify the way languages make contrasts in meaning" (Crystal 1991: 258).

to the performer's general knowledge of the work's language. The performance is at the mercy of the singer's prior knowledge or assumption of how the text should be performed. In specific situations, the performer may not necessarily have the resources or time to verify the accuracy of the speech sound reproduction. From my own practice, I would inasmuch claim that this onus should be one shared between the composer and the performer. Especially in the quantities of sung syllables, the composer can assist the performer learning the work within the notation. For example, consider the short passage in **Figure 4.1** below:

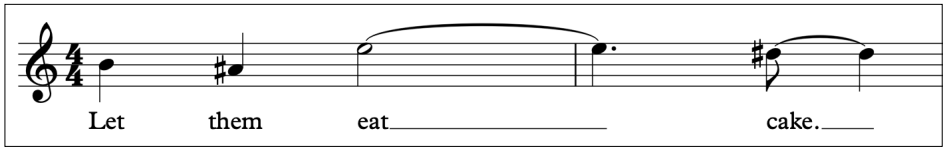



**Figure 4.1** A passage of music with no specific notation for consonant endings.

In the lyrics, each of the four words 'let, them, eat and cake' all contain a consonant ending. As the notes do not specify the onset of these ending consonants, different performers would interpret them differently. Assuming the tempo of the passage is moderate enough for the perceiver to comprehend all the words, the questions of how long the preceding vowels should be held and whether if the [m] in *them* should have an earlier onset or how much weight the [t] or [k] consonant endings should bear are all valid considerations for the composer. In the notation's current state, such decisions are left at the discretion of the performer. However, a further parsing of such micro-elements will affect the sounding result and it is the composer's artistic choice to manipulate these micro-durations. Consider the variation of the same passage in **Figure 4.2** below:

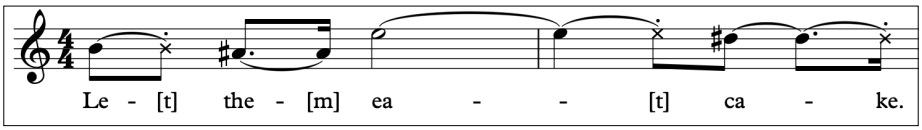



**Figure 4.2** A passage of music with specific notation for consonant endings.

This method gives the performer a clear idea of the rhythmic placement of the consonant endings. In the words of music editor Elaine Gould and author of notation manual *Behind Bars* (2011), through a "mutual understanding of the rules and conventions of notation, the composer can 'speak' effectively to the performer, who then has the best chance of achieving a faithful interpretation of the composer's intention" (Gould 2011: xvi). Music notation, although ineffective

for notating speech sounds in other contexts, is a system where quantity can precisely be notated. Therefore, a composer can view the dissecting of speech segments as a generative method for musical details. A work which makes use of this idea was Luciano Berio's *Sequenza III* (1965) for solo female voice:

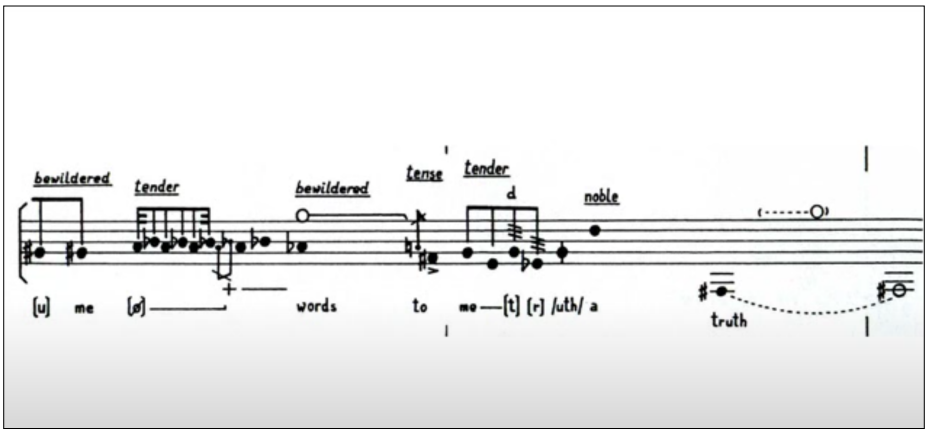



**Figure 5.0** An excerpt from Luciano Berio's *Sequenza III* (1965).

In this short passage demonstrated in **Figure 5.0**, the word *truth* is represented twice. The first occurrence is dissected into phonemic and syllabic segments and the second appeared in the word's entirety. Berio's micro-variation on *truth* drew my interest as a composer, as the segments of words (i.e. phonemes or syllables) are ready-made rhythmic materials through the text. The [r] and the coda of [uθ] both possess vastly different rhythmic qualities: while the former can be rolled, voiced and sustained, the latter — considering the voiceless of [θ] — cannot. I will first focus on the rhythmic aspect of interpreting texts this way; the subsection in **Chapter 3.3** will further discuss how reading texts in this way can inform larger musical structures such as phrases and forms.

Returning to how these ideas have affected my own compositions in the languages I work with, this concept of articulatory speech properties has played a vital role in notating rhythms. My notation of vocal music has changed considerably as I started working with the concept of segment durations. Even as early as the first creative project in this research — *natt-öö-夜* (2018) — I had considered the value of dissecting rhythms of words. This passage in **Figure 6.0** shows one of the many basic examples of how word combinations from Juhan Liiv's poetry generated contrastive rhythmic ideas:

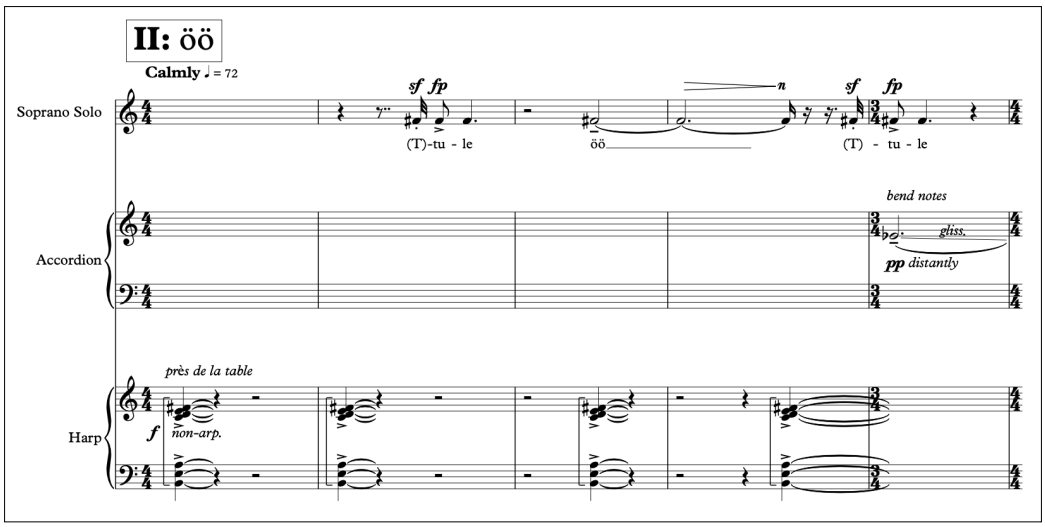



**Figure 6.0** An excerpt from the second section of *natt-öö-夜* (2018).

The Estonian word *tule* (‘come’ in its imperative from) possesses one short and one half-long syllable with the stress on [tu]. Meanwhile, the word *öö* (‘night’) is simply a long [ø] vowel. This pairing in itself is already a musical phrase in itself: two short quantity elements are followed by a half-long quantity element. I have taken this rhythm in its spoken form and embellished it further with dynamics and accents to achieve my desired phrase. The [e] in *tule* in speech is phonetically half-long. Though, I also wanted a longer note to fade out with the harp resonance. The speech-likeness is emphasised by the *fortepiano* dynamic, where sudden dynamic drop would give the [e] an illusion of a fade-out while maintaining pitch.

The previous example is a relatively straightforward consideration of quantity in the words’ speech form. It is a direct representation of what is already assumed in regular speech. There are other ways to expand on this rudimentary concept. As discussed in the Berio example, the words are broken down into the phonemic level and the higher level segments may be further considered in a composition. Not only as musical points of interest, Berio has written in his program notes for his *Sequenza III* that “emphasis is given to the sound symbolism of vocal and sometimes visual gestures, with their accompanying “shadows of meaning”, and the associations and conflicts suggested by them” (Berio 1965). In my interpretation, these gestures and “shadows of meaning” are tied to the articulation of phonemic segments. The physiological processes to produce speech sounds such as an open, dark [o] vowels or the airy hiss of [θ] are suggestive of concrete imageries and different moods evoked by onomatopoeic associations.

It can be debated whether Berio chose to repeat the word *truth* in full for sheer emphasis or for better semantic comprehension in the passage shown in **Figure 5.0**. Regardless of the reasoning, one major implication of a composer working closely in the phonemic level is that the word's semantic contents may be lost to the perceiver. The word *truth* dissected into [t], [r] and [uθ] and then further embellished by temporal processes is not easy to understand as the word *truth*. Focusing on Berio's disfigured *truth* to exemplify such a compositional play offers only a fraction of the whole truth about phonemes, however. To further this discussion, the reader must first be aware of the concept of the three phonetic segments related to quantity:

"[S]egments are articulatorily classified partly in terms of the maximum degree of construction of the vocal tract reached during the production of the segment. The period during which the maximum construction is achieved defines the medial phase of the performance of the segment. Preceding this medial phase, an onset phase embodies the approach of the vocal organs to the medial phase, and an offset phase shows the movement of the organs towards the medial phase of the next segment […] (Laver 1994: 112).

The musical parallel to this phonetic concept is "attack, sustain and release." Each segment requires an amount of time to produce. Just like in a music phrase, if the sustain is perturbed by an abrupt pause or a stop, it is interpreted as a release. There is simply no possibility to delay or elongate a release, where the sound stops. The similar can be said about an attack detached from the sustain. This is therefore why Berio's *truth*, when fragmented in different smaller combinations, is not easy to piece back together as the complete *truth* when listened to. However, let us now assume the word is extended for a much longer duration at the phoneme of [θ]. This would then not detract the listener from comprehending the word, even though a regular speaker would never say the word *truth* with such a long [θ] sound. The human speech apparatus is unlike musical instruments: our oral cavities can change formants (i.e. shape) and our tongue can transition smoothly from one place to another. Compared to many instruments, the production of sounds is more than just a 'start and stop' phenomenon, as in the case of, say, a piano note. In other words, to maintain meaning *and* to tinker with phonemic elements of a word, a composer may extend or shorten the durations of onset and offset segments. This method of composing for voice and texts is often integrated in my work. One example would be *kj/ærlige ord* (2018) for three female voices and large ensemble:

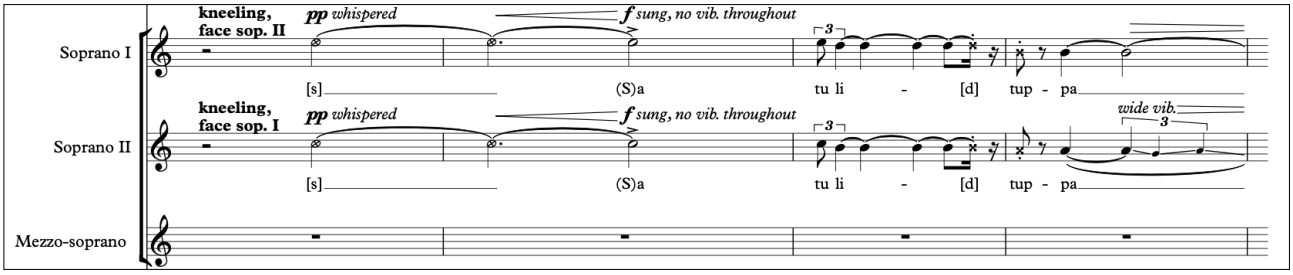



**Figure 7.0** Excerpt from the opening of *kj/ærlige ord* (2018).

The [s] segment of the word *sa* (Estonian for "you") in this passage may be infinitely extended in duration without being distracted from its meaning, as the medial phase (i.e. when [*a*] occurs) determines the word *sa*. Then later on, the word *tuppa* (Estonian for "into the room") requires a [p] consonant at the Q3 level. While it is meant to be a single word, an eighth rest bisects the word, rendering the [p] supposedly silent for a moment. While counterintuitive, this is, however, a reflection of the pronunciation of a disyllabic CVCCCV[19] Estonian word structure. Studies by phonetician Pärtel Lippus has shown that "[i]n Q3, the stressed syllable has to include a perceivable high part and a perceivable falling part" (Lippus 2011: 37). As the cross-noteheads of the score denote an almost voiceless utterance (i.e. whispering), the 'peak' intensity (or high part) of this word would be the intensity of [u] leading into [p] in *tuppa.* The falling impulse, as it is a whisper, must be silent, as there was no pitch or intensity to begin with. Like in a general practice of writing staccato notes, a silent space in time highlights the preceding element's significance. The second syllable, although weaker, is sung with pitches. Converse to the [s*a*] example which plays with the onset segment, the [*a*] vowel deals with the offset. This [*a*] is not immediately followed by another segment and therefore extending this sound would not distract the audience from understanding the word within the context of a sentence or a line.

Examples shown thus far on articulatory properties have come from my music utilising the voice. However, it must be said that I was frequently suggested by colleagues in the course of research to explore the instrumental applications of these concepts. Between making speech sounds an abstract element in form, I realised it was possible to imitate speech articulation. As mentioned earlier in this chapter, the human voice apparatus has the ability to morph itself into a different shape and affect the airstream, while most instruments do not share this capability. My string quartet

[19] C stands for a consonant unit and V stands for vowel.

*vihik (a)* is one of the artistic experiments tackling this research question. Although, this experiment has not not exactly conform to the ideas of prosody.

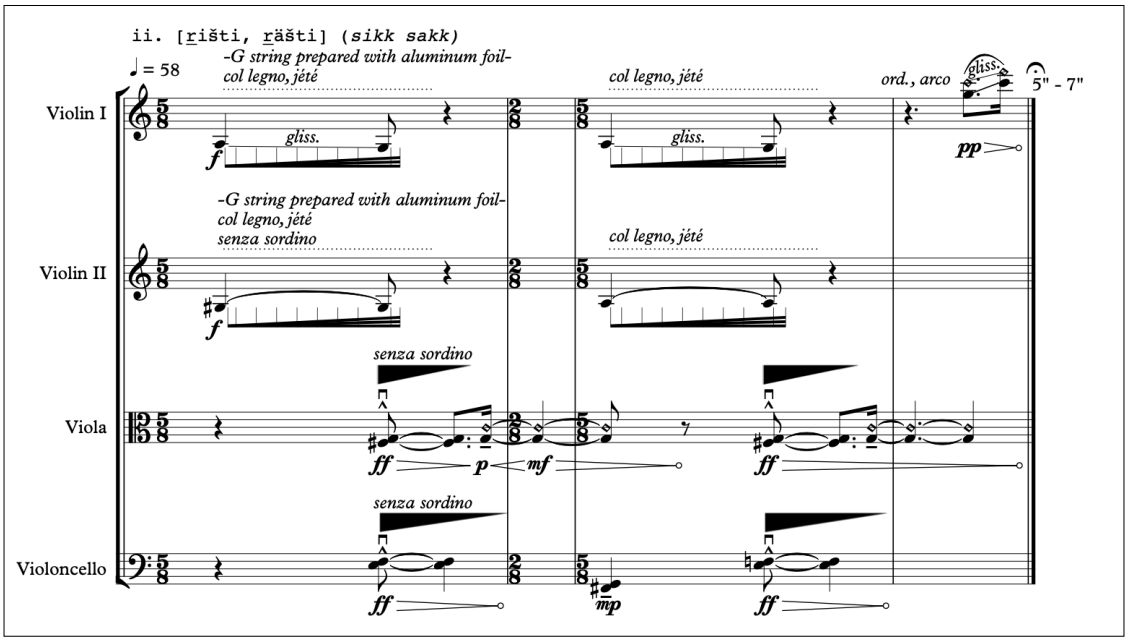



**Figure 8.0** The second movement of *vihik (a)* (2018).

The string quartet consists of ten miniature movements. Each movement is titled with a short phrase or sentence which I selected from my readings of different poetry. These fragments have been selected for their prosodic structures in addition to alliterative qualities. In **Figure 8.0**, the short Livonian phrase "*rišti, räšti*" ("crisscross" in English) is chosen from a book of new Livonian poetry.[20] The rolled trill [r] onset consonant and the [ʃ] voiceless postalveolar fricative combination is an attractive structure to work with in this movement, as the rolled [r] articulation into the [i] of this word possesses a heightening-falling intensity, which resembled the *jété* technique of the violins. The duration of a *jété* gesture is limited, which makes for the sudden outburst of the lower strings' crushed tones a metaphor for the noisy [ʃ] sound. The final closed vowel of [i] in both instances are replaced by the pinched harmonics of the viola.

I am not completely satisfied with the relationship this piece established with the quantity discussion of this subsection. This is not to say the piece has been 'unsuccessful.' It is simply that the focus of working with prosody was blurred with other compositional considerations such as timbre and pitch development. While the concept of articulation in music is fulfilled (e.g. [r] and

[20] Damberga, B., Ernštreits, V. and Kemppi, K. 2018. *Līvõ lūolkub: Tritium* 2018. Riga: The International Society of Livonian Friends.

*jété*) and this does indeed pertain to quantity, it is a sound metaphor than an imitation of the prosodic structures of the short phrases; the sounding result in the performance can be evaluated as too abstract to identify the text sources. As a result, this musical experiment of making use of phonetic segments in an instrumental context does not correspond well to rhythms pertaining to speech-likeness. The interpretation of what I thought was speech prosody like became timbral metaphors through the instruments and has created different musical results. This does not mean instrumental music cannot imitate the prosodic structures of speech but I believe that it may simply be approached through a different method. One interesting outcome which has come about from this, though, is that I began to see larger forms emerging in working with such metaphors. This will subsequently be discussed in the following subchapter *Quantity Distinction, Rhythm and Formal Construction* (**Chapter 3.2**).

Section Summary:

- Articulatory properties of speech are tied to the discussion of quantity. Articulation of a word is divided into three phases: onset, medial and offset. This is akin to the concept of attack, sustain and release in music.
- In this parallel relationship that has been mentioned in the previous point, the composer is then able to notate and specify in the score for when the three phases occur in a syllable or word segment in songs. The meaning of words can be maintained through knowing which segments' duration can or cannot be adjusted.
- In the case of Estonian, where Q2 (long) and Q3 (overlong) are distinguished, dynamics, pause in sound and pitch can be used to emphasise the distinction in held notes.
- Instrumental representation of articulatory properties in speech are theoretically possible, but may be heard as timbral metaphors rather than direct prosodic or quantity imitation.

## 3.2 Quantity Distinction, Rhythm and Formal Construction

"But suppose that sound were a simple thing: would it constitute speech? No, it is only the instrument of thought; by itself, it has no existence. At this point a new and redoubtable relationship arises: a sound, a complex acoustical-vocal unit, combines in turn with an idea to form a complex physiological-psychological unit. But that is still not the complete picture."

— from the *Course in General Linguistics,* Ferdinand de Saussure (1857 - 1913).

This subchapter expands on the concept of quantity in the context of musical form and its construction. As presented in the introduction of this chapter (**Chapter 3**), the research on quantity could be mainly quantitative data comparisons, depending on the researcher's intentions and goals. Such research can be eye-opening for the composer, since the results can potentially inform our understanding of what speech patterns in the time dimension are regular in speech. Though, highly in-depth scientific research in phonetics, on the contrary, can provide its reader with such details on certain subjects that their theories may be too specific for purposes of artistic application (e.g. specific contexts of certain words). Reflecting on the Saussure quote which precedes this paragraph, the subject of *phone* in speech therefore does not provide a composer with inherent musical significance; the physiological aspect of quantity in speech (i.e. the production of sounds of a certain duration) is one mere aspect of working with the materials. In the very same way, the concepts I have discussed in **Chapter 3.1** (i.e. articulative phonemes) do not constitute a musical structure; they are one aspect of a whole (i.e. a piece of music or a musical phrase). It is certainly valid to work with tongue taps, labial smacks and other speech sounds acousmatically to create music but it is not the sole method. Combining that with the different associations these speech sounds create, the process offers greater creative possibilities. For example, the play of interpreting texts on different segmental levels, reimagining these interpretation and redesigning the vocalisation of such utterances constitute a different creative process. I believe that many composers, including myself, find this union of 'physiological-psychological' elements in speech sounds to be generative and fruitful in composing.

Defining syllabic segments in texts is a contentious topic between literary scholars and phoneticians; this contention will be discussed later in this chapter. We must also acknowledge that the phonetic discussion of quantity may pertain to the absolute or relative durations contrasting segments (e.g. a syllable) of a larger spoken structure (e.g. a word, or a sentence). Not only that,

quantity is also a question in metrical phonology. To further this discussion, the reader must be able to distinguish phonetics from phonology: "phonology is concerned with the range and function of sounds in specific languages" (Crystal 1991: 261) while phonetics is the all-encompassing term describing "[t]he science which studies the characteristics of human sound-making, especially those sounds used in speech" (Crystal 1991: 259). Only with this distinction made, can the reader under the separate issues of language-specific problems and pure speech sound problems. In the case of a language-specific issue, it has been mentioned in **Chapter 1** of this dissertation that foot isochrony is a potential feature of the Estonian language. Two types of isochronous recurrence have been proposed for categorisation: syllable-timed languages and stress-timed languages (Laver 1994: 523). For example, just as in the Estonian language where syllable can be measured in long and short, Japanese deals with the concept of *mora,* "a term used in traditional studies of metrics to refer to a minimal unit of metrical time equivalent to a short syllable" (Crystal 1991: 223), which means that a long syllable could be inferred as the length of two short syllables, whatever duration a 'short' syllable might denote or imply. Although, contrasting Japanese and Estonian, discussions of syllable-time isochrony are present in both languages but the methods used to describe their relationships to duration are vastly different. In stress-timed languages, like English, length of speech segments depend on where the stress of a larger segment (e.g. sentence) is placed.[21] In other words, these diverging theories give their reader different methods to understand time and rhythm in speech, depending on the language. To encompass these different types of discussion, linguists have used to analyse durations in speech is metrical phonology "in which phonological strings are represented in a hierarchical manner, using such notions as segment, syllable, foot and word" (Crystal 1991: 218), or are often times referred to simply as 'prosody'.

Prosodic structures of speech can be understood in absolute durations (e.g. measurements in milliseconds) or relative durations (e.g. abstract descriptions such as 'long' or 'short'). However, the perceived sounds of different qualities (e.g. loudness, stress, speed of speech) and contexts can affect the perceiver's cognition of durations. Such studies can be traced to linguistic traditions

[21] In Peter Roach's article *On the distinction between 'stress-timed' and 'syllable-timed' languages* (1982) a conclusion was that "stress-timed/syllable-timed distinction seems at the present to depend mainly on the intuitions of speakers of various Germanic languages all of which are said to be stress-timed; examination of the subjective feelings of speakers of languages usually classed as syllable-timed should be carefully studied if the distinction is to be maintained as a respectable part of phonetic theory" (Roach 1982: 78). In other words, linguists such as Roach believe that the feeling of syllable-timed elements may be derived from intuition and bias rather than any sort of absolute distinction. This is an important debate for the reader to keep in mind that this is indeed a contentious debate within phonetics studies, and that while this dissertation uses phonetic concepts and terminologies, there are concepts which remain to be controversial when confronted with phoneticians.

during and before the twentieth-century with literature by Roman Jakobson, Ferdinand de Saussure, among many others. Prosodic studies will point to a multitude of interpretations in speech from semantics to semiotics to pragmatics. For now, I would like to focus on instances of different affective qualities in how prosodic stress is important for a reader's interpretation. For example, the English sentence "I have already done it!" can be read aloud in various ways for varying nuances of meanings (e.g. compare "I have already done it!" and "I have already done it!"). This phenomenon is salient and ubiquitous in most spoken languages. The manipulation of a text's prosodic stress — through volume, time or pitch — affects the semantic and affective comprehension of a perceiver. These traits are repeated in our speech and are not arbitrarily constructed by the speaker; the performance of these vocal-prosodic constructions is a highly codified system in every spoken language, as each tongue has its own nuances of surprise, disappointment and other affective responses. These performative moments are reproduced for their intended effects in human speech, and this can also be found in writing. Theoretical concepts on poetry in Russian literary scholar Yury Lotman's *Analysis of the Poetic Text* begin to emerge from this thinking. For example, he has argued that "'[r]hythm is generally understood as meaning correct alternation, the repetition of identical elements. (It is precisely this property of rhythmic processes, their cyclic nature, that defines the meaning of rhythm in natural processes and in man's work)" (Lotman 1976: 42). In less phenomenological terms, some languages — such as Estonian and Finnish — will emphasise sounding repetitions of 'longness versus shortness' as anchors for comprehension, while other languages emphasise other forms of stress or intensity, such as raising pitch at the end of a question in the English language. Concluding on this thought, segmental durations in varying contexts, their repetitions and the regularity of their usage affect the perceiver's emotional and semantic understanding.

Steering this conversation back towards music composition, the composer's own perception, relationship to and understanding of a spoken text play an important role: a written text can be sung in a myriad of variations for different affective results. If the composer wishes to embody a specific expression, rhythm — which as we have discussed as something closely tied to quantity — plays an important role. This question of 'rhythm' as Lotman has earlier pointed out is a valuable toolbox for composers working in all languages, regardless of their fluency in the language. In the description of Lotman, rhythm can be understood as both semantic (i.e. repetition of concepts or meanings) and pragmatic (i.e. repetition of sounds). The point is that repetition in texts — and I would opine the same in music — is a tool to seek out meanings and to create new ones for the listener.

Lotman further describes his understanding of rhythm in poetry as "a sense-discriminating element" (Lotman 1976: 42) and such "poetic structure does not merely manifest new nuances of word meanings but reveals the dialectics of concepts, that internal contradictoriness of the phenomena of life and language for whose designation ordinary language lacks special means" (Lotman 1976: 43). The 'element' is echoed in Jakobson's writings as being "purely sensualist" (Jakobson 1990: 219), or in Saussure's term: physiological (Saussure 1986: 143). In my interpretation, Lotman is insinuating that the poetic use of words in language should bear a greater freedom than its phonic value and dictionary meanings. Comprehension of poetic texts should therefore come from the human senses (i.e. touch, as in the physicality of our oral movements) as well as the abstract semantic interpretation. The composer — having the need to offer their own input as melodies, songs or abstract forms — is manipulating a layer of the perceiver's and performer's understanding. This process sometimes require the composer to identify, deconstruct and manipulate phonetic segments. In a framework between clear utterances bearing functional meanings and the play with Lotman's poetic sensibilities, the composer is endowed with a creative freedom.

Although Lotman has considered mainly examples from his native tongue, Russian, the reader must be reminded that the feeling of rhythm is found across all languages. Furthermore, the isochrony debate (i.e. 'stress-timed' versus 'syllable-timed') influences the perception of rhythm and other prosodic structures in different languages: a reader will always have prejudices according to how they have learned their languages. One example would be the contentious debate on syllabification (i.e. how does one define a syllable in the first?). Some phoneticians have argued that "native-language intuitions may influence decisions" (Crystal 1982: 76). In that sense, Lotman's poetic analysis of text may be seen as speculative in the field of phonetics, even if not called out as too Russian-centric or Indo-European languages-centric. Though, for composers whose pursuits are not to be scientists, Lotman's theories are nonetheless valid methods for constructing sound materials which ultimately serve the purpose of personal aesthetics, even if it is through accidental misinterpretation of science. I believe that his understanding of syllable segment and rhythm is generative for creating music, despite its reliance on intuition.

Despite the contentious definition of what qualifies as phonetic segments, quantity distinctions in texts and their derived rhythms have been an ongoing motivator for my own works. I would like to discuss the work *isjoriske forsteder* ('Izhorian Suburbs') because of its utilisation of

Lotman's concept of 'rhythm' on a macro-level (i.e. formal structure spanning over a large movement). In *isjoriske forsteder*, a work in two versions (one for soprano and electronics, and one for soprano and large ensemble), I have been largely working with languages I do not speak. This prompts the composer to be wary of what misinterpretations could arise from setting the text unnaturally but at the same time also offer an insight into how she/he may interpret meanings from sheerly the phonic elements. However, taking into account the repetition of texts a form of understanding 'rhythm' in Lotman's context, I have selected moments of text repetition as points of focus in the form of *isjoriske forsteder*. In the excerpt highlighted in **Figure 9.0** below, I have taken a line from Juhan Liiv's poem which writes *mu kullake* ('my dear') and placed emphasis on the [lɑ] syllable segment:

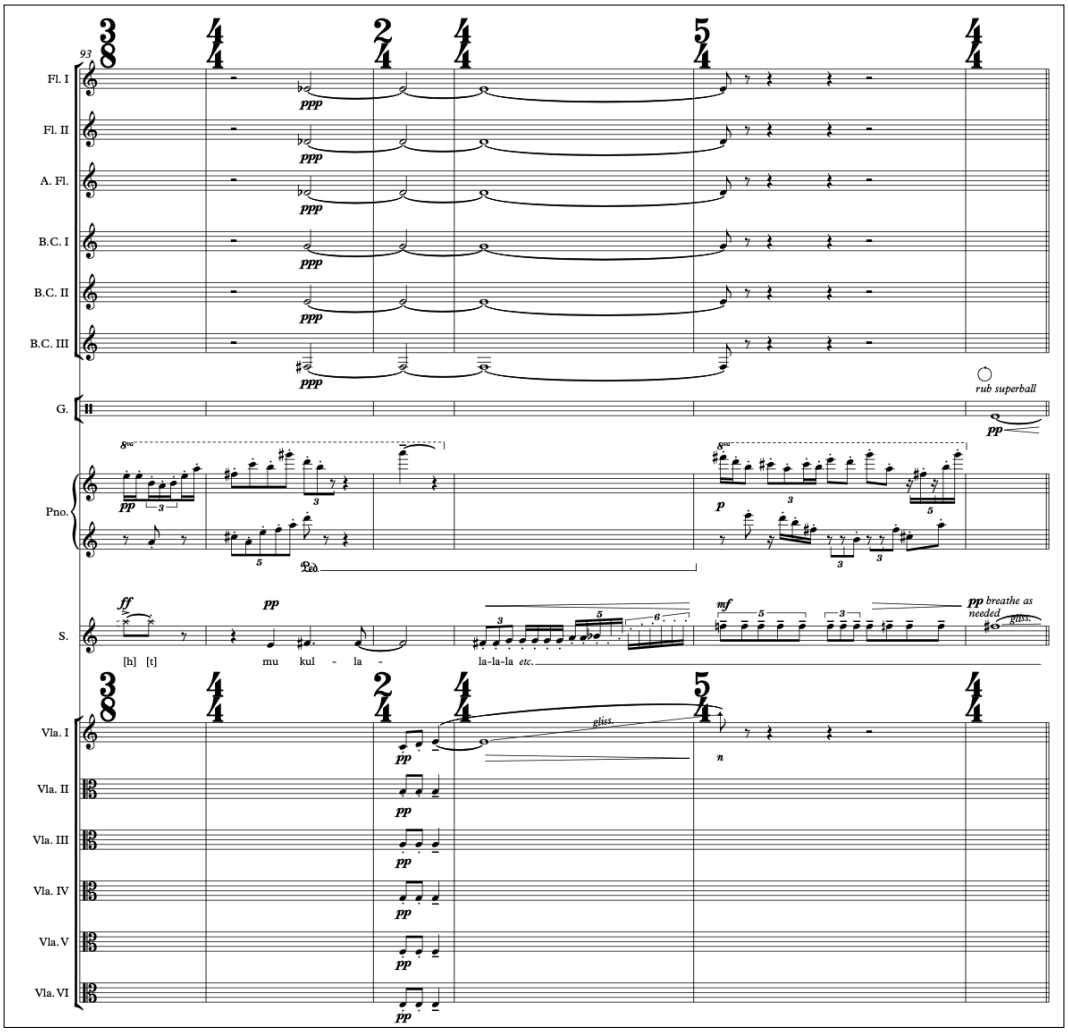



**Figure 9.0** A section from the second movement of *isjoriske forsteder* (2019).

The reason for emphasising this textual moment is because the fragment occurs late in the second movement of the piece and it serves as a return to the opening of the first movement which has highlighted the Estonian word *laht* ('bay'). Just as in the first movement, the emphasis here is placed on the *la* sound disconnected from the *ht* segment. As Liiv's text in the second movement paints the imagery of a small boat sailing over the gentle waves (*'Üks laevuke lä'eb üle vee, lä'eb üle vee, ja lainete')*, the [lɑ] syllable bears thus an associative function with water in my piece. Furthermore, as [lɑ] is often understood as a song-like syllable (i.e. 'la-la-la' and in the first consonant-vowel segment of Estonian word for 'song', *laul*), the semantic associations bind the poetic imageries together with the act of singing. These associations and lexical functions form the *semantic plane*[22] of my work. The [lɑ] syllable has proven to be a useful segment in extending passages while maintaining the same semantic plane and unity of disparate passages.

While a listener may argue that this sort of musical play distorts the listener from perceiving the text's intended meaning, I believe that in the context of my piece, it is a position in extending and exploring similar sounds for creating new semantic associations and freshness over a half hour of music. The poems are short and significations of single words are less important than the semantic plane encompassing the mood of the piece. The purpose of such associative construction over several movements of a work is not to demand the listener to make such intricate connections but rather to assist myself as a composer to orient the direction of my music and to develop form with some sort of logic in a large work. Or in Lotman's term: utilising 'rhythm'.

In this compositional thinking, the selection of fitting texts impacts the formal construction greatly. Having first decided that the word *laht* was beginning the piece,[23] my choice of Juhan Liiv's poem where many the cognate speech sounds came as a serendipitous find during my preliminary research phase for composing *isjoriske forsteder.*

—

[22] In the reference to *planes* within linguistics, the 20th century Czechoslovakian linguist Josef Vachek describes in his *Dictionary of the Prague School of Linguistics* as "the relation between the sign, and the signifié and the significant phenomenon" (2003: 131). In my example, the sign in my work is [lɑ], while the significance is implied to be both songfulness (i.e. the singing of the [lɑ] syllable) and, within the work's context, a sound relating to concepts on water (e.g. *la*ht, *la*inete or *la*evuke).

[23] To understand why *laht* was selected as the first word of the piece, refer to the section on *isjoriske forsteder* in the **Appendix**.

As I have stated in the beginning of the subchapter, we will zoom out step by step from the phonemic segment to higher level segments such as the foot and the word. As *isjoriske forsteder* can be understood as an exploration of the syllable segment, the upcoming topic is word segment level, which has influenced my string quartet *vihik (a)* (2018). The *vihik* series of works is inspired by text fragments in my sketchbooks which I have collected from casual readings of texts and in the course of my research for *isjoriske forsteder*. Being the first work of the series, I have parsed and dissected some sentences or verses from poetry I have read or written myself. The parsing of these lines would then be transformed into musical phrases or gestures and the result of the work was ten short movements, with none lasting more than sixty seconds. The short durations of each section has come from the loose interpretation of quantity distinction between each phonemic element. Consider the following passage:

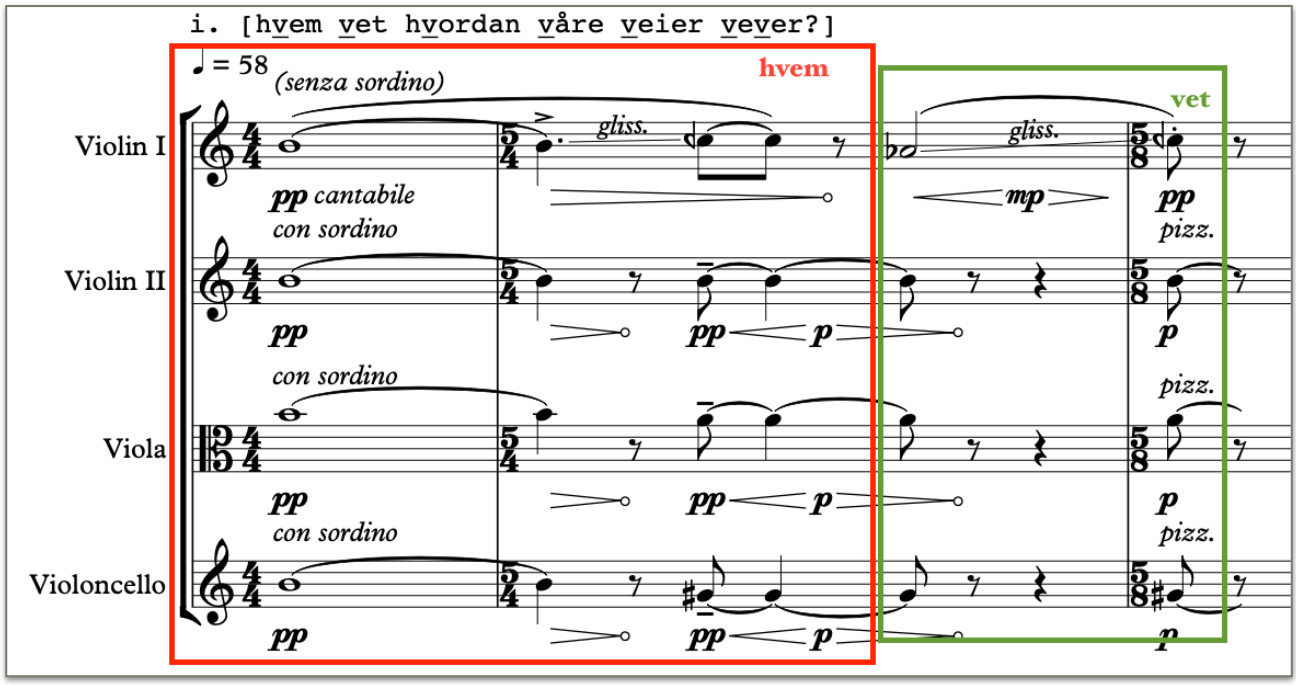



**Figure 10.0** The opening movement of *vihik (a)* (2018) highlighted with the corresponding words to the texts the passage draws from.

Note the word corresponding to the colour of the boxes in the passage. Each boxed segment indicates the word which the music attempts to represent or 'recreate' prosodically. In the Norwegian word *hvem* ('who'), the quietly emerging long notes in the first bar denote the *hv* phonemic segment, which eventually 'opens up' to the [e] vowel through dynamics in the second bar with an accent in the first violin. The lower three strings closes the word with [m] through its dynamic representation (i.e. decrescendo to *niente*) before the new word *vet* ('knows') emerges as

[ʊe] in the first violin and is articulated by a pizzicati [t] in the fourth bar.

Upon first sight, one can argue that this type of musical representation of phonetic concepts is metaphorical, especially when only four bars of the first movement is shown in the example. I would have to defend the sound world I have assembled for this piece, if that is the case. For example, the [t] pizzicato gesture is repeated throughout the ten movements, wherever the voiceless dental stop [t] articulation occurs, as I feel that it is not only the metaphor of a tongue stop, but also for its shorter durational phonic value. Various elements in this thinking build up musical representations of words, which in turn are developed into musical gestures and motives throughout the piece. It can be noted that this is one of the few pieces in my artistic research where ideas from phonetic features have been applied my instrumental writing. As well, my working method with the word segment in *vihik (a)* is far from an original thought. One major influence on my work in reflecting on the word segment would be American composer Robert Ashley (1930-2014) referring especially to his album *Private Parts* (1978) where instrumental accompaniments reflected the rhythms of the spoken texts heard throughout the collection. The blend between instrumental mickey-mousing of the spoken text and the colouring of the spoken part with instrumental colours has impacted my non-vocal works inspired by texts.

—

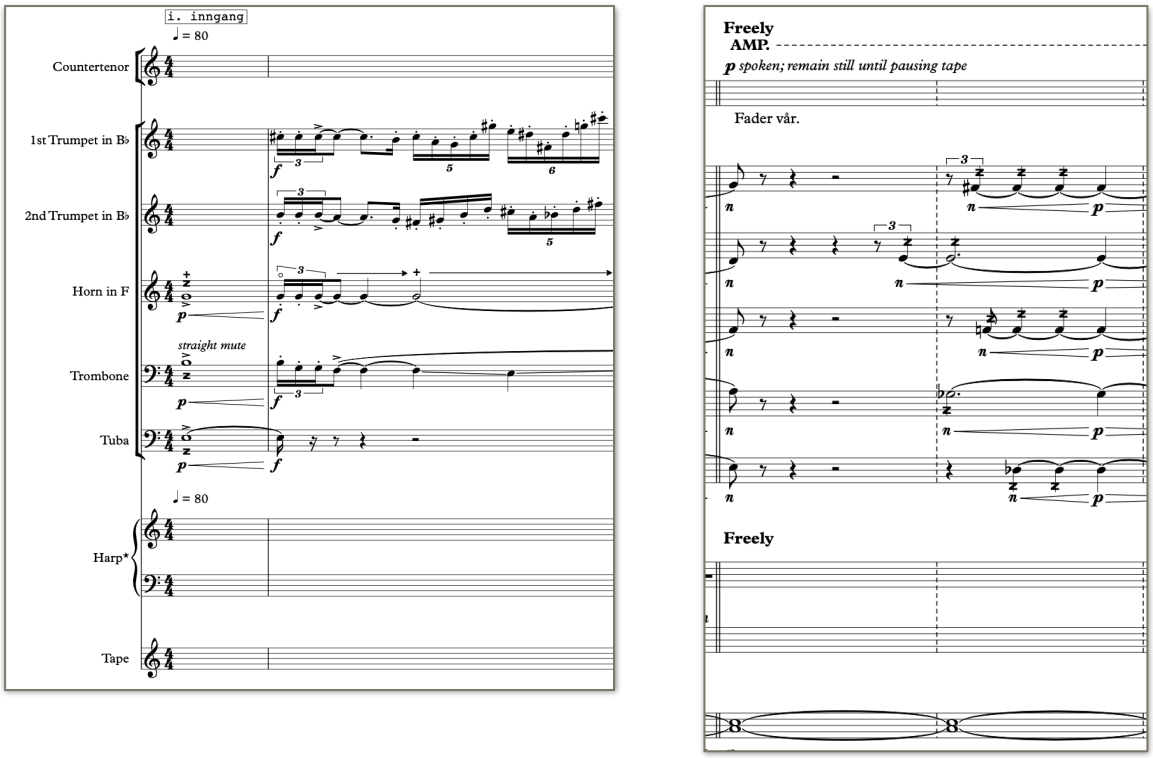
At the level of the prosodic foot, I have examined an excerpt from my work *messe norvégienne profane* (2018) as an apparatus for experimentation. It is important to initially define the prosodic foot in a speech-setting: it is about oppositions of strong and weak through phonetic features (i.e. quantity, volume, pitch and/or stress) when comparing syllables within a word or a phrase.[24] The concept of prosodic foot differs from the parsing of a word or syllable because phonetic segments are considered in a poetic context. Prosodic foot is very much a concept rooted in poetry and songs in contrast to natural speech. Hence, I believe that this a unit many composers have considered when working with texts.[25] On the other hand, the segment of the word may be less

[24] "In a sequence of syllables a relative prominence opposes one syllabic phoneme to the others of the same sequence as stressed versus unstressed. […] The greater and lesser prominence of syllabics a relative notion which can be determined only by a comparison of all syllabics pertaining to the same sequence." (Jakobson 1990: 256-7).

[25] Especially when considering Estonian texts, one major resource I draw from as a composer is the book *The Temporal Structure of Estonian Runic Songs* (2001) by Jaan Ross and Ilse Lehiste, as it details comprehensively the various phonetic segments and their functions in runic singing.

intuitive for working with poetry or prose, as recitation or performance of a text is not a robotic word-by-word utterance.

Written for solo countertenor voice, brass quintet, electronics, an optional harp and audience choir, the *messe* combines fragments of texts from the Catholic mass and a reflective prose by the Norwegian author Linda Gabrielsen. In line with *vihik (a)*, the instrumental writing for this piece takes on a major role in my play with phonetic segments. Immediately at the start of the piece, the ensemble foreshadows the utterance '*Fader vår'* ('Our father', with the trisyllabic foot of 'strong-strong-weak', otherwise known as *antibacchius*) which are the first spoken words of the countertenor:




**Figures 11.0 and 11.1** Opening bars of *messe norvégienne profane* (2018), highlighting the *antibacchius* metrical foot and the first utterance of texts by the countertenor voice a few bars later.

While it is not notated rhythmically in the score, the utterance of *'Fader vår'* is a cognate to the triplet motive which the trumpets, the horn and the trombone play in the beginning of measure 2. The texture quickly dissolves into other musical materials, calms down again and the countertenor's entry would then remind the listener of the same *antibacchius* foot, giving the piece a second beginning. The play between an instrumental interpretation of speech rhythms echoed by

actual speech allowed me to set focus on the larger mass setting, since it is the priest with the words who is in command of the ceremony and the ensemble takes on a preparatory function within the musical structure. Such a technique has not only been proven to me as useful for dramaturgical situations in music composition but it is also a way to generation variation on the same musical materials.

The metaphorical *'Fader vår'* motive in the brass quintet recurs several times over the course of the piece as a shadowy call for god without words. In the movement *Libera me*, where the countertenor leads the audience choir to sing *"libera me, Domine",* the word *domine* ('god') bears the very same prosodic foot, which unites the symbolism of godliness and the Father in this work:

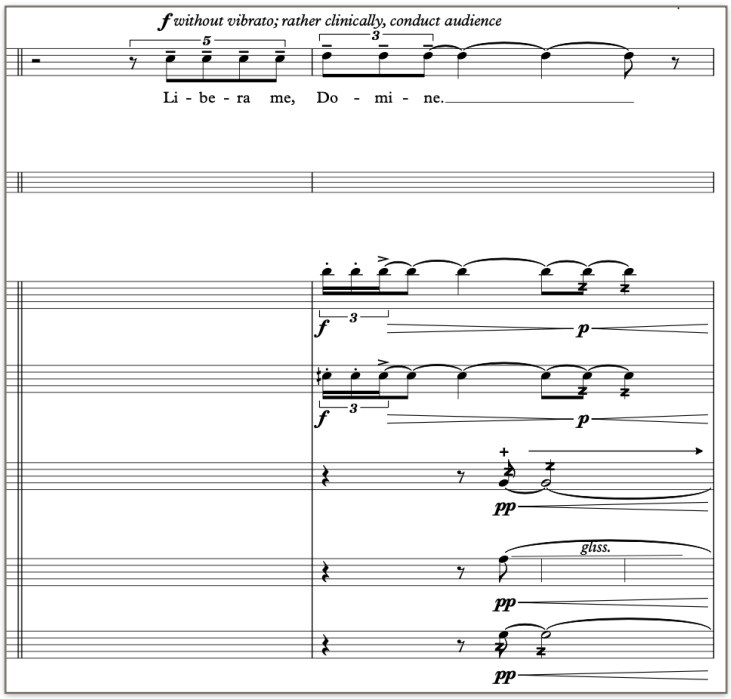



**Figure 12.0** Recurrence of the antibacchius prosodic foot in the *Domine* motive in *messe norvégienne profane* (2018).

—

I have earlier described that phoneticians and phonologists have different methods to understand prosodic relationships (i.e. stress-time versus syllabic-time) in their own language. In this thought, my trilingual work *vihik (d)* for women's choir explores the possibility of tying 'unrelated' threads together. The texts *"Mis seäl meie õue alla?"* ("What lies below our yard?") from *Vana kannel* (vol. 2), *"Du flytter"* ("You are moving") and *"Ingen har ropt"* ("No one has yelled") by Gunvor Hofmo and "牆有茨" ("The walls have thorn") from *Shijing* ("The Book of

Poetry") have been selected in the creation of the piece. In using the Cantonese reading of Chinese characters, a late 19th century form of Estonian and modern Norwegian, the work challenged me as a composer to consider different prosodic elements. In my own reading, the tempo of each individual text stood out and contrasted drastically with one another, dependent on the individual language traits in its prosodic features. This has led me to design the vertical form of the work in three layers: 1.) the Chinese text can be slow or fast, as segmental quantity at the syllable level is flexible in a monosyllabic language, 2.) the Estonian text bears a metric regularity and rhythmic repetition, this became a sort of motor for the work's overall tempo and 3.) the Norwegian text by Gunvor Hofmo, the most capricious and fluid of the three, floats melodically above the structures implied by the first two languages.

In the tradition of the German *Lied* of 19th century composers,[26] *vihik (d)* is composed with a poetic reading of the selected texts. I have paid attention not only to prosodic features but the orthographic elements of the texts have also influenced my creative choices, as I was first and foremost an interpreter of the texts. I have then imagined how I would speak the texts aloud before enforcing prosodic rules that are embedded in the language or inventing my own ways of hearing the texts. Since this piece is shorter in duration than *isjoriske forsteder*, I have focused mainly on delivering the texts by the various writers rather than inventing a new formal structure encompassing a larger span of time.

Returning to the question in the beginning of this subchapter, the Estonian language's relationship to isochrony — although still theoretical in linguistic research[27] — differs itself from the other two languages. Although this may be a weak argument for analysis, as a composer with creative choices, I have chosen to insist that the three intrinsic varying tempi exist and that juxtaposing these elements will create interesting rhythmic results within my work.

---

[26] "When you read poetry, take a minute to think about the implications of pausing with a comma instead of a semicolon or a dash instead of a period. These articulations suggest nuances of poetic meaning and, of even greater importance for poetic setting, indicate something about the tempo of the poem as either read silently or spoken aloud. Such considerations assisted the composer in both determining musical tempos for *Lied* settings and in actually shaping the lines of poetry into vocal phrases" (Stein & Spillman 1996: 34).

[27] Laver pessimistically argues that "[e]mpirical investigations [of isochrony in speech rhythm] have unfortunately failed to show any such regularity, on any absolute basis. At best, the available empirical evidence could be said sometimes to have shown the existence of timing characteristics that fluctuate around a very approximately regular rhythm, with explanations usually being offered for the deviations by appeal to the varying structural identity of the constituent parts of the utterance" (Laver 1994: 523). To paraphrase, isochrony, while research has been conducted by various linguists on different languages, has not yet been proven as a universal phonetic feature across any single language.

This layering of three varying prosodic structures — or tempi, as I will refer to it as such from this point on — in *vihik (d)* can be observed in the excerpt highlighted in the following **Figure 13.0**:

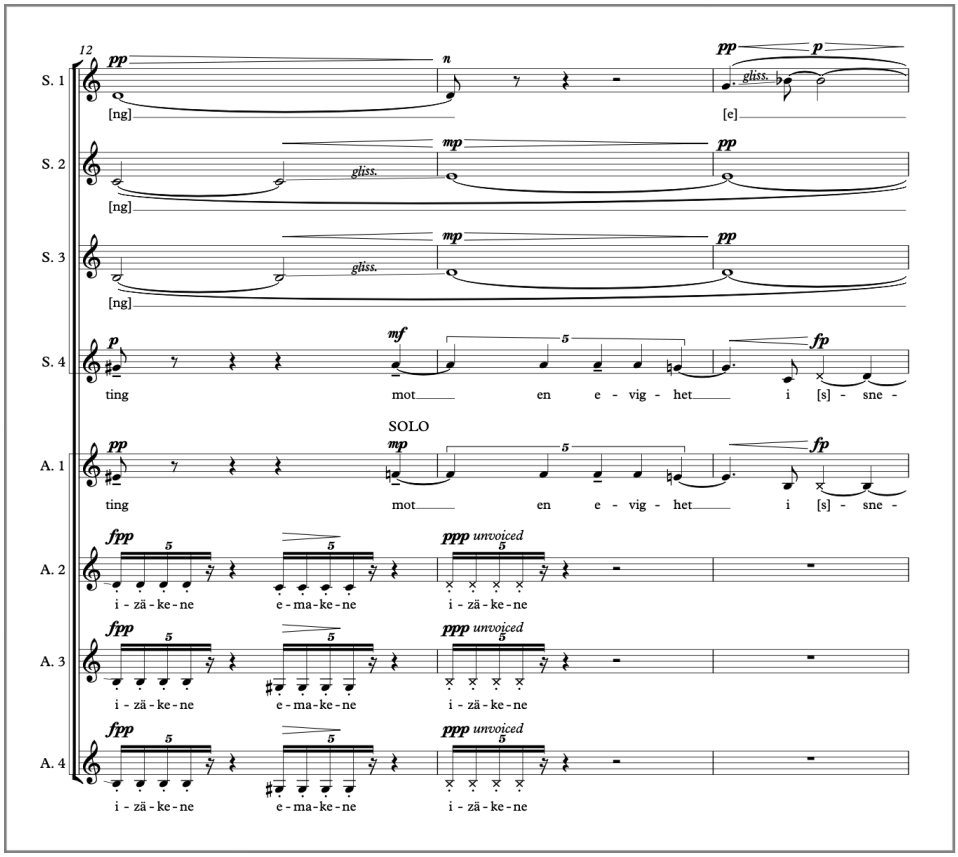



**Figure 13.0** One of the few sections in *vihik (d)* (2019) where all three language layers are sung simultaneously together.

Such a layering technique is effective for my compositions without interrupting the prosodic flow of or the meanings I have perceived in each individual text. As opposed to other forms of prose where stress, quantity and other types of accent are integral to delivering meaning, the poems selected for *vihik (d)* offer flexibility for different musical settings as they are already in, so to say, 'different tempi' thanks to the different phonological rules. I am interested in the juxtaposition of tempi aspect in the research phase of composing and I have eliminated some of the other potential text candidates which could be disruptive for the musical flow; this process is done through an intuitive read-aloud of the texts. In my sketches, the three texts in the three different languages constitute an appealing base structure which may be visualised in the following table:

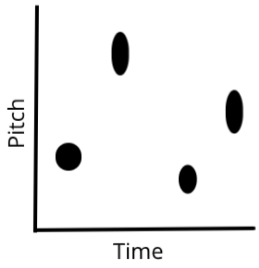
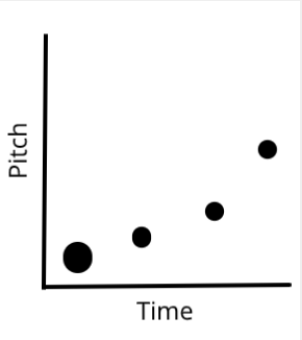
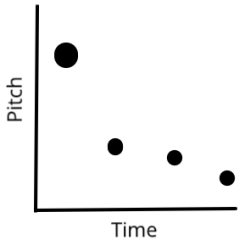
| Text | Written Language, Sung Language | Visualisation of a quadrisyllabic structure (e.g. word or phrase) of a language | Description of the Visualisation |
|---|---|---|---|
| 牆有茨 | Chinese, in Cantonese reading |  | Monosyllabic, irrelevant to quantity distinctions from utterance to utterance; each unit can be stretched or shortened in its prosodic position without affecting the semantic meaning. |
| Gunvor Hofmo's poems | Norwegian (Bokmål), in standard Eastern Norwegian reading |  | Emphatic moments are quantity dependent and prosodic structures are implied by the poet; most song-like as there is irregularity in the meter of the poems. There is a constant rising quality, which when repeated, contributing to a *rubato* sound. |
| Text from *Vana kannel* | 19th Century Estonian, in modern Estonian reading |  | Regular, syllables' quantity are distinct and supposedly isochronic in nature; the result can be recited as a chant. There is a constant falling quality, which contributes to the chant-likeness. |

**Table 3.0** Derivation of the three tempi of texts in *vihik (d)* through visualising rhythms.

On the other hand, it is arguable that texts are to perceived differently by individuals and so the rules implied by its authors can indeed be bent, misunderstood and deconstructed in different ways, regardless of its visual signs (e.g. regular rhymes, emphatic elements like capitalisation…). In the *Analysis of the Poetic Text*, Lotman opines that "[s]cansion never functions (for the contemporary adult reader) as a text; it creates the *background* for the perception of the text" (1976: 46). Indeed, the reading of a text does not necessarily equate to the vocal performance of the same passage. Lotman's argument implies that my feeling for the three different tempi across the different languages may be an imagined 'background' or a psychological preconception I have with regards to my understanding of the languages. Whether or not that is true, it has given me three musical layers in this work.

The three tempi have informed the construction of the musical form of *vihik (d)*. Between juxtaposing two or all three of the different languages, the polymetric structures that are established by the languages' rhythms have become a highlight for the two-movement work. The change between such structures serve as a focal point of the work. For example, the Chinese text and the Estonian text set up the mood in the beginning of movement one, where two stable pulses (i.e. slow wall of sound with the Chinese text and a rhythmic engine in the Estonian ostinato) create a static sound world. This structure is eventually interrupted by the yelled Norwegian line *'Ingen har ropt!'* (English: 'No one has yelled!') in the climax of the section and followed by a traditional melody and accompaniment section. The polymetric possibilities have extended the duration of my work with much needed variations and has enabled me to keep simple materials fresh.

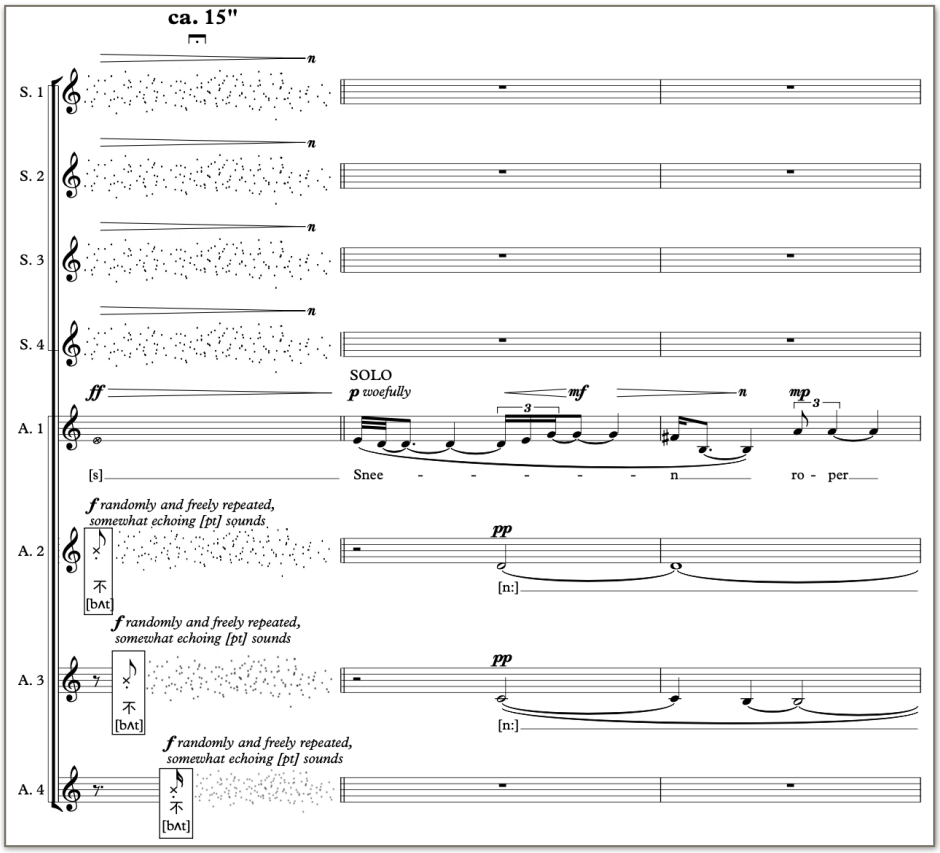



**Figure 14.0** Section in *vihik (d)* resolving the choir shouting *"Ingen har ropt!"*. The new section builds on a simple melody and accompaniment structure which is withheld up until this point of the piece.

In the final example of this subchapter, I would like to explore a third example which culminates the two concepts derived from *isjoriske forsteder* and *vihik (d)*. In *isjoriske forsteder*, I have discussed how quantity assists the construction larger rhythmic segments (e.g. syllables) that build up repetitive materials for use in longer works (i.e. formal construction), particularly in

evoking a cyclical sound world. Meanwhile, *vihik (d)* has taken the approach of considering the tempi, where quantities across texts of distinct languages are 'felt' in line with Lotman's poetic theory on rhythm in relation to different musical lines.

—

In my chamber opera *minn(i)e* (2020), the lengthy trilingual libretto written in Estonian, Cantonese and Norwegian presents a similar set of musical problems shown in the previous two works. The work is over an hour long in duration, making it a large-form work with the need to consider the unity of materials (i.e. motivic repetition and other compositional techniques), like *isjoriske forsteder*. The tempo aspect of the three different languages from *vihik (d)* is also a recurring question, as the language-respective sections of the libretto called for contrasting moods and pulses, much like the three layers in the earlier choir piece. Though, in stark contrast with *vihik (d)*, the libretto of *minn(i)e* does not bear visual elements of prosodic regularities (e.g. rhyme schemes, alliterations, etc.) as most of texts are in the form of free prose and dialogue. The solution to layer tempi according to language from *vihik (d)* is lost in this instance. Focusing on certain sections of the Estonian text by Maarja Kangro, one will find that many of the word and syllable segments are integral to the language's sound world that is reflected in many musical possibilities. Dissecting certain sections where word or syllabic repetitions do occur and parsing phonemes and vowels which repeat frequently within single words are both useful tools in constructing varying sound worlds. Done in contrast with the monosyllabic quality of the Cantonese reading of the Chinese text, I have noticed that some interesting musical results emerge:

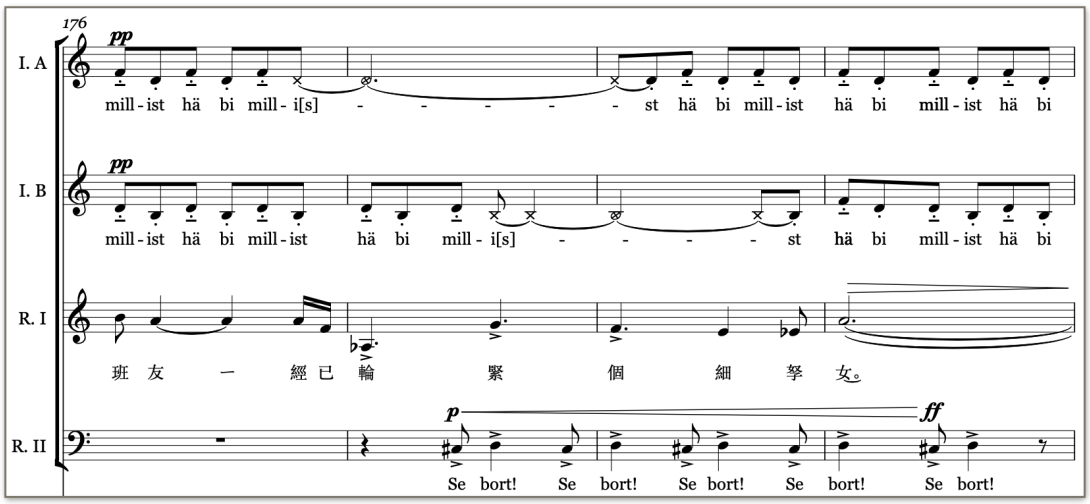



**Figure 15.0** Excerpt from the beginning of *minn(i)e.* Juxtaposition of an Estonian 'chant-like' passage and a Cantonese monologue, embellished by a Norwegian interjection.[28]

In **Figure 15.0** above, the juxtaposition of the Estonian phrase *'millist häbi'* (English: 'what shame') against one of Chapman Chen's interpretation of Minnie's diary create a metric opposition: the repeated Estonian syllables create an intrinsic strong-weak-medium-weak pattern (i.e. to beheard as a hemiola within the 6/8 meter), while each individual Chinese character fluctuates freely over the 6/8 meter. Moments like this are ubiquitous in the faster passages incorporating multiple languages throughout the chamber opera, as it creates opposition and clarity in a polyphonic texture. Furthermore, repeated elements — including the Norwegian fragment in this excerpt — maintains a background-foreground relationship which is vital to the dramaturgical aspects of staging.

Slower sections in *minn(i)e* generally do not rely on this meter-layering technique. The formal design of the piece is inspired by transitional gestures (i.e. transforming from one text to the next), since each language in *minn(i)e* has demanded various levels of focus and dramatic developments in varying sections. Returning to Lotman's concept of rhythm as a form of repetition and in order to create a sense of unity, I have focused on finding speech sounds or phonetic segments which signalled meanings or affects echoing different sections of the piece. One of such example in *minn(i)e* is the use of 'back consonants' like velar consonants like [k] to remind the audience of the coughing fit of the main character from the beginning of the work. The

[28] [28] IPA key for the Cantonese characters shown in the excerpt: 班 [pä:n˥] 友[jau˥] 經 [kɪŋ˥] 已 [jiˊ] 輪 [løn˩] 緊 [kɐn˥] 個 [kɔ:˧] 細 [sɐi˧] 孥 [nou˩] 女 [nɵyˊ]

accumulation and focus on velar articulations (i.e. [k] and [g]) among the four other onstage characters slowly devolve back into Minnie's asphyxiation and death sequence through a gradual increase of such consonants.

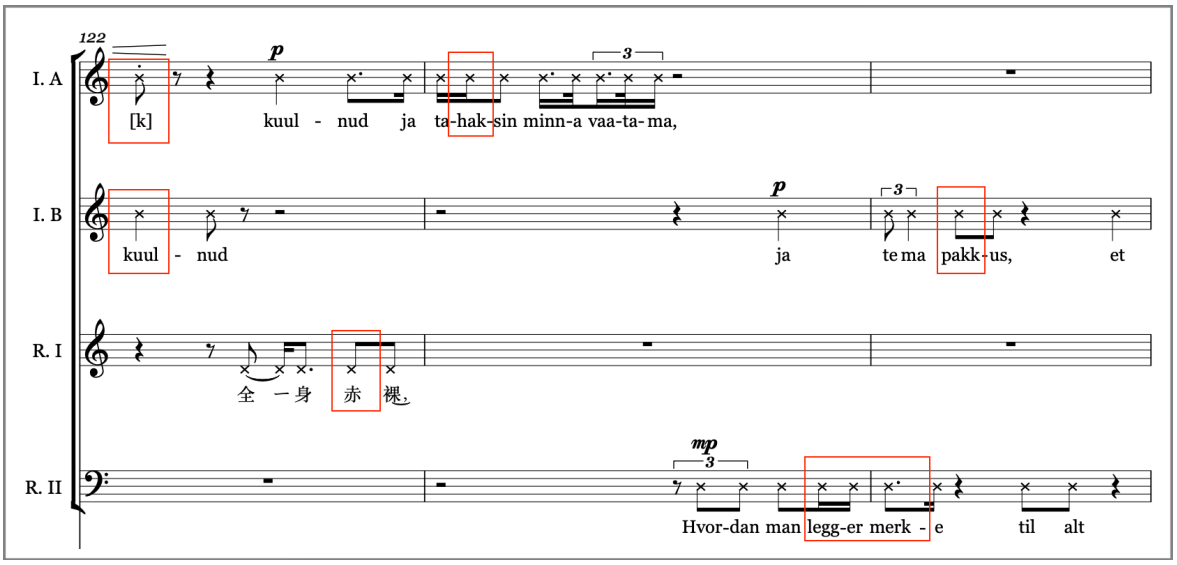



**Figure 16.0** Excerpt from the end of *minn(i)e.* Accumulating and focusing on velar consonants.[29]

In *minn(i)e*, where the texts come from distinct language families, I have been less able to rely on the physiological-psychological tie which Saussure has discussed in the quote from the start of this chapter. Taking materials from the phonemic level is most logical in working with texts across languages, since most tongues share at least some common phonemic structures (i.e. vowels or consonants). Naturally, larger phonetic segments such as the word, phrase and sentence can be viable options if the contexts permit. As this opera is expansive in temporal and semantic scopes, I am unable to cover all territories where such structures have shaped the work. Though, it can be concluded that a series of complex relationships between different moments are developed as a result of working with phonetic concepts.

[29] IPA key for the Cantonese characters shown in the excerpt: 全 [t͡ɕʰyːnɻ] 身 [sɐn˥] 赤 [t͡sʰɛːkʷ̚˧] 裸 [lɔː˧]

Section Summary:

- Above the phonemic level, phonetic segments can be identified as syllables, feet, word and sentence structures.
- The distinction of quantity segments (i.e. phonemes, syllables, feet, words) are highly dependent on the language in focus. Some languages classify timed-segments of languages through syllables while others through stress. Some phoneticians postulate that the syllable segment is classified through linguists' own intuition and understanding of a language, and are not always deemed a reliable scientific approach.
- Nonetheless, in the field of literature, Yury Lotman has argued that the repetition of speech sounds constitute 'rhythm', as the reader could feel where stress and emphasis of larger quantity segments. No experiments can produce quantitative results of such feelings but it is nonetheless a valid artistic approach in the creative spirit of poetry and music.
- Lotman's suggestion on rhythm in text as repetition of similar speech sounds has inspired a free use of repeated syllables to evoke formal unity in the work *isjoriske forsteder*. Recurring syllables tied to words with certain semantic associations constitute the concept of *semantic plane* and such structure became foundational for the work's structure.
- In *vihik (a)*, the phonetic segment of the word is explored. Words are developed into musical gestures and motives which create the work's miniature form.
- *messe norvégienne profane* has dabbled with the concept of prosodic feet of texts as a way to generate similar gestures which, in repetition, is a tool in unifying motives in a longer work.
- The trilingual choral work *vihik (d)* has taken its premise from the linguistic debate of isochrony such that the three different tempo structures are derived from each respective language of the texts. This layering of tempi has been an effective tool to create different forms of contrast.
- *minn(i)e,* a chamber opera, has combined both large-form thinking about rhythm and tempo layering from my earlier works. The result is a complex chain of relationships between different sections over the span of 70-minutes.

## 3.3 On the Pitch-Quantity Phenomenon

One phonetic feature — with a strong focus on Finnish and Estonian languages — which has interested me the most is the relationship between pitch and quantity in speech. As a non-native speaker of Finno-Ugric languages finding my way into its workings, the knowledge of durational patterns influencing meaning in speech is an attractive phenomenon which I see as potential for creating rhythmic patterns. On the contrary, two of my native tongues — Cantonese and Norwegian — possess tones or pitch patterns which play vital roles in delivering meanings and which has influenced my composing for many years. Looking at these phonetic-musical parameters (i.e. duration and pitch), my initial assumption is that the subjects occupy two separate acoustic parameters. However, this is incorrect. As I immersed myself deeper into the language learning and musical studies from the Finnic *Sprachbund*, I have begun to be aware of how quantity features of these languages are also tied to pitch. Such research is not uncommon among phoneticians. In spoken Estonian, where there is a three-way quantity distinction, several phonetic laboratory experiments by linguists have attempted to understand the pitch cues in quantity, particularly with the hypothesis that Q3 bears the biggest drop in pitch frequency over the timespan of the segment (Lippus 2011: 21), or proses as a feature in the 'overlength' phenomenon which compensates shorter preceding syllables (Ross and Lehiste 2001: 49). It must be noted that this phenomenon is a byproduct of the peak frequencies which distinguish the three quantity levels. The pitch level may be interrupted midway in an utterance, hence the regularity of the falling tone does not detract the listener from its meanings.

Meanwhile, in spoken Finnish, intonation patterns are more regular, as there are only two levels of quantity distinction: long and short. It has been claimed that "the second (unstressed) syllable is almost always lower than the first one, it can be analysed as a low tone" (Vainio et al. 2006: 121), which means that shorter segments tend to be associated with lower pitch. This highlighting of segments through pitch, duration or volume in single words is known as *lexical stress* (Crystal 1991: 328). Like in any languages, the word segment does not always reflect tones in larger prosodic structures such as sentence or paragraph. The general prosodic structure of Finnish is often heard as a gradual falling pitch pattern.[30] A similar phenomenon of falling pitch in the

[30] "Finnish intonation, at least the most "neutral pattern", can be described as a succession of declining rising-falling patterns (on content words), with an end reaching a very low F0 level (eventually containing non-modal phonation)" (Suomi, Toivanen and Ylitalo 2008: 115).

phrase structure is also observed in Estonian speech (Lippus 2011: 30), this phenomenon is noted as *downtrend* (Asu 2003: 63), where speech experiments have been conducted on questions starting with the word *kas*.

In Norwegian, it is well known that pitch accents are a distinctive feature in the spoken language. Although unlike Estonian or Finnish, where quantity correlate with pitch, the lexical stress phenomenon is nonetheless present (Kristoffersen 2015: 58). While quantity can be read through orthographical means in written Norwegian, the tonal implications are not explicit. I will return to the discussion of this topic in **Chapter 4** in greater details.

What does the knowledge on Finnish and Estonian pitch-quantity relationship impart for a composer? I return to reflect on the question of speech-likeness highlighted in the discussion of *Sprechstimme* in **Chapter 2.** The pertinent question for a composer is how speech-like do one wishes for the musical setting of a text to be. Being aware that Finnish and Estonian each have certain pitch-quantity restrictions (i.e. each Finnish line or sentence must fall in pitch with short syllables pitched lower; or each Estonian line may fluctuate in pitch dependent on the three-way quantity distinction), my compositions utilising these languages have been influenced by these phonetic assumptions. Noting that I am not a fluent speaker of the Finnic languages, these theories become important anchors for me to compare the resulting musical setting to the so-called natural speech utterance of texts. Furthermore, I find the natural patterns to be musically appealing. Dependent on the context surrounding the musical setting of a text, the pitch patterns may cause dissonance or consonance with the accompanying elements I imagined for the work. These relationships are generate musical gestures, moods and other affective qualities in a musical passage.

One of such example would be my piece for three female voices and wind ensemble, *kj/ærlige ord* (2018). In their original languages, the work utilises a Norwegian text fragment from Linda Gabrielsen's poetry in the art book collection *"Collage: Finn Aage Andersen"* and two lines from Estonian writer Juhan Liiv's poem *"Sa tulid"* ("You are coming"). In the simplest method of noting stressed versus unstressed syllables one can observe that many parts of the work follow such lexical stress rules in the musical setting of the Estonian text:

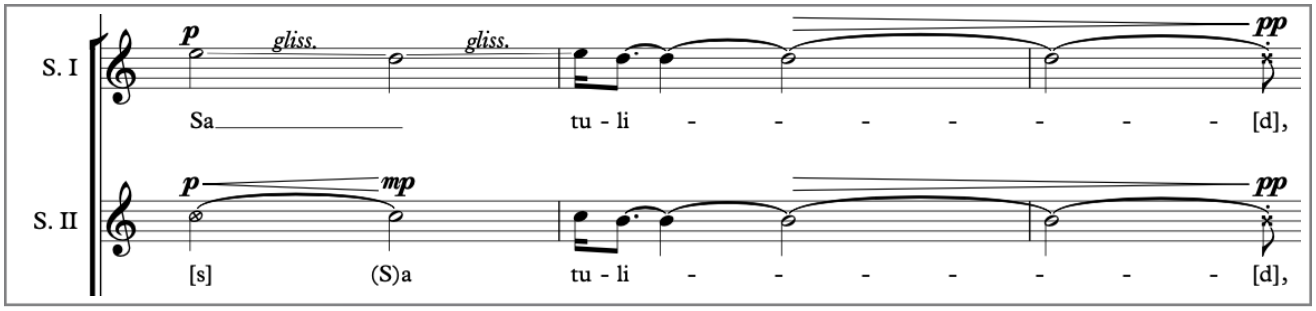
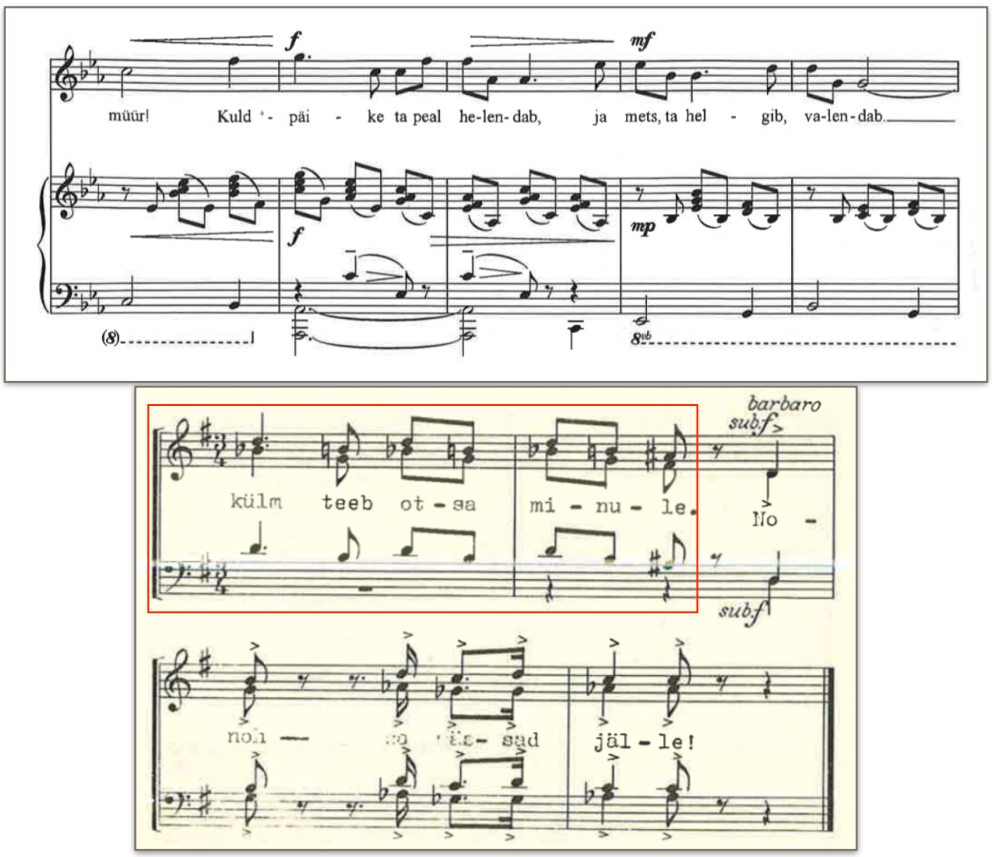



**Figure 17.0** Vocal excerpt from *kj/ærlige ord*, showcasing the lexical stress pattern of the word '*tulid*'.

This short fragment of *sa tulid*, serves as a repeated accompaniment gesture for the mezzo-soprano voice singing elaborate melodic lines throughout the piece. This fragment does not in fact imitate speech, since [s] in [sɑ] and [i] in *-lid* are stretched from their spoken durations. If the two words are to be heard for their semantic meanings without neutralising them into simply speech sounds, the focus must be placed on the most stressed segment in the fragment. The stress is on the [tu] Q1 syllable, and so it is placed on the downbeat of the bar with a note value which echoes the spoken value. To further distinguish this moment from the other syllables, according to the pitch-quantity phenomenon from earlier, *-lid* segment of the word unstressed and is therefore lower in pitch, which is once again reflected in the passage. Through understanding these details in setting music to Finnic languages, I have developed a tendency to think of musical stress and phrase structures differently whenever I work with texts in these tongues.

I would like to point out that phonetics is not the only source which has informed my compositional practice regarding the correlation between pitch and quantity. Having studied the vocal compositions of Estonian composer Veljo Tormis (1930 - 2017) as the springboard of my research, I noticed that many of his choral compositions and songs have pointed towards these phonetic hypotheses. I noted that — especially in earlier works where the composer relied on the intuitive understanding of his mother tongue — many long melodies set to complete, non-declamatory sentences (i.e. sentences which do not accent the last words for affective reasons) have a tendency to gradually fall in pitch:




**Figures 18.1 & 18.2** Excerpts from Veljo Tormis's earlier works *Neli kildu* (1955) and *Dialektilisi aforisme* (1978) conforming to the gradual pitch lowering in a non-declamatory sentence structure.

One lesson I have taken away from Tormis' score is the importance of understanding speech-likeness of a language. His repertoire encompasses a wide range of vocal works: between free song settings of Estonian poetry to choral works reflecting rules of runic sounds, his oeuvre offers the listener many possible interpretations of Estonian texts in various styles. The varying techniques he has employed in different contexts are valuable for non-native Estonian speaking composers aiming to work with texts of this language. Imitating musical patterns which reflect speech-likeness in a language which a composer is not fluent in may be done through score analysis as such. While it is possible to achieve creative results working against the naturalistic currents of speech — as in the description from **Chapter 2** of Igor Stravinsky and his treatment of English texts in *The Rake's Progress* — more creative and controlled results can be achieved through understanding the perspectives of native speakers. Orthographic representation of language is merely one aspect of interpreting speech-sounds and often the script of a language is quite deceptive (e.g. *through, thought, tough* and *ought* in English can give the reader a fair idea of this concept). Furthermore, acoustic phenomena such as the pitch-quantity correlation would not be observed if

the composer does not engage in reading scores with texts set to music by composers speaking a language natively together with comparative phonetic analyses.

I have initially failed to reflect some of these phonetic hypotheses in at the start of my research process, prior to studying this pitch-quantity phenomenon. I believe that even for native speakers of a language, such acoustic patterns are more felt than observed and analysed. On one hand, I am pointing out these instances in my work not as a corrective *mea culpa* but rather as a moment to unpack and tap into some musical potentials offered by my research work. For example, in my song cycle *natt-öö-夜* (2018) which is the first composition related to my research, I have noticed in hindsight that details regarding this falling melody pattern of Finnic languages have not been considered in my melodic construction:

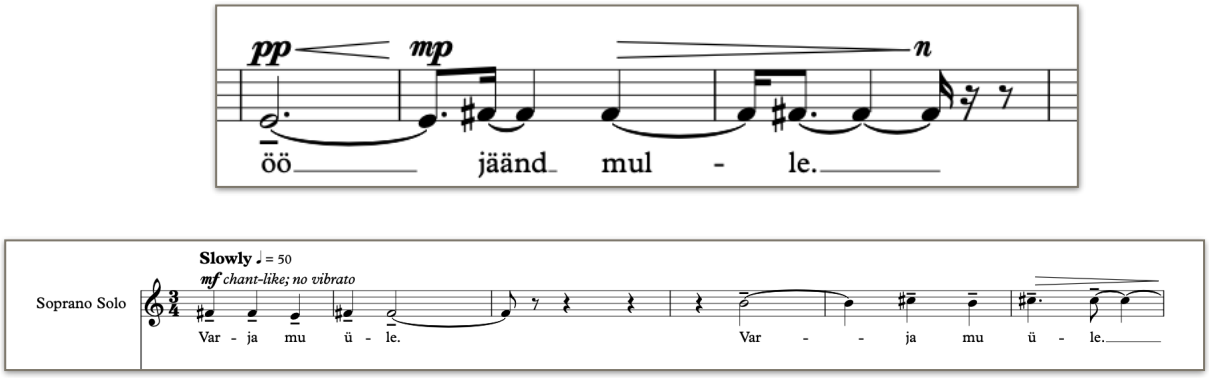



**Figure 19.1 & 19.2** Excerpts from *natt-öö-夜* (2018) showcasing unnatural pitch-quantity correlations in the setting of Estonian texts.

In the two excerpts showcased above, the reader can observe that the words *mulle* ('to me') and *üle* ('over') are endings to each their own respective non-declamatory sentences. To achieve an utterance which reflects a speech-like pattern, these disyllabic words, in accordance to the quantity-pitch hypothesis discussed, should be: 1.) lower in pitch than their preceding syllables, 2.) reflecting the stress relationship between the first and second syllable (i.e. first syllable higher in pitch or volume, more intense and the second lower, more relaxed). Indeed, one may argue that this is about song rather than speech-likeness. However, as the song cycle *natt-öö-夜* exchanges between sections of pure recitation of texts, controlled (i.e. rhythmically composed) recitation of texts and a variety of song settings, more variations could have been created with this awareness. Noting this aspect in my compositional process could have drastically changed the dramaturgy and the musical structures of the 35-minute work.

On the basis of discussing quantity, I have also disregarded the pre-boundary lengthening of utterances (Ross and Lehiste 2001: 54) in the two excerpts shown above, as these sentences are sometimes immediately followed by the next line of text with little pause. Doing so sets focus on the acousmatic listening of the texts, rather than a reflective listening of the words and their semantic content.

The relationship between pitch and segmental quantity insinuates intrinsic musical shapes embedded in the speech of Finno-Ugric languages. The falling pitch acoustic gestures make up a large portion of sentence structures in Finnish and Estonian. It can be understood as an artistic choice for composers to work with or against these acoustic implications. Since the Estonian language distinguishes three quantities, there is yet another possibility of reading pitch-quantity contents on the word and prosodic foot level, as well as the sentence and larger structures.

Section Summary:

- There is a correlation between pitch and quantity in Finnish and Estonian languages. Stressed elements within a word segment tend to begin on a higher pitch while unstressed elements tend to end on a lower pitch.
- Although it is a byproduct of the quantity element, in Estonian, the three-way quantity distinction theoretically implies that Q1 and Q2 are higher in pitch frequencies than Q3, which is 'overlengthened' by the utterance of lower pitches (i.e. it reaches a low pitch at some point in an utterance). The downtrends (i.e. falling motions) can though be interrupted by the speech pace and varying factors.
- In Finnish and Estonian, the short-long quantity distinction is not pegged to a respective pitch but makes nonetheless a falling pitch pattern in phonological phrases and sentence structures. This was noted in the research of Eva Liina Asu in her text *The Phonetics and Phonology of Estonian Intonation* (2003).
- Some vocal works of Veljo Tormis exhibit these patterns through what I speculate as intuition from the knowledge of his mother tongue; these traits could be studied by non-native Estonian speaking composers as a method to understand the habits in phrasing.
- Compositionally, these hypotheses in the points above have major musical connotations for songwriters or composers setting Finnic texts to music. In order to make a text passage more speech-like, composers could adapt pitch and rhythmic content according to these traits.

## 4. Compositional Topics on Tone

This chapter focuses on the phonetic concept of *tone* and its influence on different compositional processes. Similar to the previous chapter on quantity, I will initially outline select theoretical and scientific concepts of the topic before commenting on its connection to music composition. The chapter will offer a definition of the tonal language and detail how these languages may affect composers' creative choices. It will be followed by a discussion on how pitch affects non-tonal languages differently. After which there will be four subsections in this chapter covering specific subjects on tone pertaining to music making. I have compiled these topics based on the relevance to my artistic projects over the past few years. The resulting perspectives and methods aim to inspire other composers to explore further relationships between composition and research in phonetics.

---

Before the concept of *tone* in a language is defined, we must define the overarching concept of pitch. As one of the four major domains of perception in speech, perceiving pitch is about distinguishing the "frequency of vibration of the vocal folds" (Laver 1994: 152). A way to define pitch is through understanding the fundamental frequency ('F0'), measured in Hertz ('**Hz**') which stands for cycles per second (Crystal 1991: 265). While the study of pitch can be a quantitative practice (i.e. measuring and gathering data) within phonetics research, its descriptions across many fields — including music — is highly metaphorical. Listeners are able to compare a note with higher frequency pitch to another with a lower frequency pitch with the 'high versus low' binary, for example. This is perhaps where the science of speech sounds deviates from the artistic application of speech sounds, as performers and composers alike tend to use qualitative descriptions (e.g. a high screechy voice, a low growl…) for conveying pitch phenomena (Laver: 1994: 452). It is important to note that in the interdisciplinary approach of this research, I am alternating between interpreting such metaphors and linguists' quantitative research for the purpose of seeking out musical materials for creative work. These topics are regularly correlated with studies in cognitive sciences on acoustic perception; however, for my artistic projects, I am deriving compositional methods from my own understandings — at times purposeful misunderstandings — of such phenomena.

To shift the discussion towards tone, we must understand how the human voice apparatus produces tones. Without diving deep into the physiological studies of phonation,[31] one feature which set our speech production apart from most other sound sources is the complex ability to change the shape of the apparatus (i.e. shape of the mouth) and to control the various ways we interrupt airflow from the apparatus (e.g. muscular adjustments of our vocal folds) (Laver 1994: 187). The muscular function gives us therefore the ability to speak, sing, whistle or hum pitches. Since it is established among phoneticians that "[a]ll languages use variations in pitch to convey differences in meaning," (Ladefoged 1971: 84) these pitches function as *tones* in linguistics. There are two major classifications in the discussion of pitch in languages: 1.) "the use of pitch in *tone systems,* where it serves to differentiate units at the level of individual words and individual syllables" and 2.) "the use of pitch in *intonational systems*, where it serves to identify linguistic entities at levels higher than the word, at the phrase or sentence level" (Laver 1971: 462). Among the languages discussed in this dissertation, the spoken languages of Chinese (i.e. Cantonese and Mandarin) fall under the first category, Norwegian is a hybrid of the two with emphasis on *pitch accent* (I will return to this distinction later on in **Chapter 4.2**) and Estonian falls under the second category.

Let us first explore the first classification with Cantonese. Politically considered as a dialect[32] of the Chinese language family, the Cantonese language presents an interesting challenge for composers wishing to set texts of this tongue to music. As a tonal language with six tones (Schellenberg 2013: 7), composers are presented with a unique challenge where the music's pitch materials take into consideration of the speech's tonal contours when composing melodic lines. If the tones are not followed precisely in the musical setting of a text as in the spoken language, the meaning of the text can be altered and misunderstood. A meaningful text in its spoken form can lose its semantic values or become nonsensical if the contour is not followed.

The concept of tones in linguistics is not the same as the one in music. In music, particularly in the context of classical music, we often make the assumption that a tone is a single, unwavering

---

[31] For a thorough scientific explanation of how the human voice apparatus creates speech sounds, I would recommend studying the chapters on phonation in John Laver's *Principles of Phonetics* (1994: 184) or Peter Ladefoged's *Preliminaries to Linguistic Phonetics* (1971: 7).

[32] For further clarification, Gina Anne Tam's text *Dialect and Nationalism in China, 1860-1960* (2020) provides a thorough historical overview of Chinese *dialects,* a terminology that is both politically and linguistically controversial even today.

pitch of distinct frequencies (e.g. a clarinet holds the note A at 440Hz). However, in phonetics, tones are relative can be contours of pitches. "Tones [in Cantonese speech] are always relative, they are not absolute. Each individual speaker naturally uses different pitches for tones depending on the speaking range of his/her voice; but also, over the course of an utterance, a given tone may change its actual pitch due, for example, to the natural tendency for speech to fall in pitch over the course of an utterance" (Schellenberg 2013: 5). To paraphrase, the varying *tessituras* of different speakers do not affect how perceivers understand them and this would function similarly in song. The native Cantonese listener is able to distinguish tones in the same way, regardless of different speakers. The reader should not equate linguistic tone with an absolute pitch. This is a common presumption by non-tonal language speakers and the concept is simply false.

To visually exemplify the six distinct Cantonese tones, early twentieth-century Chinese linguist Yuen Ren Chao (1892 - 1982) developed a system where pitches are approximated with 1 as the 'lowest' and 5 as the 'highest':

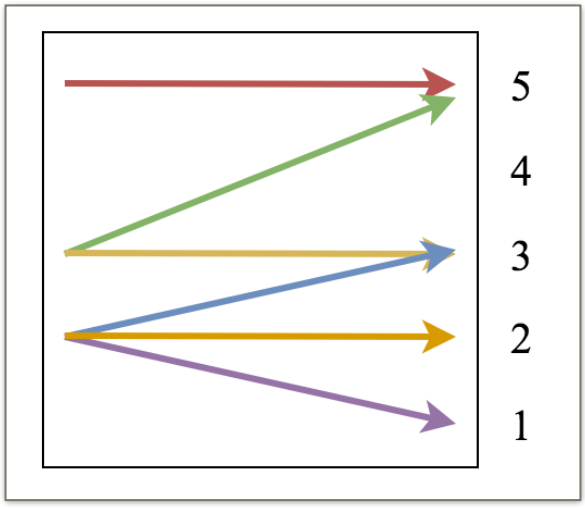



**Figure 20.0** The six Cantonese tones visualised according to the Chao system.

This system is applicable not only in Cantonese but in other tonal languages as well. But as one can observe in **Figure 20.0**, some tones are rising or falling pitches (known as contour tones) while others are static (known as register tones) (Schellenberg 2013: 7). Despite the two varying approaches to pitch, each arrow is considered to be distinct tones. Below is a chart exemplifying the six tones on the syllable [si], which can each bear different meanings:

| [si] as different characters and 'English translation' | Tone Classification | Chao System Tones | Description | IPA Symbol |
|---|---|---|---|---|
| 私 **'private'** | 1 | si55 | High level | ˥ |
| 史 **'history'** | 2 | si35 | High rising | ˧˥ |
| 嗜 **'to try'** | 3 | si33 | Mid level | ˧ |
| 時 **'time'** | 4 | si21 | Low falling | ˨˩ |
| 市 **'city or market'** | 5 | si23 | Low rising | ˨˧ |
| 是 **'yes or is'** | 6 | si22 | Low level | ˨ |

**Table 4.0** Tonal variations and possible meanings of the syllable [si] in Cantonese.

One may assume that setting Cantonese texts to music means that there are many restrictions regarding pitch. This is a correct assumption. Research by Murray Henry Schellenberg has shown that several of the world's tonal languages such as Cantonese, Thai, Ewe and Vietnamese, require a certain percentage of pitch correspondence (i.e. the melody must reflect the pitch contour of the spoken passage) must be achieved in order for the listener to comprehend the semantic meanings of the text. The samples of songs used in the research showed that Cantonese, alongside Zulu, have achieved a 92% in their correspondence of melodic phrases being restricted to the tonal contours (Schellenberg 2013: 22), the highest percentage among the studied languages. In the later subsections of this chapter, particularly in **Chapter 4.1**, I will describe the challenges of working with Cantonese texts in compositions. There are definitely many other topical discussions on Cantonese tones and, to a greater extent, tones in other Chinese languages. However, as the scope of this dissertation focuses on the relationship between phonetic features and composition, many of these phonetically interesting aspects will unfortunately not be discussed. Though, I hope that artistic researchers interested in the topic may one day continue to seek out creative potentials in this direction.

Turning to Norwegian, I have earlier mentioned in this chapter that the language's use of pitch is classified as a hybrid between the tone system (like Cantonese and Mandarin) and the intonational system (like Estonian or English). Swedish and Norwegian differentiate themselves from other members of the Germanic language family because of their utilisation of tones (Kristoffersen 2015: 66). Spoken Norwegian contains two pitch accents (described in Norwegian as *tonelag* or 'tone level'), regardless of dialectical differences found across the country: low and

high.[33] Furthermore, each dialect does not bear 'oppositional' tones (i.e. low-high *and* high-low), unlike the variations found in Cantonese. This is perhaps why these distinctions are sometimes interpreted as 'melodies' instead of tones. The two pitch accents, depending on the dialect are always low-high or high-low with pitch fluctuation in one of the tones (**Figure 20.0** below shows an example from Standard Eastern Norwegian). They are classified by Norwegian linguists as (L) 'low to high' or (H) 'high to low'.[34] In the terms of Chao system in the Cantonese discussion earlier, the tones can be visualised:

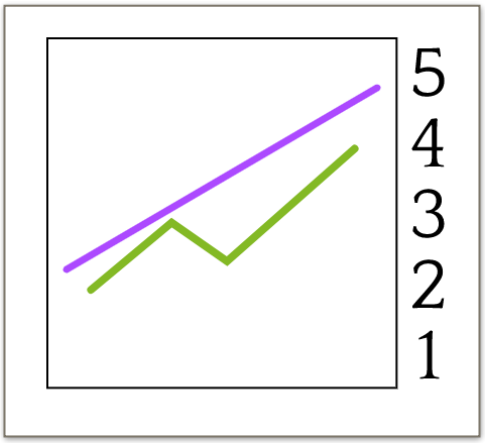



**Figure 20.0** The two tones of Standard Eastern Norwegian, tone 1, in purple, rises from low to high while tone 2 dips slightly before rising.

One of the most well-known examples of these two tones would be the *bønner-bønder* pair. Both of these words contain the same phonemic make up but are distinguished by the speaker's tones. In Standard Eastern Norwegian, tone 1 of *bønder* ('a farmer'; [bœndər¹]) would rise fairly straightforward in pitch while tone 2 of *bønner* ('beans'; [bœnər²]) dips to a lower pitch before rising again. In musical terms, the two contours can be visualised in the following way (although the intervals are approximate):

---

[33] *Vi har [med andre ord] to mulige "melodier" vi knytte til en trykkstavelse, og disse to melodiene er betydningsskillende* (Kristoffersen 2015: 58). "We have [in other words] two possible "melodies" we tie to a stress syllable, and these two melodies are differentiating in meanings" (translation by author).

[34] "'I interpret minima, that is, falling pitch followed by a rising pitch, as realizations of a phonologically low tone, and maxima, that is, rising pitch followed by falling pitch, as realizing a phonological high tone.' *Dette innebærer at en analyse av nordiske tonale forhold bare analyserer ut to mulige registertoner, L og H, og at for eksempel to tonale maksima godt kan ligge på ulike frekvensnivåer og likevel bli regnet som forekomster av H*" (Hognestad 2012: 8). "This assumes that an analysis of Nordic tonal relationships only results in two possible register tones, L and H, and that, for example, two tonal maxima can very well lie on the same frequency level and still be labeled as the occurrence of H' (translation by author).

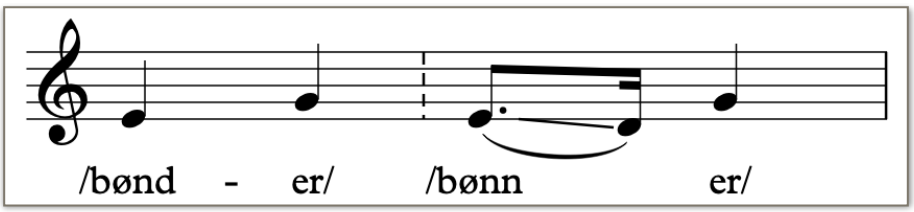



**Figure 21.0** The *bønder-bønner* pair, showcasing the tone or *tonelag* distinction of phonemically similar words.

The many dialects found in Norway make this discussion of tones socio-geographical: in most Western Norwegian dialects, the tones are mirrored (i.e. high-low predominates the two tones, contrasting with Eastern Norwegian's low-high). These variations are mostly lexically consistent, meaning that all words with tone 1 and tone 2 remain tone 1 and tone 2. The pitch contour is simply flipped. However, not all dialects are properly documented to universally verify this claim (Kristoffersen 2015: 68-9). This is partly why these 'tones' are sometimes referred to as pitch accents instead.[35] Contrasting with Cantonese, the distinction of tones in spoken Norwegian does not affect the lexical comprehension of individual syllables or words, as the listener can also derive meaning from contexts within larger structures such as the phrase or the sentence. Even if the speaker mispronounces the tones, there are other linguistic cues which will inform the perceiver of the speaker's intention and the correct word. Norwegian spoken dialects have a weak semantic correlation to the phonetic feature of pitch compared to full-fledged tonal languages. Therefore, Norwegian dialects do not fully fall under the tone system category but only contain some traits pertinent to tonal languages.

In the context of song, spoken Norwegian does not share the strictness of speech-song correspondence that has been outlined by Schellenberg mentioned earlier in this chapter. In most cases, setting Norwegian texts to song function similarly to how one would do so in English: through emphasising prosodic foot and stress. Although, in the spirit of the composer's creativity, adhering to these intrinsic melodic structures of the Norwegian language can contribute to generating a new pool of pitch materials for composition, not to mention nuances in working with this language.

[35] As exemplified in the mid-twentieth century article "Pitch Accent and Tonemic Juncture in Scandinavian" (1963) by Einar Haugen from the journal *Monatshefte* and also later linguists such as Patrik Bye in his *Evolutionary typology and Scandinavian Pitch Accents* (2004).

The upcoming subsection of this chapter (**Chapter 4.1**) titled *Relativity of Pitch in Tonal Languages and Its Impact on Song* will mainly discuss the challenges and strategies I have developed for incorporating Chinese texts in their Cantonese reading in composition. The second subsection (**Chapter 4.2**) *Composing with Norwegian 'Pitch Accents' versus Cantonese 'Tones'* compares and contrasts the different musical materials (mainly pitch materials) I have developed in working with two different languages related to tones. In **Chapter 4.3**, *Pitch in Speech and Its Musical Representation*, pitch structures in speech inform the creation of larger temporal structures in music composition will be discussed. Inherent characteristics of larger level segments (e.g. prosodic foot, word or phrase) from Norwegian 'melodies' and the tone sequences in Chinese varieties may be expanded in musical thinking. Lastly, in *the Composer and the Performance of Songs Set to Texts of Tonal Languages* (**Chapter 4.4**), answers a series of practical questions from my performer collaborators who do not speak a tonal language. The chapter breaks down presumptions about tones by guiding the performers through effective notation and rehearsal techniques.

Section Summary:

- Pitch is one of the four major phonetic domains in the discussion of human voice apparatus's ability to produce speech sounds. In phonetics, 'tones' are pitch recurrences in the phonology of a language which distinguish semantic meanings or affects for the perceiver.
- There are two classifications of tones in languages: those using a tone system (e.g. Chinese spoken varieties, Thai, Ewe) and those using an intonational system (e.g. most Germanic languages, Romance languages). Some Scandinavian languages (i.e. Norwegian and Swedish) bear hybrid features.
- Tones can be contour tones (i.e. moving from low to high, vice versa or in more complex configurations) or register tones (i.e. static on a pitch).
- Spoken Cantonese possesses six different tones to distinguish meanings in phonemically like segments, while Norwegian has two contour tones, or pitch accents, which vary from dialect to dialect.
- Spoken Cantonese requires quite strict speech-song correspondence, while Norwegian does not.

## 4.1 Relativity of Pitch in Tonal Languages and Its Impact on Song

This subchapter examines some strategies and concepts I have used or developed for composing music to Chinese texts in the Cantonese reading. The relativity of pitch in the discussion of tones have been a popular subject of linguistic studies in many of the world's languages, with much focus on languages found in Sub-Saharan Africa, Southeast Asia and East Asia (Schellenberg 2013: 26). As each language has its own set of phonological rules with varying impacts on different linguistic subjects (e.g. syntax, semantics…), the focus will be on select Chinese dialects and Scandinavian languages in the scope of this dissertation. The vastness of this subject must though be acknowledged. Composers and enthusiasts of the subject should explore the different dynamics between music and speech in other tone languages and be wary of the great variations between language sets.

The tones in Cantonese and their functions in song is peculiar compared to other Chinese dialects. Linguists Bin Li and Chung-Nin Choi from the City University of Hong Kong have stated that "Mandarin songs, especially the more contemporary ones, no longer pay close attention to the mapping between tones and tunes, whereas modern Cantonese songs may still retain such tradition" (2016: 332). The researchers argue that correspondence of tones to its musical setting is seen as a 'tradition' in Mandarin. On the contrary, linguist Murray Schellenberg from the University of British Columbia has argued that "[w]hile diverging from the speech melody occasionally does not significantly impair comprehension, a general free-for-all is not the best choice for a tone language, either. Matching melodies will certainly enhance comprehension, so it is in a culture's best interests to match, all other things being equal" (Schellenberg 2013: 36-37). These different views are interesting for a composer because one is in the position of generating content which legitimise or disprove these perspectives. I tend to share Schellenberg's opinion that composing music to Cantonese texts should adhere to the tonal structure in my own practice. Although, I do not entirely agree to the notion of the 'culture's best interests,' as unorthodox interpretations and misunderstandings of texts in music should also be legitimised and possible forms of expression.

American linguist Morris Halle (1923 - 2018) wrote in his essay *Why and How do We Study the Sounds of Speech?* that "[s]tudents of language, no matter what their field of specialization, are interested in the question of how human beings communicate by means of language in general, and

by means of a given language in particular" (Halle 1954). In the same vein, abiding to the tone rules of a language is not for the listener's sake of clarity in word-for-word comprehension. These phonological tones — evolving in the history of a language — are communicating subtexts and connotations more than merely my musical re-imagination of a found or written text. Therefore, purist disputes on what rules are to be 'respected' and what are to be 'discarded' are purely personal.

Returning to the discussion on the strategies I have employed for working with Cantonese texts, it should be noted that I have an advantage of being a fluent speaker of the language and that I could rely on my intuition when solving musical problems in this language. Intuition is insufficient for explaining why music is set to the text this way, however. This is true especially if the purpose of this dissertation is to offer insight on how other composers can work with these phenomena. Parameters are thus developed for the construction of different works.

My earliest artistic output in this research project, *natt-öö-夜* (2018), is a work for experimentation in developing strategies in composing for 'tonal' texts. One point of interest in this work is that, using the same poem by Song-dynasty poet He Zhu, I have set music to both the modern Mandarin and Cantonese readings. This has allowed me to discover two different pitch materials[36] derived from the same text. Before proceeding to comparing how these two sets of pitch materials are derived, here is a diagram showing the four tones of modern spoken Mandarin developed from the Chao system:

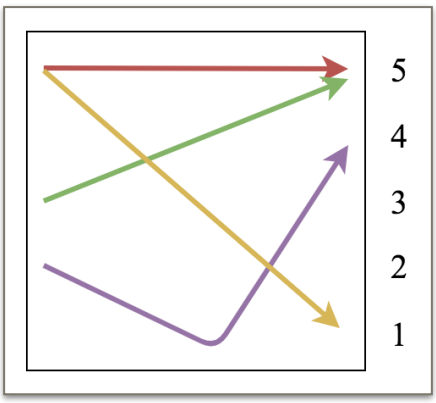



**Figure 22.0** The four tones of modern spoken Mandarin visually represented according to the Chao system.

[36] Particularly on pitch contour structures, since I have mentioned in the previous section of this chapter that the tones between the two Chinese dialects are completely different (see page 56 - 57).

Mandarin's four tones are drastically different from the six of Cantonese. One can quickly deduce the two obvious differences: the contours are more drastic (e.g. the 'dipping' third tone, indicated in purple in **Figure 22.0** is not present in Cantonese tones) and there are two fewer tones. Despite these differences, the two spoken varieties of Chinese share a common lexicon[37] and are therefore mutually intelligible in its written form but have noticeably different phonetic properties. In other words, the poem in *natt-öö-夜* by He Zhu is read in two different ways for varying results. The following chart shows the different pronunciations realised in IPA (note that the indexical symbols for tones differ between Cantonese and Mandarin):

| **Text in Traditional Chinese Characters** | **IPA in Mandarin** | **IPA in Cantonese** |
|---|---|---|
| **子夜歌 三更月 ('Song of the Midnight, Moon of the Midnight')** | [tsɻ:˥˩ je:˥˩ kɤ:˩ sæn˩ kəŋ˩ ɥe:˥˩] | [tsɪ:˧˩ jɛ:˨ kɔ:˩ sɐ:m˩ kɐŋ˩ jy:t̚˨] |

**Table 5.0** The two Chinese readings of the title of He Zhu's poem "子夜歌 三更月".

Even just looking at the title, the kinds of pitch content suggested by the phonological readings are suggesting different musical settings. Deviating from the natural setting is possible but in the case of *natt-öö-夜*, I have focused on understanding the fundamentals of phonological tones and music making. To keep the treatment simple, one method to treat these characters could be to reflect the pitch structures from character to character, as in the following passage (in Cantonese) where the register tones are reflected in the melodic writing:

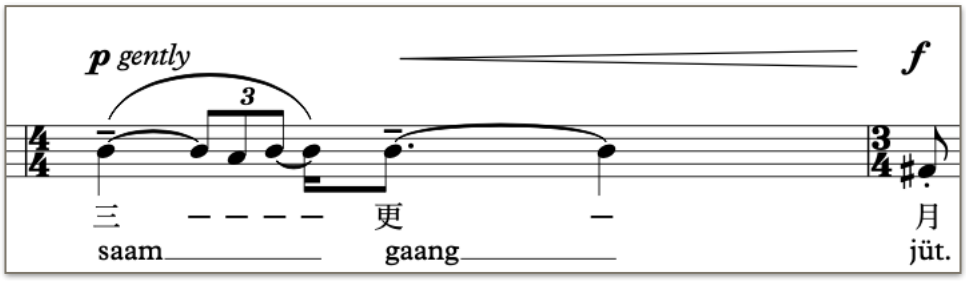



**Figure 23.0** The musical setting of "三更月" in Cantonese, reflecting pitch relativity between the three syllables. The transliteration shown below the characters is adapted into Estonian orthography.

[37] "In its most general sense, the term is synonymous with *vocabulary*" (Crystal 1991: 200).

In contrast to the register tones in Cantonese, the falling fourth tone in Mandarin in the character 月[ɥɛ:ꓥ] is reflected in the respective iteration within the piece (the cross-noteheads indicate less focus on exact pitch):

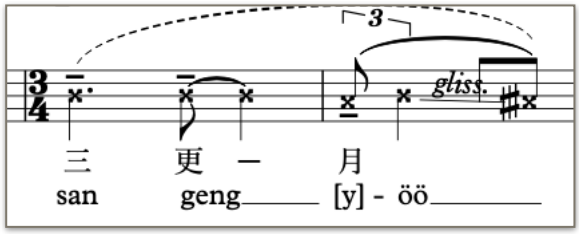



**Figure 24.0** The musical setting of "三更月" in Mandarin, reflecting pitch contours embedded in individual syllables.

The pitch materials used in contrasting the two spoken varieties are not drastically different, as the two first characters are register tones (i.e. on a single pitch) and the only variation occurs in the third character 月[ɥɛ:ꓥ], where its Cantonese reading is a checked tone[38] and its Mandarin reading is an open falling pitch contour. Though, these short passages have sparked a new idea in my melodic expansion of register tones, especially in maintaining its semantic comprehension. Whenever a speech tone is unwavering (i.e. a register tone), it can first be established and any following melismatic passages then will not disturb its meanings. One example could be the character 鵑 [ky:nl] read in Cantonese, as exemplified in the following passage (for practical purposes before the premiere, the transliteration in the score is realised according to Estonian orthographic approximations, and does not reflect the IPA transcription in anyway):

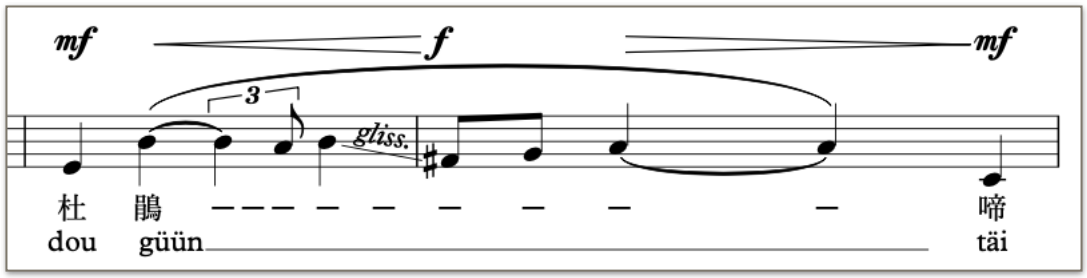



**Figure 25.0** A melismatic passage taking advantage of semantic consistency in the established high register tone.

[38] "[S]yllables ending in an occlusive coda (-p,-t,-k)" (Xu, et al. 2010: 459).

Naturally, the longer the melismatic passage, the more likely the semantic comprehension will be affected within the larger prosodic segment (e.g. word or sentence). This holds true for any texts, with or without tones. While the horizontality of melodic structures is the simplest focus, it is possible to incorporate tones vertically (i.e. as harmonic structures). In the following example in *vihik (d)* (2019), my trilingual work for women's choir, I have developed the first chord out from the pitch contours of a text fragment from the ancient Chinese collection 詩經 (*'Shijing'* or the 'Book of Poetry'):

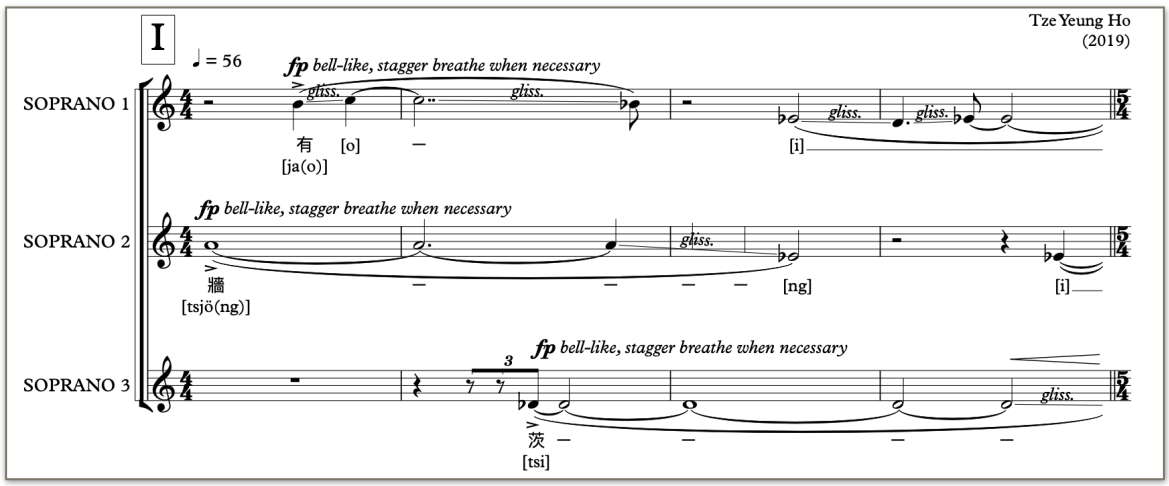



**Figure 26.0** The opening bars of *vihik (d)* (2019) where the first three Chinese characters from "牆有茨" build the harmony of the opening chord.

The first three characters of the text "牆有茨" (roughly translated as 'the walls have thorns') are read in IPA as such: [tʃœŋɻ], [jɐuɻ] and [tsɪɻ].[39] The pitch sequence calls for a low tone then a rising tone from low then back to a low tone again. The opening chord in *vihik (d)* reflects such a relative contour motion: soprano 2 voice begins with the pitch A on the first syllabic segment, then soprano 1 joins one whole tone above with a further glide upwards and finally soprano 3 fulfils the third syllabic segment a major seventh below. As mentioned in the article *Tone Contour Realization In Sung Cantonese* (Schellenberg 2011), there are no absolute intervals when discussing the levels of low, mid or high. Spoken tones are relative and even though the first and third sung pitches of the excerpt in **Figure 26.0** are not the same pitches on the score, the contour will be perceived in the same way. The separate attacks with ample time between each syllable also keeps the semantic information intact. Through this method of stacking syllables in Cantonese, a composer could consider voicing of chords or harmony through the lens of tonal contours.

[39] It must be noted that the transliteration in **Figure 26.0** are adapted to a Swedish orthography instead of a purely IPA one, as the piece was written for a Swedish-speaking choir.

In this voice stacking method for Cantonese, the composer should beware of the intersegmental coordination of phonemic structures in each syllable or character. Each Chinese character bear distinct phonological rules. “Even as single segments, a segment has contextual neighbours of utterance-marginal silence on both sides. In all cases, the articulatory events that make up the segment concerned have to be co-ordinated with those of the neighbouring context” (Laver 1994: 339). Therefore, in cases of diphthongs or other complex structures, the tone (or voicedness) should be placed on the correct phoneme, otherwise the correct Chinese character will not be reflected. For example, the character 有[jɐuˇ] in **Figure 26.0** contains an [ɐu] diphthong which is parsed due to the need for a long note but the focus must ultimately fall on [u] in order for the perceiver to understand 有. Stressing on and lengthening the [ɐ] is not possible from the semantic cognition perspective.

———

The previous examples have focused mainly on Cantonese examples and have tendencies to zoom in on relatively simple vertical (e.g. harmony) and horizontal (e.g. melody) structures of song. The looser ‘rules’ of the tongue’s *tonelag* in Norwegian has allowed me opportunities to musically interpret speech tones in other manners. Norwegian tones are more like melodies over multi-syllabic words, unlike the Chinese varieties. Therefore, the segmentation process of Norwegian phonetics can be considerably different from those in Chinese. While one may focus on phonemic level in Cantonese, the same cannot be said for spoken Norwegian varieties. The focus on syllable and prosodic foot levels makes the discussion of speech tones and song is different between the two languages. Spanning a tone over a word segment means that the melodic content of pitch involves longer durations. Different studies on the subject have claimed differing opinions on the two-tone distinction. It has been claimed that only tone 1 in Norwegian requires lexical cues[40] and there are researchers who even claim as far as that there is only one true tone (i.e. tone 2) in spoken Norwegian, since tone 1 is simply a part of the the language’s intonation and lacks lexical significance.[41]

---

[40] "*I et arbeid (2005) lanserer Aditi Lahiri og hennes medarbeidere tanken om at det er tonelag 1 som krever leksikalsk spesifikasjon og dermed er det markerte tonelaget*” (Hognestad 2012: 2). “In a work by Aditi Lahiri and her collaborator (2005), it is stated that *tonelag* 1 is which that requires lexical specification and thereby is the marked *tonelag*.” (Translation by author).

[41] *“Tanken utvikles av Haugen og Joos (1952), som langt på vei sier at norsk egentlig bare har ett tonelag, nemlig tonelag 2, mens tonelag 1 kan analyseres som ”rein intonasjon*” (Hognestad 2012: 3). “The thought is developed by Haugen and Joos (1952), which goes as far as saying that Norwegian only has one *tonelag*, namely *tonelag 2*, while *tonelag* 1 can be analysed as “pure intonation.” (Translation by author).

As phonetic segments above the phonemic level tend to occupy longer durations of time, the effect of playing with its tone structures bears another set of implications. Unlike spoken Chinese varieties, the deviation from speech tone structures can be a useful idea for creating different musical characters. Thanks to the freedom of Norwegian speech tones, utterances can be contrastive between speech and song without altering the semantic content of the text. In *messe norvégienne profane* (2018), I have utilised the contrasting treatment of speech tones to create two 'characters' in the piece. The phrase *Fader vår* ('Our Father') is first spoken and then repeated in song:

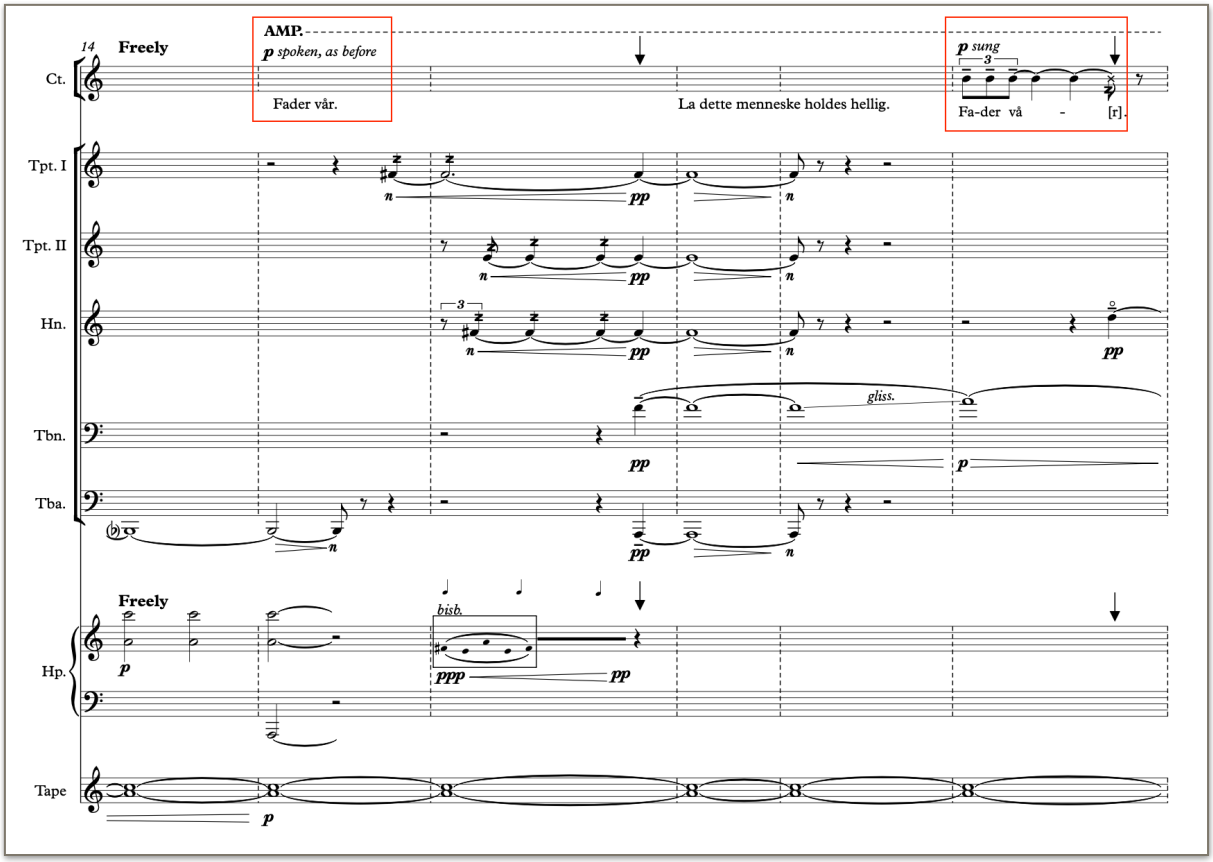
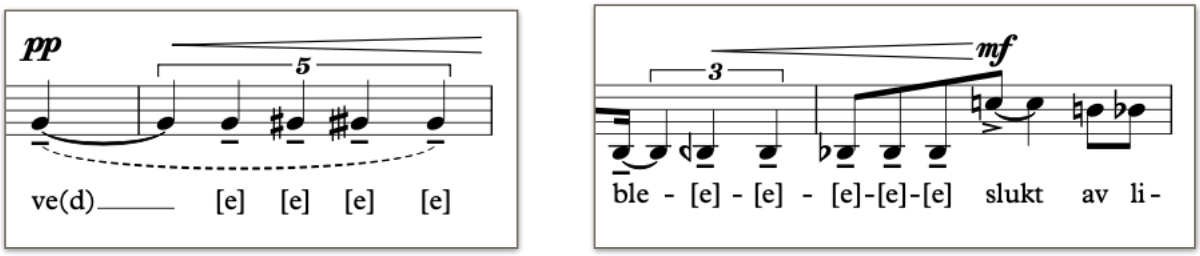



**Figure 27.0** An excerpt from *messe norvégienne profane* (2018) with contrasting treatment of the same phrase.

Through the repeated utterances, the speech tone pattern is distinguished from the sung pattern. *Fader vår* in the first red box **Figure 27.0**, when read in Standard Eastern Norwegian, could be spliced as such: [fa] is low, [dr] (or *-der*) is low and [vɔːr] (or *vår*) is where the rise of tone 1 occurs. This is contrasted with the single, unwavering pitch treatment of the sung phrase in the second red box. I have chosen to set the text in such a manner because the song quality sets focus to or zooms into the seance quality of the opening of an imagined church ritual. The *messe*, based on my muddled childhood impressions of church services, blends spoken passages and sung lines in a way where spoken passages represent the pastor's sermon and the sung passages of the mass. As

prosodic structures in liturgical texts in Latin have often been freely and musically realised by composers throughout the ages,[42] I have attempted to keep this style of text treatment even in Norwegian. The alternation between Norwegian speech and Norwegian song gives the work an inherent contrast and refreshes the focus of the listener such that isolate semantic content could be emphasised.

Although tones are considered in multi-syllabic segments in spoken Norwegian, it is not rare to find monosyllabic words of significance in the language. I have come across some words in my artistic output which carry an interesting tone-pitch features. For example, the words *ble* ('became'; [ble:¹]) and *ved* ('by/at' [ʋe:¹]) often play vital roles in some poems used in my music. As single syllable words ending on a vowel without consonant codas, they must be realised as tone 1, as monosyllables cannot bear lexical tones (Bye 2004: 4). In other words, words as such are supposedly rising in pitch for an indefinite duration of time, due to the lack of a coda segment. One of such example would be from the work *vihik (c)* (2019) for mezzo-soprano and orchestra. I have used a text by Norwegian writer Hanne Bramness (b. 1959) in this work. Unlike the other Norwegian texts I have used for the other works in my artistic research, Bramness favoured longer verses and prosodic segments in her poetry for detailed imageries and moods. Monosyllabic words have a different significance in this type of poetry, as single syllables tend to stand alone in the larger prosodic patterns. The temporal quality of such syllables (i.e. stress emphasised by longer duration) has an interesting impact on song:




**Figures 28.1 and 28.2** Two excerpts from *vihik (c)* (2019) showcasing the treatment of monosyllabic words of focus.

As the rising tone 1 in the words *ved* and *ble* do not have a coda segment indicating a strict point of release in the sound, the glissando quality that is implied by the words have no limits of an

[42] "Composers have tended to treat Latin texts in accordance with the prosodic patterns of their own language: for instance, French composers have usually paid little or no attention to Latin word stress in settings of liturgical texts; […]" (Randel (ed.) 1999: 533).

ending pitch. I have noticed that it has become a habit for me to glide these extended vowel segments by quarter tones up to a minor second. Though, this could be extended to a much larger interval and the semantic element would still remain undisturbed. In **Figure 28.1**, the glide upwards on the word *ved* is a naturalistic setting of the text. While the downward glide in **Figure 28.2** on the word *ble* is artificially created to give the line movement. Though, neither embellishment would majorly change the semantic integrity, since they are monosyllables with only one possible pitch accent or tone (i.e. sliding upwards).

There are many more linguistic phenomena the composer could develop and tinker with on a lexical basis than the ones I have exemplified. Furthermore, different tonal languages and their phonological implications offer different possibilities which may not be ubiquitous across languages. For example, the three Cantonese register tones (i.e. high, mid and low distinction) can affect the composer's consideration of tessitura but the two rising (or falling, depending on the dialect) tones of Norwegian may not. The scope of this subchapter is not to comprehensively highlight every instance I have considered these features of the languages. Rather, I hope that the cases discussed will spark thoughts on how to consider the relationship between speech tones of tonal languages and how they impact the composition of sung passages.

Section Summary:

- Phonological rules of each language guide the composer in maintaining semantic integrity of a text when it is set to music. Such rules vary greatly between spoken Cantonese, Norwegian and Mandarin, as each language has a different relationship to pitch.
- The six tones of Cantonese lend themselves to stricter rules: three register tones (i.e. low-mid-high) must be distinguished in relative pitches. Meanwhile, the two contour tones of Norwegian may more likely be altered without disturbing semantics.
- In the trilingual song cycle *natt-öö-夜*, I maintained the tone structures of Cantonese in the melody crafting, although the intervals were relative. It is also possible to maintain semantic integrity with melismatic passages as long as the core of the syllabic is first established. Checked tones (i.e. 'shorter' tones cut off by coda consonant segments) are also governed by a temporal aspect (i.e. shorter duration with an interrupted pitch contour or an unwavering pitch).

Section Summary (cont'd):

- *vihik (d)*, a choir work, showcases the harmonic verticality developed by consecutive Cantonese syllables of a text passage. The interval relativity once again allows the composer to freely pitch the low-mid-high distinctions but the vectors (i.e. going from one point to another) inform the shape of the chord.
- *messe norvégienne profane* exemplifies the possible melodic variations between spoken and sung Norwegian. As the correspondence of tones between speech and song is not as strict, the contrast of speech and singing could be used as a tool for making variations on the same text.
- Monosyllabic Norwegian words with no consonant codas tend to be tone 1 (i.e. contour tone, rising without a dip). The lack of indication for temporal interruption means that the pitch glide can be indefinitely extended.

## 4.2 Composing with Norwegian 'Pitch Accents' versus Cantonese 'Tones'

The previous subchapter has touched upon differences in composing with speech tones of Cantonese and the 'tones' (or more commonly known as pitch accents) in spoken Norwegian. This section further comments on the difference between the two phonetic features and dissects the different contexts of pitch content from speech used in music composition. I have discovered that, despite both languages deal with pitch as a phonetic feature, the approaches and results in music vary greatly and there are challenges in putting the two languages together in a single work. This subchapter will focus on generating musical materials through understanding these speech pitch features. It does not focus on the language-specific phonological rules in songs discussed in **Chapter 4.1**.

The hybridity of intonational and tonal pitch-use in Norwegian speech makes it an interesting language to create music. It is unlike Cantonese where the pitches of each individual syllable are immediately established and the perceiver are able to compare one syllabic segment to another. The rules that are outlined in the previous chapter make composing for Cantonese readings of Chinese texts a technical task with little variation on pitch contours (i.e. song-speech correspondence). Spoken Norwegian, on the contrary, gives space for bending these rules.

While individual Norwegian word segments bear intrinsic tones (e.g. tone 1 *sæter* [se:ter] versus tone 2 *seter* [se:ter] in Standard Eastern Norwegian), the pitch elements of speech segments above the phrase level (i.e. phrase, sentence and paragraph) shift in pitch (Hognestad 2012: 10). This shift of pitch stress at higher prosodic levels qualifies spoken Norwegian as an intonational language rather than a tone language. It has been argued that the feature of *pitch peak delay* within multisyllabic words make for a similar shift even at the syllabic level. "Shifts in the alignment, and the phonological association of tones bring about reassignments to function" (Bye 2004: 10). To paraphrase, depending on the context of where a word of a certain tone is placed in a sentence structure, the contours (i.e. the rises or the falls in pitch) may be altered. While these details are likely relinquished in the process of composition, the phenomenon can be important for compositional considerations in music theatre with Norwegian texts. These features have the potential to be understand as quantifiable elements of emotive nuances or a play with the speech-likeness within the passage's context. For example, in cases of quicker utterances in a recitative, a

composer could design the phrase such that the pitch-peak of a phrase or a word to match with the high point of a phrase. The opposite could also be done perhaps for a subversive meaning or an ironic effect in portraying an affect which comes through from being less speech-like. For example:

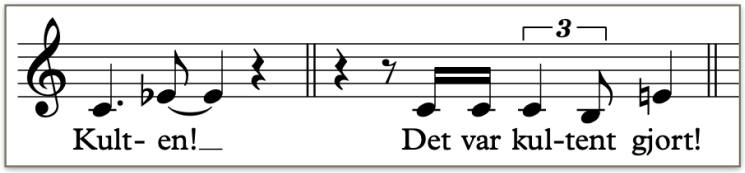



**Figure 29.0** Two short music fragments differentiating the two Norwegian tones or pitch accents; *'Kulten!'* (Tone 1 on the word segment, 'The Cult!') and *'Det var kultent gjort!'* (Tone 2 on the word segment *kultent*, 'That was nastily done!').

In the first fragment, where the singer only utters 'the cult!' (a noun), the pitch rise of tone 1 is reflected. In the case that this word is the focal point of a musical phrase, following this tone contour is enough to reflect the spoken inflections. In the second fragment, the word *kultent* is an adverb and the focal point of the melodic phrase. Let us assume that this word will bear the most dramatic weight in the passage and the composer aims to give this phrase some significance in a work. As the word bears tone 2 and to achieve speech-likeness, the pitch dip must occur at some point of the disyllabic segment. In linguist Patrik Bye's paper *Evolutionary typology and Scandinavian pitch accent*, it was postulated that "Accent 2 has a single peak *late* in the stressed syllable" (2004: 13), meaning that the dip in **Figure 29.0** should technically occur on *kul-* rather than the second unstressed syllable of *-tent* in the word *kultent* ('unpleasantly' or 'nastily'). However, how much of such precise detail can music notation possess comes into question: in theory, *kul-* should be notated as a glissando downwards and the lowest pitch moment should come right before the utterance of *-tent*. Similarly in Cantonese, a research has shown that there are 'sacrifices' to be made to the tone structures of a higher level segment when the tempo is quick.[43] Depending on the tempo, such detailed notation is impractical in the standard notation context; it is better to verbally explain the phenomenon to non-native speakers of the language. Temporal exactness from phonetic analyses in music notation must be ultimately sacrificed. Although, the approximate shape of *kultent*'s tone 2 is more or less maintained, particularly with the word *gjort* following on a higher pitch to highlight the pitch fall.

[43] "[I]n faster-paced songs, the tendency is for the tonal contours to be levelled out. This is accomplished by sacrificing the initial pitch rise when time is short" (Chan 1987: 140).

It can be said that working with Norwegian tones or pitch accents is quite different from working with Cantonese through the example above. In the latter language, the composer establishes the pitch register levels from the onset of a syllable and cope with the consequential phonemic elements without extensive consideration of temporal elements (e.g. duration of phonemes, etc.). Each Chinese character is an individual, monosyllabic word segment and considerations of prosodic flow and structures of larger phonetic segments have other considerations. Norwegian tones, meanwhile, shift from syllable to syllable dependent on the larger phonetic segments in discussion. Especially when considering different styles of texts (e.g. poetry versus prose versus dialogues), the composer is challenged to interpret or 'translate' the speech sounds to music at different phonetic levels (e.g. am I interested in focusing on the word, the phrase, or the sentence in a musical moment?).

The musical problem becomes: how do I read, imagine and hear a text passage in Norwegian? Depending on the context of different types of works (e.g. song versus theatre), which phonetic segment level should I focus on? Comparing the resulting works with Norwegian texts in my projects, my works tend focus on phonetic segments between the syllable and word level. This has presumably something to do with my interest in miniature forms in the style of György Kurtág (b. 1926). In earlier works of my research process such as *vihik (a)*, *natt-öö-夜* and *kj/ærlige ord*, I have focused on the two contour tones of Norwegian. As early works, they not always accurate in its speech-likeness, though the intonational elements of speech (i.e. "pitch accentuation, stress, focus and emphasis"[44]) have been intuitively felt throughout my compositional process.

[44] As defined by D. Robert Ladd in his book *Intonational Phonology* (2008).

| Works with Norwegian Texts | Sections with Focus on the Phonemic Level | Sections with Focus on the Syllabic Level | Sections with Focus on the Prosodic Foot Level | Sections with Focus on the Word Level | Sections with Focus on the Sentence Level and Above |
|---|---|---|---|---|---|
| *natt-öö-夜* | x | x | x | x | |
| *vihik (a)* | x | x | | | |
| *kj/ærlige ord* | | x | x | x | |
| *messe norvégienne profane* | | | x | x | |
| *vihik (c)* | | x | x | x | x |
| *vihik (d)* | | x | x | x | x |
| *minn(i)e* | x | x | x | x | x |
| *hvorfor pusen?* | x | x | x | x | x |

**Table 5.0** Table of artistic works highlighting the specific phonetic segments.

Unsurprisingly, the longer works later in the process — the chamber operas and my only choral piece in the project — have passages focusing on larger speech segments, as these works have a narrative function. They require more than just word or poetic means (i.e. focus on smaller sound segments). Why are smaller phonetic segments simpler to work with? Why is it challenging to consider phonological tones in larger contexts in music composition? To answer these questions, I have considered some possible musical settings of Cantonese and Norwegian texts in passages where both languages are present together in the same musical space:

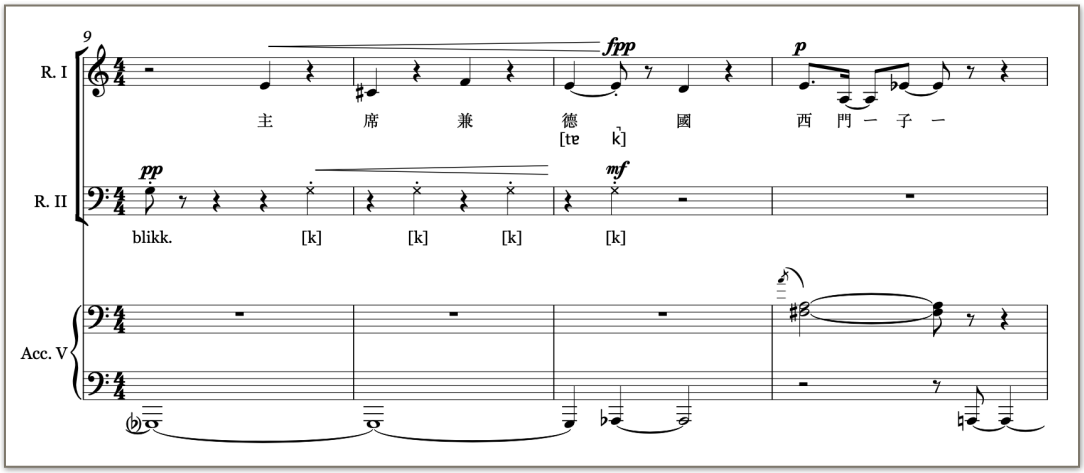



**Figure 30.0** Play on the Cantonese checked tones in *minn(i)e* (2020).

This duet (**Figure 30.0**) between the Norwegian reader (R. II, or reader II as indicated in the score) and the Cantonese reader (R. I, or reader I) serves mainly as passive storytellers of Minnie Vautrin's diary entries stylised by the Cantonese librettist, Chapman Chen. Reader I pushes forward the narrative through a song-recitation of longer prose passages and reader II is responsible an accompanying role with complementing commentaries in Norwegian. Although the drama's main focus is on the sentence level in the Cantonese text, there are moments in this aria where the Norwegian reader's phonemic segments coincide (i.e. are similar to) with the Cantonese reader's checked tones (i.e. tones that are 'cut short' by a consonant coda). In **Figure 30.0**, this occurs in bar 11 on the Cantonese syllable of 德 [tɐk̚˨] ('German'), where the preceding Norwegian word *blikk* ('moment') is already reduced to a repetitive [k] sound. The coda consonant of Reader I 'cancels' the almost cough/hacking-like [k][45] phoneme of Reader II. This is where the polyphony between the two languages resolve and checked tone of 德 [tɐk̚] becomes a musical comma in the languorous and rhythmically flat storytelling.

A few bars later, the tonal content of the two languages meet again in a phonemically similar syllable. While Reader I's context of storytelling through larger sentence structures, I notice that the Norwegian word *barn* ('children') coincides well with the Cantonese reading of 班 [bɐ:n˥], which is a high tone. The accordion accumulates intensity towards this moment and the rising pitch quality of *barn* would make for this downbeat a cadential structure in the music, since it metaphorically 'resolves' upwards to a higher pitch, which makes both musical and phonological sense:

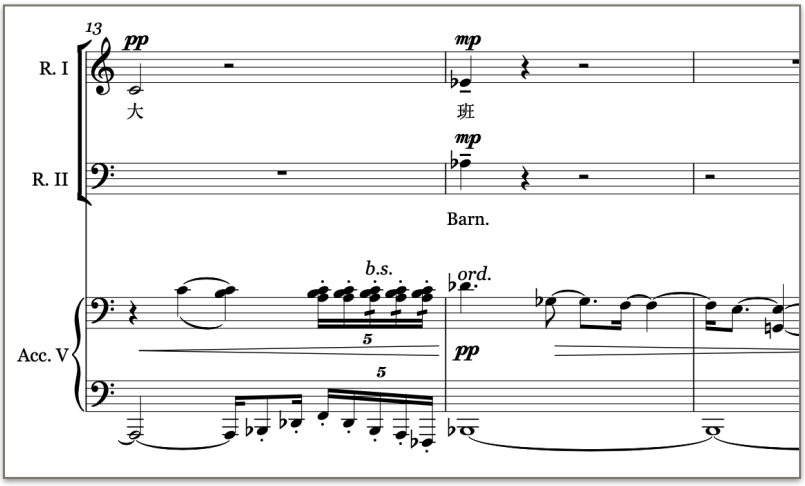



**Figure 31.0** Phonemically similar syllables between Cantonese and Norwegian in *minn(i)e* (2020).

[45] The [k] sound and cough metaphor is important to the opera, as it represents the asphyxiating main character, Minnie. This phonemic-motivic figure recurs often throughout the piece.

Another moment of phonemic similarity within the prosodic structure of the Cantonese narrative occurs at the end of the duet (**Figure 32.0**). At this point, the Norwegian singer, together with the onstage accordionist (Acc. V in the score), are uttering numbers. The accordionist hesitates and repeats *ti* ('ten'), which is naturally an upward glide in the Norwegian tone, until Reader I concludes with the line starting with the syllable 睇 [tɐi˧] ('to see') with a rising tone similar to the preceding Norwegian tone as well as the same initial consonant of [t]. The resulting effect is a unifying ending where both languages meet once more before the scene concludes and the lights dim on stage:

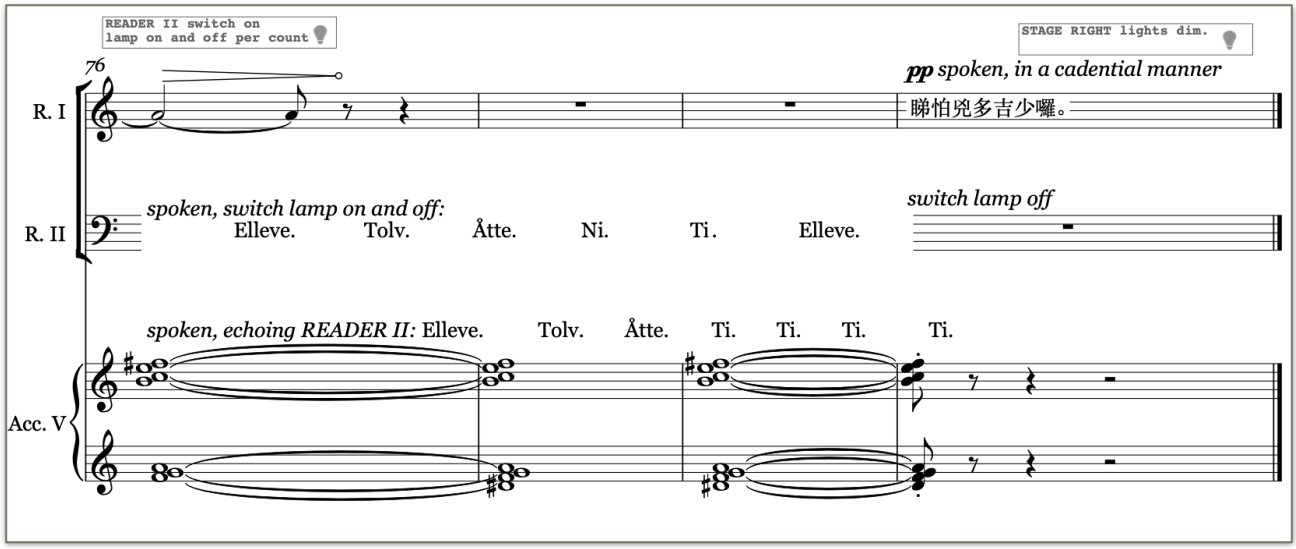



**Figure 32.0** Syllables with similar tone contours between Cantonese and Norwegian in *minn(i)e* (2020).

These moments have led me to understand that combining texts of two different tonal languages presents a couple of challenges: 1.) when the contour structures of tones in larger segments do not match well, it is difficult to unify the melodies without pitch-material interference and this is why the secondary voice (i.e. Reader II) must take on an accompanist role with focus on small segment levels in these passages, and 2.) similarities of tone materials between languages tend to occur on a smaller segmental level (i.e. phonemes, syllables and prosodic feet), which is why specific moments in the duet are highlighted as 'cadences' for where languages meet.

I have attempted a more complex blend of tones between the languages in the choral work *vihik (d)* (2019). Here in this musical moment, I have interlocked two Cantonese syllables with one in Norwegian because of the similar speech tone structures (the transliteration on the score are shown in Swedish orthography, as the piece is written for a Swedish choir):

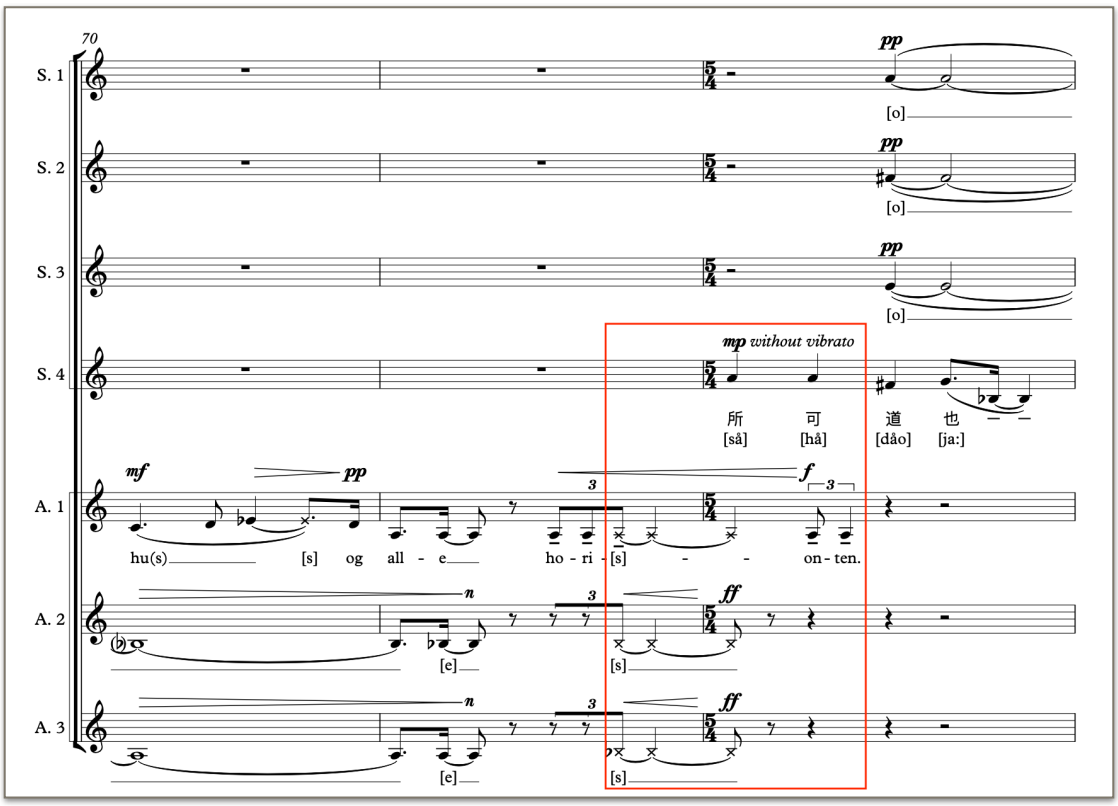



**Figure 33.0** Interlocking Cantonese syllables with a similar syllable of a Norwegian word in *vihik (d)* (2019).

In the word *horisonten* (‘the horizon’), I deliberately delay the utterance of *-son-* through hesitating on the sibilant [s] in the three lower voices. When the Cantonese syllable 所 [sɔ˧] (‘that which’) enters, the Norwegian *-son-* segment becomes fulfilled by the Cantonese phonemic subset on the same pitch (i.e. the A) before proceeding to the next syllabic segment. This creates an illusion that these two languages have merged momentarily. The stickler with this effect, though, is that most audience members are unlikely to speak both languages fluently enough to perceive this moment of play between the languages.

The differences in pitch contour between the two phonologically ‘tonal' languages (i.e. spoken Cantonese and Norwegian) have made the combination of texts in both languages a challenge. Most larger phonetic segments (e.g. words and sentences) in my works are parsed into smaller segments (e.g. syllables and phonemes), such that the pitch content of speech and song remain clear. Usually, Norwegian tones are more lenient due to the language's intonational tendencies in quicker speech of longer text passages.

Section Summary:

- The difference between the intonational leanings of spoken Norwegian versus the fully tone-based feature of spoken Cantonese makes for compositional challenges when the two languages are used in the same work.
- The *pitch peak delay* feature of most Norwegian dialects make for a speech-like representation in music notation a challenge. The glides of the two pitch accents or tones are pushed towards the end of a syllabic segment of multisyllabic words, which may not be practical to notate in music. Verbal instruction for non-native speakers is preferred from my own practice.
- Tones of Norwegian speech are dependent on which phonetic segment level the composer deals with in setting music to texts. In quicker speech passages, certain word or syllabic segments are likely emphasised over others. The remnant of the text will be intonation-based and the emphasis of stress on the temporal plane would suffice for the most part.
- Reflecting on most of my artistic output in working in the two languages, I have discovered a tendency to work with lower segment levels. In cases where higher segment levels are highlighted (i.e. storytelling through prose or dialogues), the second line with the other language tends to revert to a phonemic focus in complementing the first line.
- In my works, highlighting phonemically similar elements — pertaining to tone, especially — is often done through cadential or significant musical moments, as shown in *minn(i)e*.
- The composer must have a good understanding of the tonal pitch contours in the different languages for identifying and isolating possible moments in two texts where the blending of two languages can occur. This happens mostly on a syllabic or phonemic level in my works.

## 4.3 Pitch in Speech and Its Musical Representation

This subchapter discusses the pitch materials in speech and how the abstraction of these concepts are used in my composition. The two previous subchapters have addressed issues of how speech tones influence music making on a direct and technical level, such as notating pitch or tone contours reflected in speech and considering the semantic functions of maintaining certain phonemic features of a text.

The consideration of speech tones in compositions may be discussed on larger compositional scopes such as musical forms, dramaturgy and even the internal semiotics of a composition. Canadian linguist Murray Schellenberg opines in his research on singing in Cantonese and Mandarin that "[c]ultures choose to manifest tones to varying degrees and in varying ways in music" (2013: 44). This claim is noteworthy because the music repertoire discussed in his dissertation and papers has not addressed certain genres such as *Lieder*, contemporary art songs and 'Western opera' traditions (I undermine the concept here, as different genres of music theatre should technically also fall under this category). Granted, these discussions are niche subjects even among established composers, musicologists and linguists. Schellenberg's point is though a positive motivator for my creative work, as the experiments will open up undiscovered territories between spoken language and music. It must be said that, especially at this level of abstraction, scientific theories of phonetics are *interpreted* rather than executed or reflected in the compositional practice. In this interpretative lens, the 'artistic researcher disclaimer' must be made again such that linguists' and phoneticians' criticism of how misinterpretation of phonetics for the sake of generating composition tools may be forgiven. For this, I quote the philosopher Hans-Georg Gadamer in his response to his critics: "the power of science develops the power of a genuine delusion" (Gadamer 1990: 273).

I believe that this is where some of Yury Lotman's thoughts on poetics return: we are dealing with scientific systems through the means of art. "The relationship of system to text in a work of art is significantly more complex than in non-artistic sign systems" (Lotman 1976: 120). There is much to be said and speculated about what is done to speech-sounds and their metaphysical processes than simply proving or disproving the relevance of speech-sounds in music. Lotman further argues that "art *deviations* from the structural organisation can be as meaningful as the

[transmission of textual information from author to reader]" (Lotman 1976: 120), which is another point to be made when I begin to conceptualise musical ideas from linguistic ones.

I confess that the development of larger abstract musical concepts from tone languages in the research has come more frequently from Chinese texts than Norwegian texts. Although speech tone or pitch accents play an important role in the language, the intonational tendencies and features of spoken Norwegian make its use in the compositional process one more akin to most other intonation-based Germanic languages. At this point the languages are not discussed in semantic manner which gives the listener "'evidential' and 'regulative' information" (Laver 1994: 14), like in terms of discussing individual verses or words. The abstraction or re-imagination of phonetic information derived from speech tones become the musical material pertaining to form.

One of the most straightforward examples of how speech tones have influenced my form building in composition is harmonic syntax. I have mentioned in **Chapter 4.1** that the choral work *vihik (d)* (2019) relied on the tone contours of Cantonese for the construction of chords. It should be no surprise that, with such common practice concepts of chords and harmony involved in the work, there are formal consequences in creating musical materials this way. Contrast the pitch content of the three repetitions of the lyric "牆有茨" ('the wall has thorns'):

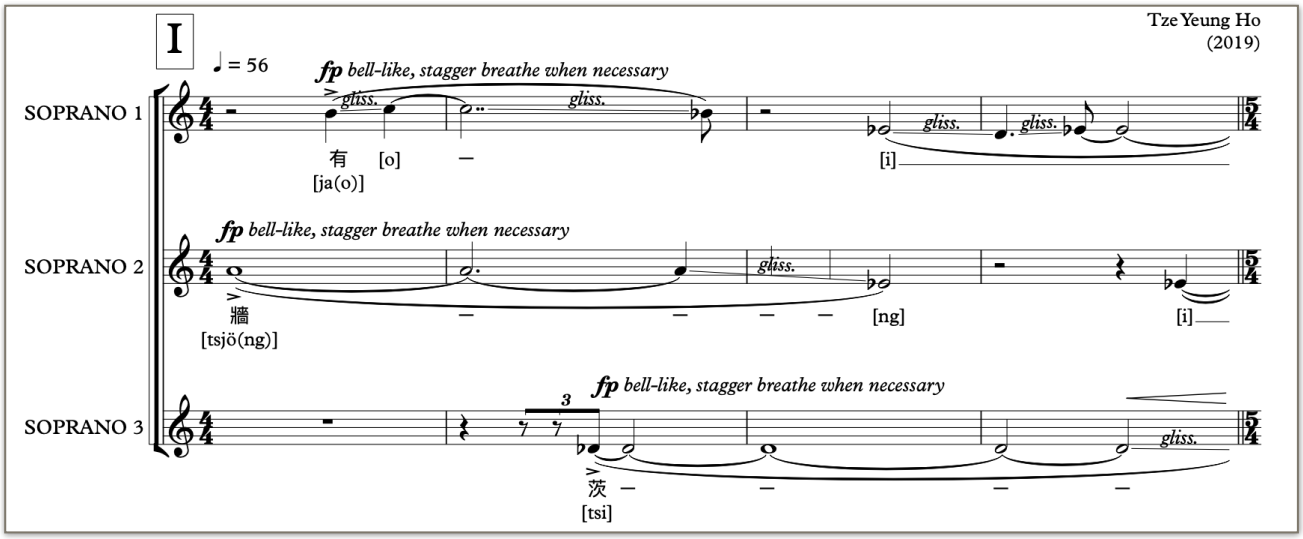



**Figure 34.1** The first occurrence of the "牆有茨" chord in *vihik (d)* (2019).

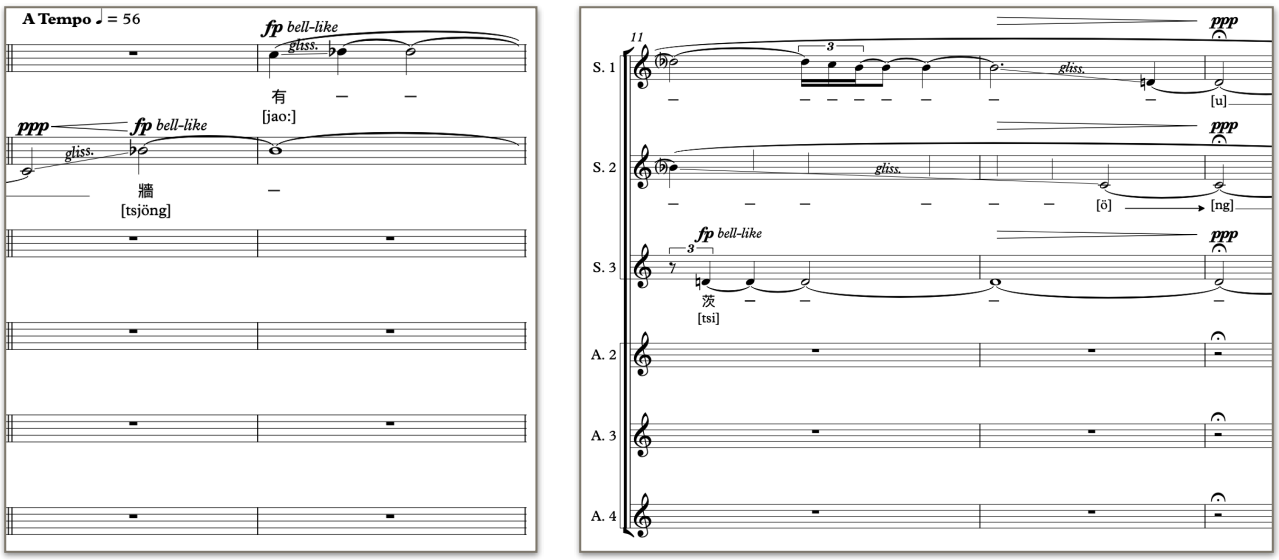



**Figure 34.2** A reoccurrence of the "牆有茨"chord approximately in the middle of *vihik (d)* (2019).

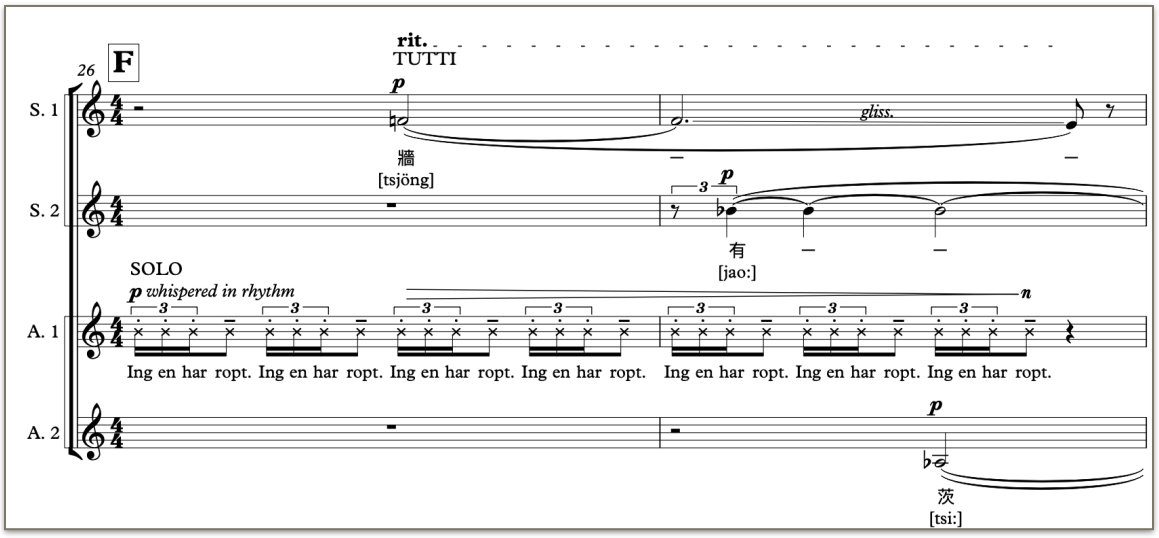



**Figure 34.3** The final occurrence of the "牆有茨" chord at the end of *vihik (d)* (2019).

In this choral work of approximately 9 minutes, the harmonic structure must be varied in one way or another to maintain musical freshness for the listener. The tones of the three syllables has inspired the first chord in **Figure 34.1**, and this was repeated several times in its original state in the first movement of two. In Yury Lotman's term on poetry, such a type of repetition was "to ascertain the internal structure of a work" (Lotman 1976: 120). The repeated material must further transform over the temporal duration of a work in order for a system (i.e. a musical form) to emerge. Hence, the glissando down to B-flat is sometimes altered to a B-natural for its later iterations. This wavering pattern between B-flat and B-natural led to the 'goal' of the beginning of

the second movement in **Figure 34.2**: a semitone above the original. This transposition will naturally maintain the speech tone integrity of the text. As noted in **Chapter 3.2**, this choral work has three tempo layers derived from the three languages. The Cantonese text, being monosyllabic, governs the slowest layer and the chord in discussion is the springboard providing a harmonic backdrop for the 'faster tempo' elements (i.e. the Norwegian and the Estonian texts). In the end of the work, the chord returns once more (see **Figure 34.3**), this time a tritone below the version heard at the start of movement two. The other singers whisper or speak the texts unvoiced. The harmonically distant iteration of the chord serves as an unsettling coda which echoes earlier musical elements. The chord vanishes as the voiceless elements of texts overwhelm the last resonance of the chord.

*vihik (d)* is a straightforward example in considering form, since the tones of Cantonese have already informed the harmonic structure of a work. The harmonic language is developed as such so that it is technically achievable by a semi-professional choir for which it is written. More sophisticated processes can be developed from speech tones in soloistic works, as the performers' abilities can be extended and integrated into more intricate composition processes. As large-scale processes require longer durations, it should not come as a surprise that much of the references to complex use of speech tone structures are found in my chamber operas *hvorfor pusen?* (2019) and *minn(i)e* (2020).

The Norwegian libretto by Linda Gabrielsen in *hvorfor pusen?* suggests much of the musical materials I have utilised to create the music for the work. Beyond the upward gliding quality of the two spoken Standard Eastern Norwegian tones seen in the melodic writing, the tone contour shape represented in other musical lines serves as a musical allegory or reference to the text throughout the piece. This upward slide shape is further amplified every line of the libretto as they are basically all questions. Taking into account of the intonational quality of the language, the rising pitch contour is a vital factor for both the storytelling and the music. One of the instrumental intermezzi is built entirely on the illusion of a rising figure, played in the strings and inside of the piano:

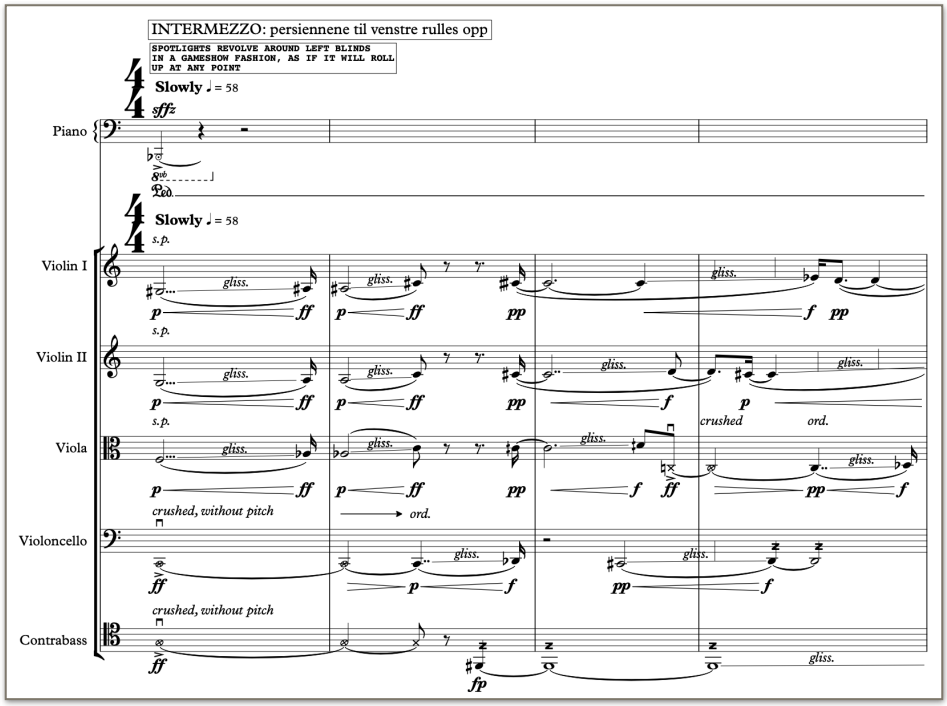



**Figure 35.0** An intermezzo from *hvorfor pusen?* (2019) inspired by the contour tones of Standard Eastern Norwegian.

Not only is the music of the intermezzi nodding to the tone structures of the spoken Norwegian language, the rising figure predominates the scenography of the work. This idea of the shape is also reflected in the imagined staging of the work with automated blinds that are requested in the technical rider of the score:

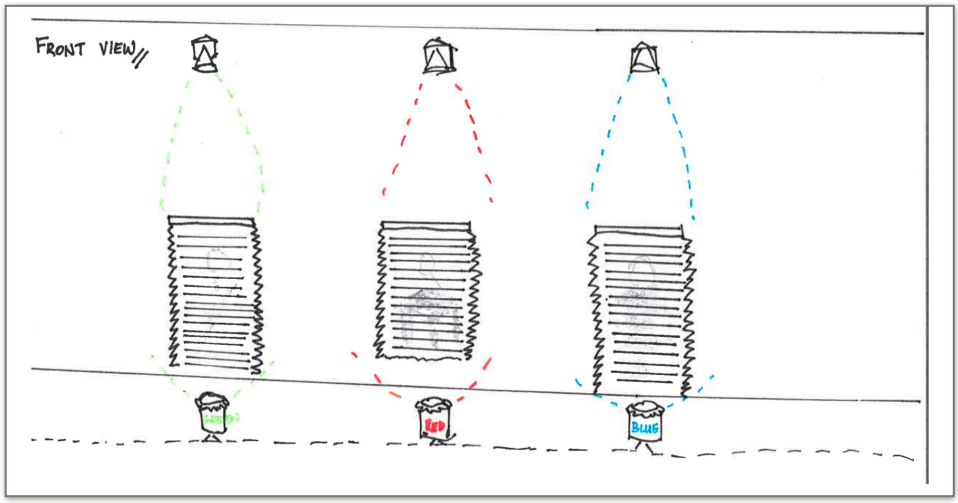
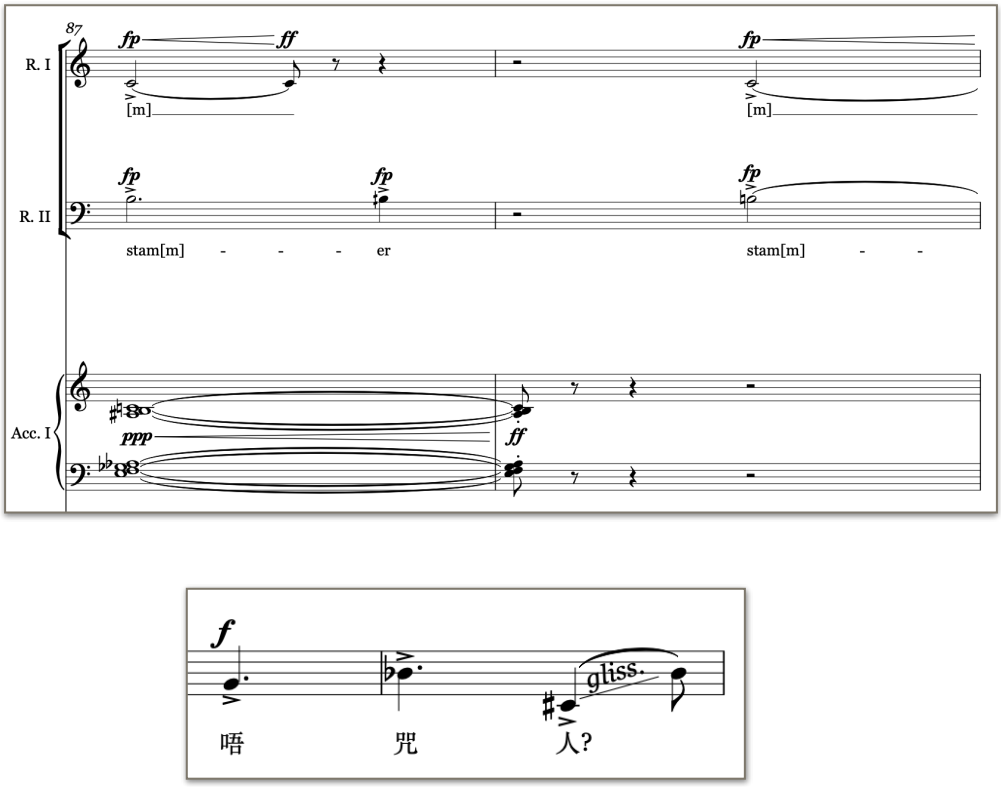



**Figure 36.0** Visual sketch of the stage set-up with automated blinds revealing the three characters in *hvorfor pusen?* (2019). The musical and linguistic rising gesture ubiquitous to the work has led to the choice of this scenographic consideration.

The raising of the blinds with its musical representation can be observed throughout the work until the very last moment. While the blinds lower after each of aria the individual characters, the intermezzi following each solo immediately propels the next sets of blinds to rise (like in **Figure 35.0,** where the string section's clunky upward glissandi portray the weight of raising the blinds). This creates an illusion of the constant revelation of a new character. Such a reflection on a similar significance or motif has been noted in poetry by Yury Lotman: "[t]he resemblance of sign and object constitutes the minimal act of repetition. Artistic texts (literary, pictorial and even musical) are built according to the iconic principle" (Lotman 1976: 114). This multidimensional derivation of physical and musical motifs from speech tones in Norwegian has been a fascinating object to develop for me as a composer. The concept may be repetitive and ubiquitous but the subversion of the message in the different aspects of a musical theatre is generative and does not restrict the perceiver from open associations.

In another moment similar to the one described in *hvorfor pusen?*, I recall the use of Cantonese tones in my second chamber opera of the project, *minn(i)e*:




**Figures 37.1 and 37.2** Two pitch contrasting passages in *minn(i)e* (2020) between the Cantonese and the Norwegian texts.

In the passages shown in **Figures 37.1** and **37.2** (treble clef) taken from different moments in the same section of the work, the two similar pitch motives (G - Bb), which are derived from coinciding speech tone elements, deviate into different materials over time. This is an important point to note, as the two characters singing these passages emerged from the same humming [m] phoneme as shown below:

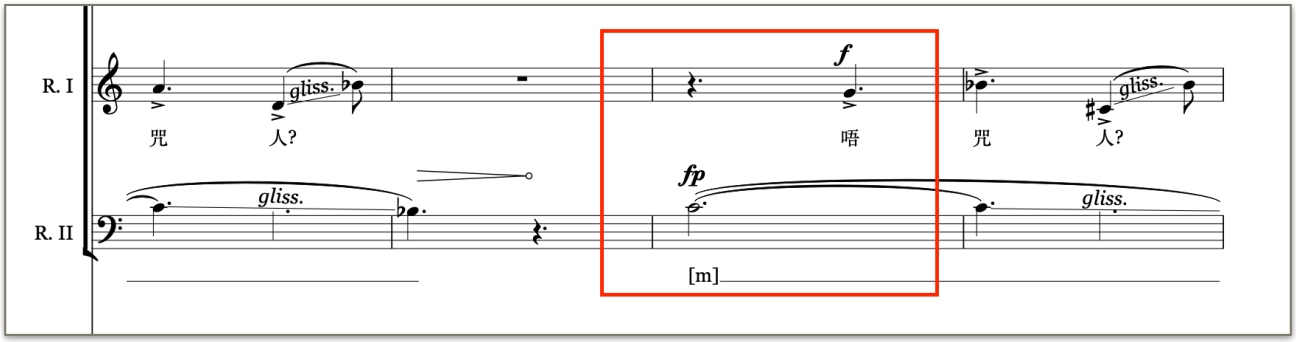



**Figure 38.0** A passage from *minn(i)e* showcasing the slow, phonemic deviation between the two Reader characters through the humming phoneme of [m]

This transformational process has major dramaturgical implications for the work, as the two nameless, generic characters are reading the same book (i.e. Minnie Vautrin's diary) and occupying the same side of the stage. In order to reflect the theme of distance and trauma, the characters' lack of identity creates schisms between many facets of the work: the languages, the occupied space, the portrayed time period and theparallel personal trajectories. The actors' unique characterisation are stemmed from the deviation of emotional responses to the diary which is highlighted throughout the work. While Reader I responds to the text as a reactive and agitated reader, Reader II is presented as reflected and collected. The pitch and phonemic content development of the two characters play therefore an important role in how these characters meet briefly and deviate in the same theatrical space. After Reader I breaks out into the tantrum-like outburst after the passages shown in **Figures 37.0** and **38.0**, the two readers on stage never return to fully singing together until the end of the performance, where a choral texture on 'amen' reminds the audience of the [m] phoneme passages in the work's opening. Lotman's idea of 'resemblance of objects and signs' mentioned earlier is a vital idea for binding phonetic sounds and musical form into a single work. One crude way to visualise the relationship is as following:

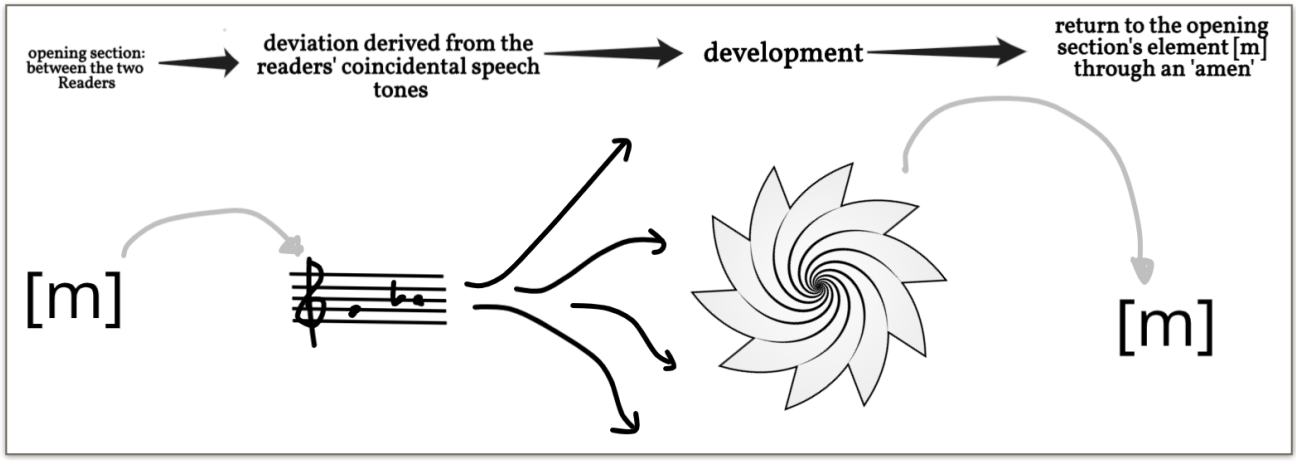



**Figure 39.0** An abstract map structuring the trajectory of phonetic elements of the two onstage readers which propel the musical form in *minn(i)e* (2020).

Working through my chamber operas, I have discovered that speech tones in Cantonese and Norwegian are not merely pitch-related compositional tools. The formal design of theatrical elements in larger works and in their extramusical elements are also taken into account from the tone structures. Yury Lotman has postulated that "the artistic text reacts with (sometimes completely accidentally) co-located texts and enters into semantic relations with them" (Lotman 1976: 123) and this statement holds true even in the relationship between text and music composition. Through a rising motive from the speech tones, *hvorfor pusen?* generated music on a rising motive and a theatrical contraption or prop (i.e. the automated blinds) used in its production. Lotman has also claimed "[t]his phenomenon may be seen in the composition assemblages ranging from an anthology, an almanac or an album, all seen as structural unities to the relationship of various pictures in an exhibition or buildings in an architectural ensemble" (Lotman 1976: 123). Developing large ideas from a series of small concepts — like the tone structures of speech — can therefore lead to larger collections of associations beyond the immediately perceivable range and scope of ideas. I have discovered such an 'anthology' through the humming [m] sound in *minn(i)e* is an implication of Minnie's name, suffocation, a grunt of frustration and the phonetic meeting point in 'amen'. This is particularly important for composing music theatre works, as the composer's imagination should not end with the music: it is about envisioning a world located between the text, musical materials and characterisation. After having produced two works within the chamber opera genre, select phonetic concepts are seen as potent sources which bring about fresh ideas across disciplines and practices of a composer and her/his collaborators.

Section Summary:

- Speech tones of Cantonese and Norwegian have influenced formal structures of my music, particularly in larger theatrical works.
- One must accept and appreciate the deviation from scientific understandings of speech tones and music in this instance of discussing phonetic concepts as allegories to music making.
- Yury Lotman's poetic theories serve as a bridge uniting the elements of poetic structures (i.e. use of texts) and sound phenomena, a territory from which I draw my inspiration when working with texts across different languages.
- Form in harmony derived from speech tones of *vihik (d)* provides a self-evident musical transformation from one section to another.
- The rising pitch accents or tones of Standard Eastern Norwegian serves as a fundamental symbol for the opera *hvorfor pusen?* Not only is the work musically influenced by the speech tones but the scenography and dramaturgy is similarly derived from the simple rising motive. The motive unravels a method in telling the libretto's story which originally contained only a series of questions.
- *minn(i)e*'s influence by speech tones of Cantonese and Norwegian is a complex relationship: two main characters on stage share moments of coinciding speech tone. This coincidence is where the deviation of the two different characters begin to emerge. The onset of this pitch motive is preceded by a humming [m] phoneme, which returns at the end of the work to create a sense of completion and unity within the musical form.

## 4.4 The Composer and the Performance of Songs Set to Texts of Tonal Languages

The final subtopic of this chapter will shift focus from theoretical compositional considerations of speech tones to the more pragmatic subject of performance preparation. I have noticed that many of my vocalist-collaborators have requested extra materials (e.g. recordings, IPA transliterations…) for learning the scores with tone language texts. This beckons the question: what are some effective tools which a composer could use to help performers achieve a satisfactory performance of works realising speech tones? This subchapter studies the experience from my collaborations with vocalists.

In the course of composing *natt-öö-夜*(2017), I have worked closely with the Canadian soprano Elisabeth Hetherington who later premiered the work. I frequently communicated with her about her process of learning the work and discussed the intonational and tone qualities of the text. These talks led me to making a questionnaire with questions pertaining to how she approached rehearsing works with texts of an unfamiliar language and particularly those with tone features:

> **Tze Yeung Ho:** Tell me a little about your previous experiences singing in a language which is not unfamiliar to you.
>
> **Elisabeth Hetherington:** For as long as I have been singing (I began singing in a professional children's choir when I was 8 and I'm 25 now) I have been exposed to different languages and sung in non-native tongues. At first, these were merely a collection of sounds without meaning. I would learn a German piece as "aacchh veee zoo baaalt"[46] and remember the melody with it's relationship to these seemingly random collection of sounds. It wasn't until I was much older (probably 15, or 16) that I started to look into the meaning of these foreign words and using the translations as a means to assist my memorisation.
>
> As a professional singer, I find myself most often singing in non-native languages and this process of translating the text is vital to not only understanding the music/motivations of the composer, but to developing my interpretation and later on memorising. I cannot commit a piece to memory if I do not have a word for word translation, as I fall back on the habits of to my 10 year old self, treating the language as a random collection of sounds.

[46] Author's addendum: Other languages should also be considered here.

> In the western classical tradition, singers are required to sing in the four major languages of the tradition (French, Italian, German and English)[47] as if they are the singer's mother tongue. Anything less than flawless diction is unacceptable. As such, I have gained an ear for mimicry and can easily imitate a native speaker rather convincingly. Two of the most useful tools in this regard are: a thorough understanding of the International Phonetic Alphabet, and access to an individual who can record the text for me to listen to."
>
> (Hetherington, email correspondence February 28, 2018)

Hetherington's answer offers a glimpse into the rehearsal methods of vocalists, as it reveals age-old practices (e.g. choir rehearsal techniques, 'knowing the text', rote memorisation, listening to recordings…) which the composers often take for granted. The notational factor, which is the primary agent of transmitting information to the performer, neglects the ontological difference from performers who are working towards physically understanding, interpreting and relaying the composer's ideas. Noticing that Hetherington focused on the idea of isolating patterns in languages — be it the narrative of a piece or the semiotic 'motivation of a passage' — the composer could assist or collaborate with the performer in identifying these patterns correlating to the vocalist's native language. This concept is not original and has often been a subject of discussion among linguists and phoneticians engaged in second language acquisition: "[c]omparisons [in such terms] of the phonetic and phonological patterns of the target accent and that of the learner allow the teacher to predict typical errors likely to be made by beginners" (Laver 1994: 80). In that, I would go as far as suggesting that it is the composer's responsibility to point out such speech patterns. Consider also the response of Hetherington below:

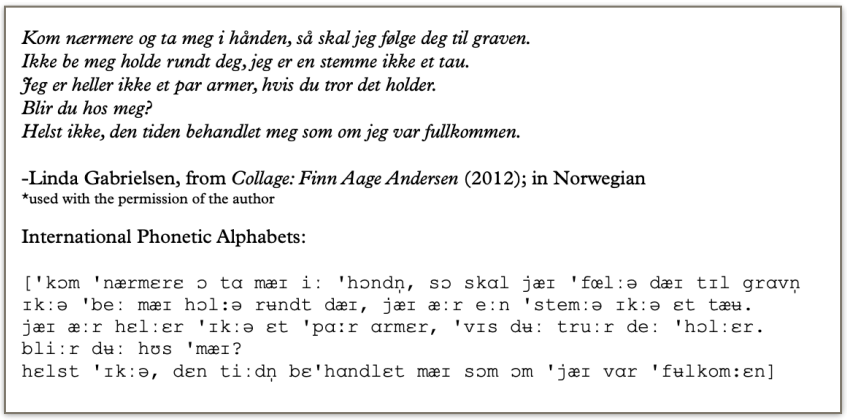
> **Tze Yeung Ho:** What are some challenges you foresee with working in a [completely] foreign language?
>
> **Elisabeth Hetherington:** If it is a language that I have no exposure to, the most obvious challenge will be trying to imitate a language of which I have no grounding or basis. It is incredibly difficult to recreate a pronunciation you have never heard. If the pronunciation of the language is not phonetic, that presents a whole new challenge, namely learning to associate sounds with phonetic combinations.
>
> **Tze Yeung Ho:** How do you think you will resolve the challenges? What are some techniques you have used when you sing in foreign languages?

[47] Here Hetherington omits other languages used historically in opera, as her affiliated institutions (i.e. the University of Toronto and Music Conservatory in Amsterdam) focused mainly on the four mentioned tongues.

> **Elisabeth Hetherington:** The best way to approach these challenges is by arming one's self with assistance. I would never deign to sing in a language I have not received ample coaching. Even when singing in French, a language I myself speak, I get coaching from native French speakers - their experience of the language is vital.
>
> I always begin by writing out the pronunciation and translation of the language. In this step, I will have a native speaker either coach me live, or make a recording for me. Then I practice saying the text as if in conversation. Then I start to say the text in the rhythm of the piece of music. Only when I feel completely comfortable do I put in the pitches.
>
> (Hetherington, email correspondence February 28, 2018)

The primary tool which I have found to be quintessential when working with a vocalist in such a situation is the voice recording. As Hetherington has concluded above, being able to hear the speech sounds uttered by a native speaker cannot be replaced with theoretical understanding of a language (e.g. IPA) in the learning process of a new work.Whenever I work with vocalists who will be performing in a language with which they are not comfortable, a recording of the lyrics/libretto read aloud will be made available. This has proven to be useful with Hetherington in *natt-öö-夜*, with the Netherlands-based vocal students of Finnish and Portuguese backgrounds in the premiere of *kj/ærlige ord*, with an Estonian vocalist in *vihik (c)* and with Chinese soloists in *hvorfor pusen?* the opera. I have further noted that vocalists requested the texts to be read in different tempi, which I suspect has something to do with listening to and familiarising with the different segmental levels of speech. For example, when I was asked by the Chinese soloists to record the reading of the Norwegian texts in *hvorfor pusen?*, the soprano vocalist asked me to provide an extremely slow version of the spoken text such that she can understand the intonation of individual verses, especially when it involves naturalistic speech in certain moments of the work. Meanwhile, the mezzo-soprano and baritone soloists with only sung parts were content with the quicker readings focusing on larger prosodic structures.

Beyond the recording, Hetherington pointed out the need for IPA (International Phonetic Alphabet) as part of her learning process, which has seeped into my compositional practice:



*Kom nærmere og ta meg i hånden, så skal jeg følge deg til graven.*
*Ikke be meg holde rundt deg, jeg er en stemme ikke et tau.*
*Jeg er heller ikke et par armer, hvis du tror det holder.*
*Blir du hos meg?*
*Helst ikke, den tiden behandlet meg som om jeg var fullkommen.*

-Linda Gabrielsen, from *Collage: Finn Aage Andersen* (2012); in Norwegian
*used with the permission of the author

International Phonetic Alphabets:

['kɔm 'nærmɛrɛ ɔ tɑ mæɪ i: 'hɔndn̩, sɔ skɑl jæɪ 'fœl:ə dæɪ tɪl grɑvn̩
ɪk:ə 'be: mæɪ hɔl:ə rʉndt dæɪ, jæɪ æ:r e:n 'stem:ə ɪk:ə ɛt tæʉ.
jæɪ æ:r hɛl:ɛr 'ɪk:ə ɛt 'pɑ:r ɑrmɐr, 'vɪs dʉ: tru:r de: 'hɔl:ɐr.
bli:r dʉ: hʊs 'mæɪ?
hɛlst 'ɪk:ə, den ti:dn̩ bɛ'hɑndlɛt mæɪ sɔm ɔm 'jæɪ vɑr 'fʉlkom:ɛn]

**Figure 40.0** IPA guide in the technical notes of *kj/ærlige ord* (2019).

The tool of IPA, however, is problematic. Debates on IPA typology are prevalent in the field of phonetics and phonology, as it is not particularly 'natural' to create representations of sounds without the semantic or semiotic contexts of a text.[48] Phonetic data which resulted in the studies of phonology vary from speech collected in the laboratory to recordings of speakers in controlled environments (e.g. news anchors) to normal speech. The context of song is often not the focus in the studies. On one hand, a composer setting music to a text is not a 'natural' occasion of speech. Though on the contrary, the composer may wish to exploit speech-likeness for a dramatic or musical purpose. While these are valid concerns with the use of IPA in a creative compositional process, it is not the main concern here in this subchapter. The question at hand is if IPA is at all helpful for the vocalist in performing the work. The conclusion through my personal collaborations is that it should be made available only if specifically requested. In the case of my collaboration with Hetherington, it has been requested as it is helpful for her to visualise the sounds in the recordings of the spoken texts. This beckons the question if the use of IPA is merely a visualisation of speech sounds for discovering patterns. Particularly in the context of tonal language, the typology of tones, as shown in the Chao system and in IPA, are relatively unintuitive upon first glance; both systems require that the reader have a prior understanding of the possible register or contour tones of a language. Although, teaching the two tones in Norwegian or the six tones in Cantonese to the performer are neither time consuming nor complicated tasks for a composer. It does require at least a *vis-a-vis* session between the two parties, which may not always be possible in professional situations like working with large ensembles with minimal rehearsal possibilities.

[48] "In the absence of naturalness conditions on phonological development the space in which languages can vary is in principle open, substantively speaking, and the typology of attested languages can't be taken as approximating the shape of this space" (Bye 2004: 2).

One final tool I have found useful is to provide some form of semantic or meaning ‘cues' for the performers. This can be as straightforward as a direct translation of a text to the performer’s native language. Though, when the drama or poetic features do not allow for that, an instruction of affective directions should be given in the music or as verbal explanations. In a direct translation of the text, the semantic features are not enough to represent the cultural cues or the subtexts of a passage of music, especially in different genres of music theatre. Linguist Colin Biggs opines that “[e]ven where there is agreement that linguistics has something of value to say about meaning, there have been those who have questioned the extent to which discussion of meaning has anything to do with the language faculty as such” (Biggs 1982: 108). This thought traces back to the notion that the vocalist’s performance is about embodying the work in its emotive and physical aspects through the notation theoretically conjured by a composer. While it is not possible for a score to deliver all nuances to the performer, all possible addenda should be considered in the rehearsal process, where time or resources permit so. The composer may consider highlighting the vital semantic and affective elements incorporated in the score in the role of a dramaturge, even if it may seem reductive to do so in verbal explanations. For example, after completing the score of *natt-öö-夜* and translating the texts to English, I have emphasised the compositional concept of phonemic likeness to Hetherington. This context was not found in the original set of poetry, since the curation of the three texts is my own artistic choice. In the process of working with Hetherington, I have to showcase how the form and semantic transformations of the *öö* > *yö* > *月* (read in Mandarin as [ɥø]) sequence emerged from the three different languages:

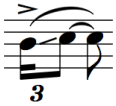
| Lang. | 🇪🇪 | 🇫🇮 | 🇨🇳 |
|---|---|---|---|
| **Word** | öö | yö | 月 |
| **IPA** *+Pinyin for Mandarin | [ø:] | [yø] | [ɥø] / ‘yúe’ |
| **English meaning** | Night | Night | Moon |
| **Musically notated** |  |  |  |

**Table 6.0** Table of the *öö* > *yö* > 月 sequence in the work *natt-öö-夜*.

While the English translation of the poems provided Hetherington with instructions on the mood and character of the piece, they are insufficient in demonstrating the musical dramaturgy built in the elements between the phonemes, the poems and the musical lines. As shown in **Table 6.0**, the three phonemes discussed are semantically as well as phonemically related. Furthermore, focusing on the ‘musically notated’ row, the reader may also observe that there is an emergence of a motive from a single note, where the contour tone of the Mandarin reading of 月 is partly an affective resolution to the static *öö*. All elements are symbiotic to the formal development of the work. Providing this chart or an explanation of where these significant moments occur for the performer proved to be useful. Hetherington’s mapping and interpretation of the score in her learning process has then focused on these phonetic feature inspired moments.

Although the three tools detailed in this subchapter may be obvious, my experiences have shown that composers are not proactive in providing the information for performers. This culture has led to some of these vocalist-collaborators opining that the composer should not feel obligated to provide anything beyond the preliminary score. After spending several years researching the subject and working with different artists, this is a habit I believe both performers and composers

should reconsider. The linear collaborative trajectory does not generate the dialogue between performers and composers which is much needed when musical materials are drawn from languages across cultures and practices.

Section Summary:

- Vocalists in my past collaborations have shown that they are cautious about working with texts in unfamiliar languages, with emphasis on tone languages for those who are trained in Eurocentric traditions.
- In the case speech tones are present in the works, three major resources a composer may provide a vocalist are: 1.) recordings of the spoken text by a native speaker; 2.) IPA transliteration or some form of orthographic or typological visual cues; and 3.) instructions on where meanings, phonetic features and/or musical elements intersect.
- Recordings should focus on different phonetic segments, paying attention to the speed which the text is read aloud.
- IPA transliterations may be insufficient, as phonological differences between languages may present misunderstandings. Verbal instructions and demonstrations are necessary to clarify all speech sounds.
- It is useful for the composer to engage with the performers for complex structures or ideas. While score instructions may seem to be sufficient, the interpretation of the performer may be better informed in the course of a dialogue.

## 5. Evaluating the Impact of Phonetics on Composers: Conclusive Thoughts

The seven subtopics of the two body chapters have covered the territories of the subject which I find to be the most urgent in the context of my own creative output. This dissertation has focused on the speech-likeness aspect within the use of spoken texts in music. Having taken Schoenberg's *Sprechstimme* technique as a springboard in formulating the two-part question "what is speech? And what is song?", I have analysed some possible interpretations of the concept of 'speech' from the perspective of composers, vocalists, phoneticians and linguists. The definition and its nuances vary from practice to practice. The grey area between the song-speech dichotomy emerges once the dialogues between the practices begin. The method of my work is mainly an interplay between phonetic theories and compositional practices, with underlying phenomenological approaches embedded within the historical and theoretical framework of linguistics. The central methods have included analysis of other composers exploring the relationship between music and spoken language, applying phonetic theories in compositional and artistic thinking and hands-on practice of composing and working with performers.

My research, taking the perspective of a composer rather than a phonetician, has explored a spectrum of possible composition techniques in a categorical approach of looking at the human voice apparatus. At least for the languages I have been exploring for my music (i.e. Scandinavian, Finno-Ugric and Chinese languages), I have determined that the parameters of quantity (or duration) and pitch are the core elements which differentiates the phonetic functions of speech and song.

I have chosen to omit methods of spectrographic analysis because my work focuses on the generation of compositional ideas: it is about phonetic concepts influencing compositional thoughts and not the *reproduction* of phonetic phenomena. For example, in the creation of musical phrase shapes taken from tone and quantity structures of a word, I am most interested in the overall structure. On the contrary, through spectrographic analysis of the same word, another composer could very well emulate these structures even more precisely (i.e. in exact durations, frequencies, etc…). However, the precision of such an act is not a part of my personal practice in composition; I find a more malleable consideration of the materials to be more interesting my creative process.

In the question of quantity, I have pointed towards the phonemic level of speech functioning as an approach for dealing with composing details such as articulation and phrasing. Not only applicable to vocal music, this method of working at the micro-segment levels is useful in considering instrumental music, as acoustic profiles of some phonemes can be imitated and emulated through instrumental and electronic means. The syllable and word levels of speech segments can be informative of larger musical forms, as the inherent durations of these segments may be augmented into larger musical units occupying longer durations. I have also discussed the phenomenon of quantity and tones within speech bearing strong correlations and thus affecting compositional decisions; this was particularly relevant in the contexts of composing music for Estonian and Finnish texts where phoneticians have largely proven strong pitch-quantity relationships within the speech of the languages.

Within the domain of pitch, I have outlined the effect of tones in tonal languages such as Cantonese and Mandarin on musical setting of texts. Previous researchers on the subject, particularly of Murray Schellenberg's doctoral dissertation, have shown that the restrictive traits of such languages do indeed create hindrances for the composer's creative choice making. The assumptions that there are strict pitch rules to music making do not hold true in composers of different genres (particularly in cases of experimental and contemporary art music). The understanding the phonetic nature of Chinese tones within the context of a text or a speech fragment prevails this assumption. This common misconception of strictness has led to a brief investigation of how performers perceive tonal languages as something 'difficult' to master in the performance of a work. From this perspective, I have been able to develop tools and rehearsal methods for collaborative efforts, with particularly successful results on the matter of using recordings emphasising different phonetic elements (e.g. slower recordings of spoken passages emphasising stress, and regular speech-space recordings for emphasising melodic contours). Finally, perhaps the most open-ended discussion within the realm of pitch, I have defined and compared the pitch accent of Scandinavian languages with their Chinese counterparts and some of their effects on the compositional process. The result of this pursuit has shed light on the composer's semantic play through concepts of phonetic likeness and syllable transformation processes and through deemphasising the absolute need to satisfy the phonetically 'correct' tones of a language.

I believe that this research is not an exhaustive representation of what more can be discussed within the framework of phonetics and composition. Shall a composer be interested in a different

set of languages, I am certain that the phonetic subjects of focus could differ greatly, even though the ontological scope (i.e. within the terminologies and knowledge of phonetics) would remain the same. Although I have not touched upon these topics, I suspect that an investigation in composing music for texts in some large language families (e.g. Afro-Asiatic or Turkic languages) and their subgroups (e.g. Romance or Slavic languages within the Indo-European language family) could offer composers an array of stylistic insights in comparing habits of composers from different linguistic backgrounds. Such explorations could be useful for both the composer's personal understanding of different creators' craft and for deconstructing any culture-specific or monoculture-centric thoughts on music composition. A reading through phonetics can serve as an alternative context in which one can discuss similar parameters in music. At the time this conclusion is being written in 2020, discussions of decolonisation of hegemonic musical genres have become heated topics of debate in the new music world, with particular acknowledgment to concepts noted in lectures and texts by scholars Edward Said, Homi Bhabha[49] and George Lewis. I believe that the discussion of language is among one of the many means to unpack different historic views on composition as an art form. Meanwhile, this dissertation embodies a personal dimension of my work as a composer, the set of Finno-Ugric, Scandinavian and Chinese languages is put into focus and it does not suggest that these tongues are in any way more noteworthy than the ones which are not mentioned in this dissertation.

I believe that further research on this topic from both composers' and phoneticians' perspectives should be continued in the future. Not only through another set of languages or phonetic parameters, the compositional aspect can be zoomed in from varying perspectives. For example, after having worked on my composition *isjoriske forsteder*, I became more knowledgeable in lament and runic singing of Finno-Ugric cultures and of works by poets such as Larin Paraske, Juhan Liiv and Lydia Koidula. The processes I have discovered in the course of this research could have been expanded to a comprehensive discussion of setting music to Finnic texts. They have led me to speculate that one alternative approach for this research's premise can be taken from the perspective of a composer working with specific writers or texts in a certain language. From my focus on studying and creating several works by Estonian poet Juhan Liiv, I realise that stylistic idiosyncrasies of an author may impact how a composer perceives a language, given that the historical and cultural context would highlight specific features of a language.

---

[49] Namely the books *Orientalism* (Said 1978), *Edward Said Continuing the Conversation* (Bhabha 2005) and many of Bhabha's articles on topics of the Other, and cosmopolitanism.

On the topic of language-specificity, I have noticed a lack of literature on music setting of texts from a Scandinavian tradition, as most of my sources in this dissertation on Norwegian have been taken from phonetic studies, as sources specifically on phonetic tones and music were scarce. Despite Norwegian composers' involvement within different song traditions tracing back to the nineteenth century with composers and writers of the time, it seems that musicologists, linguists and composers have yet to engage the subject in a comprehensive way as Chinese-speaking scholars have been doing in recent years.[50] This is likely due to the fact that *tonelag* in Norwegian dialects vary greatly across Norway while studies on tones and pitch in Mandarin and Cantonese are standardised in epicentres of the Chinese-speaking sphere.

From the phonetic discussion, further investigations on the phonetic parameter of stress could be fruitful for composers and vocalists. Unlike pitch and quantity as discussed throughout in this dissertation, the concept of stress does not share a direct or linear equivalent concept within music. While one may argue that phonetic stress is correlated to musical stress, the different perceptions of 'time' in speech and music make stress a difficult concept to fully reconcile between the two ontological frameworks. This is particularly contentious in different genres of music where speech-likeness and its criteria for such an evaluation varies (e.g. rap music's emphasis on speech is unlike that of 19th century German *Lieder*). It is important to note that the topics between phonetics and composition are where creative possibilities lie. While it might be interesting to look at composition from a scientifically phonetic perspective, my work has depended on looking at phonetics from a compositional perspective.

I hope that this interdisciplinary work has outlined a possible framework for composers, writers working with composers and phoneticians alike to seek out a language to communicate commonalities across the different fields. One particularly interesting byproduct of this research is a gravitation towards the issues of semantics. The study of sounds in both speech and music does indeed beckon the question: what does a sound imply and why do we listen? This question is unfortunately not what I had set out to answer, although — as a creator — the urge to do so was strong in every creative moment. While I had the ambition to summarise my creative processes

[50] Here I am noting the abundance of articles by the infrequently discussed authors in Chinese-speaking spheres such as Bin Li, Ling Zhang, Juanjuan Liu, Marjorie Chan and others within my own bibliographic references.

within a strict theoretical framework, my process was by no means the maximally effective research method, as each creative project would eventually steer its own course towards different end goals and results. I believe that artistic researchers are not obliged to feel that the result must find its home in the initial hypothesis proposed at the start of the research process, particularly in music and language where both fields are constantly evolving. The interdisciplinary approach will sometimes bring the reader to meander in a field of unrelated ideas in what seems to be a void of factoids and minutiae. Nonetheless, these seemingly irrelevant pieces of knowledge and serendipitous findings between disciplines are worthwhile for the composer to examine, as they have enriched my compositional toolbox.

## Sources


Interview

Hetherington, Elisabeth  February 28, 2018

Scores

*Arnold Schoenberg. Pierrot lunaire* 1914. Vienna Universal Edition.

*Luciano Berio. Sequenza III* 1965. Vienna: Universal Edition.

*Tze Yeung Ho. hvorfor pusen?* 2019. Oslo: NB Noter.

*Tze Yeung Ho. kj/ærlige ord* 2018/9. Oslo: NB Noter.

*Tze Yeung Ho. messe norvégienne profane* 2018. Oslo: NB Noter.

*Tze Yeung Ho. minn(i)e* 2020. Private collection.

*Tze Yeung Ho. natt-öö-夜* 2018. Oslo: NB Noter.

*Tze Yeung Ho. vihik (a)* 2018. Oslo: NB Noter.

*Tze Yeung Ho. vihik (c)* 2019. Oslo: NB Noter.

*Tze Yeung Ho. vihik (d) 2019.* Private collection.

*Veljo Tormis. Neli kildu* 1955. Helsinki: Fennica Gehrman.

## Bibliography


Adorno, Theodor and Livingstone, Rodney (trans.) 1998. *Quasi Una Fantasia: Essays on Modern Music*. London: Verso Books.

Asu, Eva Liina 2003. *The Phonetics and Phonology of Estonian Intonation*. Cambridge: University of Cambridge.

Asu, Eva Liina and Nolan, F. 2006. Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. — *Proceedings of Speech Prosody 2006*. Dresden: ISCA Speech.

Bang, Anne Louise; Krogh, Peter Gall; Ludvigsen, Martin and Markussen, Thomas 2012. The Role of Hypothesis in Constructive Design Research. — *Proceedings of The Art of Research IV, Helsinki 2012*. Helsinki: Aalto University School of Arts, Design and Architecture.

Biggs, Colin 1982. In a word, meaning. — *Linguistic Controversies,* pp. 108 - 121. London: Edward Arnold Publishers.

Bye, Patrik 2004. *Evolutionary typology and Scandinavian pitch accent*. Dordrecht: Kluwer Academic Publishers.

Byron, Avior 2006. The Test Pressings of Schoenberg Conducting *Pierrot lunaire: Sprechstimme* Reconsidered. — *Music Theory Online.* http://www.mtosmt.org/issues/mto.06.12.1/mto.06.12.1.byron_frames.html, accessed on April 27th, 2018.

Cahill, Lynne 2019. *Discovering Phonetics and Phonology.* London: Red Globe Press.

Carlson, Marvin A. 2006. *Speaking in Tongues: Language at Play in the Theatre.* Ann Arbor: University of Michigan Press.

Carter, Chandler 2010. *The Rake's Progress* and Stravinsky's Return: The Composer's Evolving Approach to Setting Text. *Journal of the American Musicological Society,* Vol. 63, No. 3; pp. 553 - 640. Berkeley: University of California Press, Journals and Digital Publishing.

Chan, Marjorie, K. M. 1987. Tone and melody interaction in Cantonese and Mandarin songs. — *UCLA Working Papers in Phonetics* Vol. 68, pp. 132-169. Los Angeles: University of California, Los Angeles.

Crystal, David 1991. *A Dictionary of Linguistics and Phonetics.* Oxford: Blackwell Publishers.

Crystal, David (ed.) 1982. *Linguistic Controversies.* London: Edward Arnold Publishers.

Daitz, Mimi S. 2004. *Ancient Song Recovered: The Life and Music of Veljo Tormis.* Hillsdale: Pendragon Press.

Everett, Yayoi Uno 2015. *Reconfiguring Myth in Contemporary Opera.* Indianapolis: Indiana University Press.

Gadamer, Hans-Georg 1990. The Universality of the Hermeneutical Problem. — *The Hermeneutic Tradition.* Ed. Gayle L. Orniston and Alan D. Schrift. Albany: State University of New York Press, pp. 147 - 158.

Gadamer's Aesthetics 2007, revised 2016. *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/gadamer-aesthetics/#Play, accessed April 27, 2018.

Halle, Morris 2002. *From Memory to Speech and Back: Papers on Phonetics and Phonology 1954 - 2002.* Berlin: Mouton de Gruyter.

Hannula, Mika; Suoranta, Juha and Vadén, Tere 2005. *Artistic Research — theories, methods and practices.* Espoo: Cosmoprint.

Harbeck, Rebecca 2016. The State of Cantonese Language in Hong Kong. — *US - China Today.* June 24, http://www.uschina.usc.edu/w_usct/showarticle.aspx?articleID=20282, accessed April 21, 2018.

Haugen, E. 1963. Pitch Accent and Tonemic Juncture in Scandinavian. *Monatshefte, 55*(4), 157-161, www.jstor.org/stable/30156267, accessed July 24, 2020.

Heller, Eric J. 2013. *Why You Hear What you Hear: An Experiential Approach to Sound, Music and Psychoacoustics.* Woodstock; Princeton: Princeton University Press.

Hognestad, Jan K. 2012. *Tonelagsvariasjon i norsk: Synkrone og diakrone aspekter, med særlig fokus på vestnorsk.* Kristiansand: Unversitetet i Agder.

Jakobson, Roman; Wauge, Linda R. (ed.) and Monville-Burston, Monique (ed.) 1990. *On Language.* Cambridge: Harvard University Press.

Johnston, David (ed.) 1996. *Stages of Translation.* London: Oberon Publishers Ltd.

Johnson, Keith 2003. *Acoustic and Auditory Phonetics.* Cornwall; Hong Kong: MPG Books.

Kramer, Lawrence 2012. *Expression and Truth: On the music of knowledge.* Berkeley: University of California Press.

Kristoffersen, Gjert 2015. *Innføring i norsk fonologi.* Bergen: Universitetet i Bergen.

Krull, Diana 1999. Foot isochrony in Estonian. — *Proceedings of XIVth International Congress of Phonetic Sciences*, pp. 1063-1066. San Francisco: ICPhS Archive.

Laanest, Arvo 1975. *Sissejuhatus läänemeresoome keeltesse.* Tallinn: Eesti NSV Teaduste Akadeemia.

Ladefoged, Peter 1973. *Preliminaries to Linguistic Phonetics.* Chicago: The University of Chicago Press.

Lau, Elaine 2010. Tone-Melody Relationship in Cantonese. — *Working Papers in Linguistics.* Vol. 41(3); pp. 1 -12. Honolulu: University of Hawai'i at Manoa.

Ladd, D. Robert 2008. *Intonational Phonology.* Cambridge: Cambridge University Press.

Laver, John 1994. *Principles of Phonetics.* Cambridge: Cambridge University Press.

Lerdahl, Fred and Jackendoff, Ray 1985. *A Generative Theory of Tonal Music.* Boston: MIT Press.

Lehtonen, Mikko 2000. *The Cultural Analysis of Texts.* London; Thousand Oaks; New Delhi: SAGE Publications.

Li, Bin and Choi Chung-nin 2016. Singing Tones in Cantonese Opera and Pop Songs. — *Speech Prosody,* pp. 322-325. Boston: ISCA Speech.

Liiv, Juhan 1989. *Sinuga ja Sinuta.* Tallinn: Eesti Raamat.

Lippus, Pärtel 2011. *The Acoustic Features and Perception of the Estonian Quantity System.* Tartu: Tartu University Press.

Lippus, Pärtel and Ross, Jaan 2014. Temporal Variation in Singing as Interplay between Speech and Music in Estonian Songs. — *Expressiveness in Music Performance: Empirical approaches across styles and cultures.* Oxford: Oxford University Press.

Lippus, Urve and Tormis, Veljo 2008. *Lauldud sõna.* Tallinn: Folger Art Ltd.

Liu, Juanjuan 2003. 羅忠鎔藝術歌曲中人聲旋律的處理方法 — *Proceedings of the Symposium at the 2003 Chinese Composers' Festival,* pp. 156-169. Hong Kong: Hong Kong Composers' Guild.

Lotman, Yury 1976. *Analysis of the Poetic Text.* Ann Arbor: Ardis.

Mateo, Marta 2014. Multilingualism in opera production, reception and translation. —*Linguistica Antverpiensia, New Series. Themes in Translation Studies, 13,* pp. 326-354. Oviedo: University of Oviedo.

Ormiston, Gayle L. (ed.) and Schrift, Alan D. (ed.) 1990. *The Hermeneutic Tradition.* Albany: State University of New York Press.

Patel, Aniruddh 2007. *Music, Language, and the Brain*. Oxford: Oxford University Press.

Pavlenko, Aneta 2005. Bilingualism and Thought. — *Handbook of Bilingualism: Psycholinguistic Approaches,* pp. 433 - 453. Oxford: Oxford University Press.

Pedersen Gagiu, Elena 2015. Symbolism, the beginning of modern poetry. — *The 6th International Conference Edu World 2014 "Education Facing Contemporary World Issues,"* pp. 593 - 599. Pitesti: University of Pitesti.

Pickering, Michael (ed.) 2008. *Research Methods for Cultural Studies.* Edinburgh: Edinburgh University Press.

Randel, Don Michael (ed.) 1999. *The Harvard Concise Dictionary of Music and Musicians*. Cambridge: The Belknap Press of Harvard University Press.

Roach, Peter 1982. On the distinction between 'stressed-timed' and 'syllable-timed' languages. — *Linguistic Controversies.* London: Edward Arnold Publishers, pp. 73 - 80.

Roederer, Juan G. 2008. *The Physics and Psychophysics of Music.* New York: Springer Science+Business Media.

Raju, Marju and Asu, Eva Liina and Ross, Jaan 2010. Comparison of rhythm in musical scores and performances as measured with the Pairwise Variability Index. — *Musicæ Scientiæ.* Vol. XIV, number 1; pp. 51 - 71. Tallinn: Sage Publications.

Ross, Jaan and Lehiste, Ilse 2001. *The Temporal Structure of Estonian Runic Songs.* Berlin: New York: Mouton de Gruyter.

Rubin, Hertbert J. and Rubin, Irene S. 2012. *Qualitative Interview: The Art of Hearing Data.* London: Sage Publications.

Rupprecht, Phillip (ed.) 2013. *Rethinking Britten.* Oxford: Oxford University Press.

Sagart, Laurent 1999. The origin of Chinese tones. — *Proceedings of the Symposium/Cross-Linguistic Studies of Tonal Phenomena/Tonogenesis, Typology and Related Topics* pp. 91 - 104. Tokyo: Tokyo University of Foreign Studies.

Sammallahti, Pekka 1998. *The Saami Languages: An Introduction*. Kárášjohka: Davvi Girji.

de Saussure, Ferdinand 1959. *Course in General Linguistics.* Ed. Taylor, Mark C. New York: McGraw-Hill Book Company.

Schellenberg, Murray H. 2013. *The Realization of Tone in Singing in Cantonese and Mandarin.* Vancouver: The University of British Columbia.

Smith, David Woodruff and Zalta, Edward N. (ed.) "Phenomenology", *The Stanford Encyclopedia of Philosophy* 2018. https://plato.stanford.edu/archives/sum2018/entries/phenomenology/, accessed April 20, 2020.

Stein, Deborah and Spillman, Robert 2010. *Poetry into Song: Performance and Analysis of Lieder.* New York: University of Oxford Press.

Suomi, Kari; Meister, Einar; Ylitalo, Riikka and Meister, Lya 2012. Durational patterns in Northern Estonian and Northern Finnish. — *Journal of Phonetics* Vol. 41; pp. 1 - 16. www.elsevier.com/locate/phonetics, accessed 14 July, 2020.

Suomi, Kari; Toivanen, Juhani and Ylitalo, Riika 2008. *Finnish Sound Structure: Phonetics, phonology, phonotactics and prosody.* Oulu: University of Oulu.

Tam, Gina Anne 2020. *Dialect and Nationalism in China, 1860 - 1960.* Cambridge: Cambridge University Press.

Taylor, Mark C. (ed.) 1986. *Deconstruction in Context.* Chicago: The University of Chicago Press.

Toussaint, Godfried T. 2012. *The Pairwise Variability Index as a Tool in Musical Rhythm Analysis.* Abu Dhabi: Faculty of Sciences, New York University Abu Dhabi.

Tsur, Reuven 2008. *Toward a Theory of Cognitive Poetics.* Sussex: Sussex Academic Press.

Tunbridge, Laura 2010. *The Song Cycle.* Cambridge: Cambridge University Press.

Vachek, Josef 2003. *Dictionary of the Prague School of Linguistics.* Amsterdam: John Benjamins Publishing Company.

Vainio, Martti; Aalto, D.; Järvikivi J. and Suni, Antti 2006. Quantity and tone in Finnish lexically stressed syllables. — *Proceedings of the 2nd International Symposium on tonal aspects of Languages,* pp. 121-124. La Rochelle: Université de la Rochelle.

Vurma, Allan 2007. *Voice Quality and Pitch in Singing: Some Aspects of Perception and Production.* Tallinn: Estonian Academy of Music and Theatre.

Wong, Ying Wai 2006. Contextual Tonal Variations and Pitch Targets in Cantonese. — *Proceedings of Speech Prosody 2006.* Dresden: ISCA Speech.

Xu, Xiaoying; Tao, Jianhua; Zhang, Ling and Lu, Ying Chao 2010. The Duration Analaysis of Checked Tones in Cantonese Speech. — *Proceedings of the 7th International Symposium on Chinese Spoken Language Processing,* pp. 459 - 464. Tainan: National Cheng Kung University.

Yung, Bell 1989. *Cantonese Opera - Performance as creative process.* Cambridge: Cambridge University Press.

Zhang, Ling 2015. Tunes of different speech styles in Cantonese. — *Chinese Language and Discourse 6:2,* pp. 183-204. John Benjamins Publishing Company.

## List of Doctoral Concerts

1. *natt-öö-夜*

Date: 31.05.2018

Venue: *Ajaloomuuseum (Suurgild),* Tallinn, Estonia

Performers: Elisabeth Hetherington (soprano), Matti Pulkki (accordion), Liis Viira (harp)

Program: Ho (WP), Dowland, Britten, Moulinié

2. *isjoriske forsteder*

Date: 02.05.2019

Venue: *Rahvusraamatukogu suur konverentsisaal,* Tallinn, Estonia

Performers: Elisabeth Hetherington (soprano), Hans-Gunter Lock (sound diffusion)

Program: Ho (WP), Gubaidulina

3. *tvil**

*messe norvégienne profane*

Date: 09.06.2019

Venue: *Jakob kirke*, Oslo, Norway

Performers: Sean Bell (countertenor), Liis Viira (harp) and Nynorsk messingkvintett (brass quintet)

Program: Ho (WP), Bach, Viira (WP), Britten

*hvorfor pusen?*

Date: 17.09.2019

Venue: Shanghai Conservatory of Music, Shanghai, China (as a part of Shanghai New Music Week Festival's Composition Competition)

Artists: Chengjie Zhang (conductor), Shiru Wang (stage director), Yufei Xiong (soprano), Junxi Mao (mezzo-soprano), Rongjun Xu (Baritone), Shanghai Conservatory New Music Ensemble

Program: Ho (WP), Norman (WP), Gao (WP)

*Submitted as a recording in two parts.

4. *minn(i)e*

Date: 24.04.2021

Venue: Sakala3 Theatre, Tallinn, Estonia

Artists: Lodewijk van der Ree (conductor), Anselmi Hirvonen (stage director), Magnus Pind (scenographer), Annely Leinberg (mezzo-soprano), Aule Urb (mezzo-soprano), Kristel Marand (alto), Ka Bo Chan (countertenor), Eilert Egil Taugbøl Hasseldal (baritone), Theodore Parker (electric guitar), Momir Novakovic, Henri Zibo, Mikk Langeproon, Live Berger Brekke, Kristina Farstad Bjørdal (accordions), Opera Veto

Program: Ho (WP)

## Töö lühikokkuvõte

Doktoritöös „Keele foneetilised omadused helilooja töövahendina. Helilooja perspektiiv" („Phonetic Features as a Tool in Music Composition: From the Perspective of a Composer") vaadeldi keele kahe foneetilise omaduse – kvantiteet ja toon – rakendamist heliloomingus. Uurimistöö teoreetilises osas tugineb autor kantoni, norra ja eesti keele foneetilisele analüüsile. Valik tuleneb autori isiklikust kokkupuutest nimetatud keeltega, töös kasutatud muusikanäited pärinevad autori enda loomingust.

Töö piiratud mahtu arvestades on autor keskendunud seitsmele alateemale: 1) kõne artikulatoorsete omaduste interpreteerimine muusikas; 2) väldete eristamine, rütm ja vormikonstruktsioon; 3) helikõrguse ja välte fenomen; 4) toonikeelte helikõrguste suhtelisus ja selle mõju laulmisele; 5) norra keele rõhutüüpide ja kantoni keele toonide kasutamine heliloomingus; 6) kõne helikõrgus ja selle esinemine muusikas 7) toonikeelsete laulude esitamine helilooja vaatevinklist. Tegemist on interdistsiplinaarse tööga, milles vaadeldakse keele foneetiliste omaduste (välde ja toon) mõju helilooja loomeprotsessile uurimistöö autori laulutsüklis *natt-öö-夜*, elektroakustilises helipoeemis *isjoriske forsteder,* keelpillikvartetis *vihik (a),* miniatuurses kontserdis metsosopranile *vihik (c),* kolmkeelses kooriteoses *vihik (d),* kakskeelses miniatuuris kolmele häälele *kj/ærlige ord,* ilmalikus missas *messe norvégienne profane* ning kammerooperites *hvorfor pusen?* ja *minn(i)e.* Doktoritöös arutletakse kompositsiooniteooria ja -praktika üle avaramas taustsüsteemis, millesse olid hõlmatud ka teised distsipliinid nagu lingvistika, poeetika, partituuri analüüs, fenomenoloogia ning muusikateadus.

**Esimeses, sissejuhatavas peatükis** tõstatati järgmised loovuurimuse küsimused:

- Mil moel aitab foneetiliste nähtuste nagu välde ja toon tundmine heliloojal genereerida muusikalist materjali, tööriistu ja ideid?
- Kuidas saaks helilooja arendada tööriistu, mis aitaksid mõista häälikute mõju tema loomeprotsessile?

Töö interdistsiplinaarne metoodika tugineb nii muusikateaduse kui häälikuõpetuse fenomenoloogilistele aspektidele. Ka töö koostamisel kasutatud põhilise kirjanduse nimistu pärineb erinevatest valdkondadest ning hõlmab autoreid nagu Roman Jakobson, John Laver, Murray

Schellenberg, Jaan Ross, Ilse Lehiste and Juri Lotman. Erinevaid õpetusi ühendab Hans-Georg Gadameri filosoofiline käsitlus mängust (ingl *concept of play*; saksa *Begriff des Spiels*), kus teaduslikke ja kunstilisi kontseptsioone käsitletakse läbi loova isikuga (antud juhul heliloojaga) toimuva dialoogi. Mängu teooria rakendamine põhjendab ka teineteisest kaugete teadus- ja kunstivaldkondade kaasamine muusikateoste loomisse.

**Teises peatükis** tegeletakse süvendatult kõnelähedase laulumaneeriga helilooja vaatevinklist, tuginedes peamiselt Arnold Schönbergi *Sprechstimme* tehnikale. Ehkki laul ja kõne kuuluvad mõlemad inimese hääleaparaadi väljendusvahendite hulka, räägitakse muusika kontekstis neist kui kahest eraldiseisvast kategooriast. Nii muusikalisi kui kõnehelisid iseloomustavad erinevad võrreldavad ning mõõdetavad parameetrid nagu helikõrgus ja kestus. Autori tõi näitena Arnold Schönbergi tsükli „Pierrot lunaire“ (1912), kuna teos on aastakümnete jooksul tekitanud interpreetide seas poleemikat selle üle, missugused hääle „kõnetaolised omadused“ kuuluvad kokku *Sprechstimme* tehnikas laulmisega. Töös tuuakse välja Pierre Boulezi ja Darius Milhaud' kriitilised seisukohad antud küsimuses. Eelmainitute poolt kasutatud metafooride põhjal võib järeldada, et kõnelähedase laulumaneeri iseloomustamiseks võiksid sobida foneetilised terminid nagu toon (helikõrgus) ning välde (segmentide kestus), mis on antud töö kaks peamist teemavaldkonda.

Töö **kolmas peatükk** käsitleb välte (ingl *quantity*) mõistet keeles. Välte mõistega tähistab autor kõne segmentide pikkust (nt pikk ja lühike silp). Nende segmentide kestust võib mõõta absoluutarvudes. Foneetikud võivad koguda väldete kohta andmeid kõne salvestamisel ja mõõta kestusi kõnelejate võrdlemiseks. Vältuste pikkust mõõdetakse absoluutväärtustes. Foneetikas tegeletakse kõne kvantitatiivsete andmete kogumisega (salvestamise ja mõõtmisega) näiteks erinevate kõneviiside võrdlemiseks. Enne vältustesse süvenemist helilooja vaatevinklist toob autor välja vältusi käsitleva teoreetilise raamistiku eesti ja norra keele häälikuõpetuse põhjal (alapeatükid 3.1–3.3). Nii tuntakse näiteks eesti keeles kolme põhiväldet: lühike (Q1), pikk (Q2) ja ülipikk (Q3), norra keeles aga ainult kahte väldet: lühike ja pikk. Eesti keeles langeb sõnarõhk enamasti sõna esimesele silbile, seevastu norra keeles võib rõhk sõnades asetuda erinevatele silpidele. Välte käsitlemine luules ja muusikas on seotud prosoodiliste rõhkudega, mille kaal võib sõltuda kas osaliselt või täielikult silbile antud pikkusest. Foneetiliste segmentide erinev rõhutamine võib nihutada sõna semantilise tähenduse mõistmist ning olla seega helilooja kunstiline taotlus, nagu näitab autor Igor Stravinski ja Benjamin Britteni teoste põhjal.

Väldetega seonduvad ka kõne artikulatoorsed omadused (ingl *articulatory properties*), mida käsitleb töö **peatükk 3.1** pealkirjaga „Kõne artikulatoorsed omadused muusikalises interpretatsioonis“ („Musical Interpretation of Articulative Properties in Speech“). Sõna artikuleerimine jaotatakse kolme faasi: algus, keskosa, lõpetus (ingl *onset*, *medial* ja *offset*), millele muusikas vastavad heli tekitamise kolm faasi: atakk, heli kestus ja lõpp (ingl *attack, sustain and release*). Neid kahte nähtust kõrvutades lõi autor erinevatele tekstifragmentidele tuginevast materjalist koosneva keelpillikvarteti *vihik (a)*, milles konkreetsetele foneemidele on omistatud sarnase akustilise profiiliga mänguvõte, nt [t] võrdub *pizziccato*'ga. Samas rõhutab autor, et kõne artikulatoorsed omadused võivad instrumentaalmuusikas avalduda pigem tämbriliste metafooride kui prosoodiliste või välteliste imitatsioonidena. Kõne ja laulu artikulatoorsed omadused võivad osutuda heaks abivahendiks heliloojale, kui tarvis on värvida, eristada või võimendada eesti keele kolme vältust. Näiteks kasutab autor oma teoses *kj/ærlige ord* kolmele häälele ja puhkpilliansamblile lühikese eestikeelse sõna *sa* muusikalist pikendamist hääliku [s] pikendamise kaudu, seejuures jääb sõna tähendus samaks.

**Peatükk 3.2** jõuab foneemide järel järgmise astme kõneüksusteni ning vaatleb helilooja silmade läbi silpide, jalgade, sõnade ja lausestruktuuri väljenduslikkust. Kõne kvantitatiivsed üksused (st foneem, silp, prosoodiline jalg, sõna) sõltuvad tugevalt konkreetse keele omapärast. Temporaalsed üksused võivad erinevates keeltes avalduda kas silpide või rõhkudena. Autor vastustab nt keeleteadlase John Laveri propageeritavat intuitiivset lähenemist, mille kohaselt silbiüksuste eristamise aluseks on uurija enda keelevaist. Kirjandusteadlane Juri Lotman on väitnud, et häälikute kordumisel moodustub rütm, kuna lugeja tajub, kus on suuremate kvantitatiivsete üksuste puhul rõhud. Antud subjektiivset väidet ei kinnita küll ükski teaduslik eksperiment, kuid autor peab seda piisavalt põhjendatuks luule- ja muusikaloome kontekstis. Lotmani idee sarnaste helide kordumisel tekkivast rütmist oli aluseks uurimistöö autori teosele *isjoriske forsteder* (2019), milles kasutatakse korduvaid silpe teose vormilise ühtsuse loomiseks. Korduvad silbid on seotud sarnaseid semantilisi seoseid tekitavate sõnadega, mida nimetatakse semantiliseks väljaks (st muusikas erinevaid tähendusi ühendavad žestid). Tekkiv keeleline struktuur määrab ka muusikateose struktuuri. Teoses *vihik (a)* (2019) uuritakse sõna kui foneetilist segmenti. Sõnade põhjal arendati välja muusikalised žestid ja motiivid, mis kujundasid teose miniatuurse vormi. Seevastu teoses *messe norvégienne profane* (2018) katsetas autor prosoodilise jala alusel loodud muusikaliste žestidega, mille kordustest moodustus pika kestusega muusikaline vorm. Kolmkeelse teose *vihik (d)* (2019) aluseks on lingvistiline isokroonsus (samaaegsus), st igast töös käsitletavast

keelest tuletas autor erineva tempostruktuuri ning saavutas nende kohakuti asetamisega teose polüfoonilise kontrasti.

**Peatükk 3.3** kõneleb helikõrguse ja välte seosest eesti ja soome keeles. Mõlemale keelele on iseloomulik, et võrreldes rõhutu silbiga algab rõhuline silp kõrgemalt helikõrguselt. Ehkki helikõrgus on fonoloogilisele pikkusele alluv kaasnähe, saab väita, et eesti keele vältuste puhul on lühike (Q1) ja pikk (Q2) silp kõrgema helisagedusega kui ülipikk silp (Q3), mida väljendatakse madalamal helisagedusel (st madalam helisagedus saabub teatud hääldushetkel). Heli laskumine võib katkeda kõnetempo ning muude asjaolude toimel. Soome ja eesti keeles ei ole pikad ja lühikesed silbid küll seotud konkreetsete helikõrgustega, kuid nii fonoloogilistes fraasides kui lausestruktuurides on täheldatav, et helikõrgus on laskuv. Sama trendi kinnitab Eva Liina Asu oma uurimuses „The Phonetics and Phonology of Estonian Intonation“ (2003). Käesoleva töö autori arvates on võimalik, et ka Veljo Tormis on oma emakeelevaistu ajel taolist nähtust oma teostes intuitiivselt kasutanud. Siinse töö autor on teadlikult rakendanud keele nimetatud omadusi enda vokaalteostes nagu näiteks *kj/ærlige ord* (2019). Pole välistatud, et eelmainitud hüpoteesid võivad mõjutada heliloojaid, kes kirjutavad soomekeelsetele tekstidele. Suurema kõneläheduse saavutamiseks võivad heliloojad rakendada nimetatud foneetiliste fenomenidega seonduvaid helikõrguslikke ja rütmilisi iseärasusi oma teostes. Vaadeldud peatüki eesmärk oli juhatada sisse järgmine, foneetilisest helikõrgusest kõnelev peatükk.

Käesoleva uurimistöö teine peamine foneetiline teema käsitleb tooni. **Neljandas peatükis** tutvustab autor üldisemalt akustilisi sageduste tajumist kuulaja poolt. Toon või helikõrgus on üks neljast peamisest valdkonnast, mida käsitletakse inimhääle ja suulise kõne kontekstis. Sõltuvalt keelest võib kõnetoon mõjutada teksti tähenduse mõistmist. Vastavalt kõnetooni kasutamisele jaotatakse keeled kahte suurde gruppi: toonisüsteemil põhinevad keeled (nt hiina keele erinevad dialektid, tai keel, eve keel) ja intonatsioonisüsteemil põhinevad keeled (enamik germaani ja romaani keeltest). Osa skandinaavia keeli (nt norra ja rootsi keel) on hübriidsed keeled. Toonid jagunevad omakorda kontuurseteks toonideks (st helikõrgus muutub madalast kõrgeks või vastupidi) või registritoonideks (st helikõrgus püsib muutumatuna). Suulises kantoni keeles on foneemiliste tähenduste eristamiseks kuus erinevat tooni, norra keeles aga kaks kontuurset tooni, mis varieeruvad dialektide lõikes. Kantonikeelset vokaalmuusikat luues on seega oluline laulu ja kõne täpne vastavus, norrakeelses vokaalloomingus pole taoline sidusus oluline. Keelte ja toonide

käsitlemine foneetilises kontekstis lubab autoril järgmises peatükis liikuda edasi nende rakendamise juurde heliloomingus.

**Peatükis 4.1** uuritakse mõningaid strateegiaid ja käsitlusi, mida autor kasutas kantoni häädusega hiinakeelsetele tekstidele komponeerimisel. Peatükk võtab luubi alla semantilise ühtsuse säilitamise toonikeelel loodud muusikas. Kuna kantoni, norra ja mandariini keeltes on semantiline suhe helikõrgusse väga erinev, erinevad vastavalt ka komponeerimisstrateegiad. Kantoni keele kuue tooni kasutamine allub rangetele reeglitele, suhtelised helikõrgused peavad olema eristatavad vastavalt keele kolmele registritoonile (madal, keskmine ja kõrge). Seevastu norra keele kahe kontuurtooniga võib helilooja käia ringi vabamalt, ilma et keeleüksuste tähendus muutuks. Kolmkeelses laulutsüklis *natt-öö-夜* (2017) kasutas autor kantoni keele toonistruktuuri meloodiajoonises, ehkki intervallid olid suhtelised. Samuti oli võimalik säilitada melismaatiliste lõikude semantiline ühtsus juhul, kui silpide põhitoon oli üheselt määratletav. Lühendatud toone (st toone, mille lõpetasid *coda* konsonantsed segmendid) mõjutas ka temporaalne aspekt (st kõrgusliku kontuuri või kindla helikõrguse katkestamisest tulenev vältuse lühenemine). Kooriteoses *vihik (d)* (2019) kasutab autor tekstilõikudes järjestikuseid kantonikeelseid silpe vertikaalse harmoonilise plaani arendamiseks. Tänu intervallide suhtele sai autor vabalt otsustada, millisele helikõrgusele asetada madal, keskmine ja kõrge toon, ehkki vektorid (st mineku suund) pidid jääma muutumatuks. Teoses *messe norvégienne profane* (2018) näitlikustas autor kõneldava ja lauldava norra keele võimalikke meloodilisi erisusi. Kuna kõne ja lauldava heli omavaheline suhe oli lõdvem, kasutas autor variatiivsuse loomisel ära sama teksti kõnelemisel ja laulmisel tekkivaid kontraste.

**Peatükis 4.2** käsitletakse norrakeelsete helikõrguslike rõhkude (tooniaktsentide) ning kantonikeelse tooni erinevusi keeles ja muusikas. Erinevused norra keele kõneintonatsioonide ning täielikult toonipõhise kantoni keele vahel esitasid autorile väljakutseid mõlema keele üheaegses kasutamises samas teoses. Norra keele enamikes murretes eksisteeriv nähtus, mille kohaselt tooni kõrgpunkti saabumine ei ole sünkroonis silbi endaga, vaid saabub hilinenult (ingl *pitch peak delay*), muudab keerukaks norrakeelse muusika noteerimise. Silbirõhu või tooni muutumine lükkub mitmesilbiliste sõnade puhul silbi lõpuosale. Traditsiooniline notatsioonisüsteem on selliste nihete ülestähendamiseks ebapraktiline ning autor pidas paremaks norra keele mitteräkijate tarbeks lisada partituuri sõnalised seletused. Norra keele kõnetoonide kasutamine sõltub ka sellest, missugusel foneetilisel astmel toimub teksti komponeerimine. Kiiremates kõnelõikudes muutuvad teatud sõna-

või silbiosad teiste suhtes rõhutatumaks. Ülejäänud tekst põhineb intonatsioonil ning temporaalne rõhk osutub enamasti piisavaks. Autor märkas, et oma kakskeelsetes teostes on ta eelistanud töötada madalamatel foneetilistel tasemetel. Teostes, kus määravaks on saanud kõrgemad tasemed (nt sisu edastamine proosa või dialoogidena), kipub teise hääle (sekundaarse keele) foneemiline fookus nihkuma vastavusse esimese häälega. Teisisõnu, foneemiliselt sarnaste elementide markeerimine, eriti tooni säilitamine, toimus sageli muusikaliselt olulistel hetkedel nagu kadents jms nagu nähtub teosest *minn(i)e* (2020). Et kaks erinevas keeles teksti ei sulanduks, vaid säilitaksid oma lingvistilised eripärad, peab helilooja piisaval määral tajuma erinevate keelte helikõrguste kontuure. Selles tuleb abiks häälikutesse süvenemine ja teksti valjusti lugemine.

**Peatükis 4.3** „Helikõrgus kõnes ja selle esinemine muusikas“ („Pitch in Speech and Its Musical Representation“) uuritakse helikõrguse kui foneetilise omaduse rakendamist suuremate muusikaliste vormide loomisel. Autori teoste, eriti suuremate lavateoste vormistruktuuri on mõjutanud kantoni ja norra keelte kõnetoonid. Algselt foneetikast pärit mõistena muutusid kõnetoonid autori loomingus muusikalisteks allegooriateks, mis hõlbustasid mõtlemist pikemates ajakategooriates. Näiteks on Juri Lotman luulet käsitledes sageli kõnelenud sõnades esinevatest kõlalistest sildadest, mis aitavad ühtlustada teksti. Siin saab tõmmata paralleeli motiivide kordusega muusikas. Viimase edasiarendus on kõnehelide erinev interpreteerimine, tehnika, millel põhineb uurimistöö autori teos *isjoriske forsteder* (2019). Vastupidist tehnikat kasutas autor teoses *vihik (d)*, milles iseeneslikult lõigust lõiku kulgev muusikaline vorm kujunes kõnetoonidest tuletatud harmoonia põhjal. Ooperi *hvorfor pusen?* (2019) aluseks said Ida-Norras kõneldavale murdele omased tõusva intonatsiooniga rõhud või toonid. Viimased mõjutasid lisaks ooperi muusikale ka selle lavakujundust ja dramaturgiat ning tegid võimalikuks loo jutustamise algselt vaid küsimustest koosneva libreto põhjal. Kammerooperi *minn(i)e* (2020) aluseks oli foneetilise sarnasuse (st teineteisega sarnanevate häälikute) printsiip. Teose loo jutustamiseks kasutati erinevates keeltes esinevate sarnaste foneemide kordusi.

**Peatükis 4.4** arutletakse, kuidas õpetada toone toonikeeli emakeelena mitterääkivatele lauljatele. Ka siin lähtub autor helilooja vaatevinklist ning keskendub ühelt poolt partituurijuhistele ning teiselt poolt lisaproovide läbiviimisele. Autori koostöö sopran Elisabeth Hetheringtoniga oma teoste ettekandmisel viis järelduseni, et lauljad suhtuvad võõras keeles laulmisse ettevaatlikkusega. Kui teoses kasutatakse kõnetoone, võib helilooja lauljat aidata järgmiselt:

1) salvestada emakeelse rääkija kõnet;

2) kasutada IPA (ingl *International Phonetic Alphabet*) transliteratsiooni või mingit muud ortograafilise või tüpoloogilise tähistamise viisi;
3) tähenduste, foneetiliste omaduste ja/või muusikaliste elementide ühisosal põhinevaid juhiseid.
4)

Salvestiste puhul tuleks tähelepanu pöörata erinevatele foneetilistele segmentidele (*phonetic segment*), sh teksti valjusti lugemise kiirusele. IPA transliteratsioonid osutusid üldiselt ebapiisavaks, kuna keelte fonoloogilised erinevused võivad põhjustada vääriti tõlgendamist. Sõnalised juhised ning näited on otstarbekad kõikide häälikute selgitamiseks, kuna heli selgitamine visuaalselt pole võimalik. Partituuri lisatud selgituste puhul tuleb heliloojal valmis olla muutusteks, mis võivad tekkida dialoogis esitaja ning tema interpretatsiooni muutumisega.

Viimane, **viies peatükk** pakub võimalusi, kuidas käesoleva uurimistöö teema oleks laiendatav teistsugustele keelegruppidele kui antud töö raames esitatud. Autor on arvamusel, et teistsuguste keelekombinatsioonide puhul on ka kunstilised probleemid erinevad ning seega huvipakkuvad edasiseks uurimistööks, ehkki autori poolt välja pakutud foneetilised tööriistad on üldiselt kohaldatavad erinevates valimites. Autor peab oluliseks interdistsiplinaarse raja jätkamist muusikaloomes, kuna nt antud töös käsitletud mitmekeelsus, aga ka muusikute kirev keeleline ja kultuuriline identiteet on tänapäevases muusikamaailmas muutumas üha tavalisemaks nähtuseks.

Töö lisast leiab doktorikontsertide detailsed kavad ning info loovuurimusliku doktoriprojekti raames valminud üheksa teose kohta.

*Tõlkinud Linus Ganman ja Saale Fischer*

## Appendix. Descriptions of Compositions Derived from This Research

This appendix outlines the major works that were developed as I conducted my artistic research. While the following descriptions are not thorough analyses of the compositions, they aim to provide the reader with a better understanding of the works' origins and relationships to the phonetic concepts explained in the body chapters of this dissertation. The text box at the end of each work's description refers the piece to the body chapters' subsections in which it has been discussed or mentioned. The works are are in the order of the year of completion:

---

***natt-öö-夜*** (2018) — song cycle for soprano, accordion and harp

Languages: Estonian, Cantonese, Mandarin, Norwegian and some Finnish

*natt-öö-夜* is the first composition associated with this thesis. The song cycle of 19 movements, just shy of forty-minutes in duration, juxtaposes the poetry of writers Juhan Liiv, Sigbjørn Obstfelder and He Zhu in the original languages (i.e. Estonian, Norwegian and Old Chinese). One important trait to note is that the Old Chinese text was interpreted in both modern Mandarin and Cantonese; the score provides the performer with instructions on which spoken form is used.

As the initial experiment in the beginning of my research, I freely chose the texts I desired to work with and disregarded the limitations of the different languages. All selected poems were tied to the theme of night. In this, I sought out words within these three pieces of writing which bore similar sounds and had connections to the theme. In moments I was unable to seek out the similar words tied to the theme, I sought out sheerly similar sounding utterances which served as 'portals' moving between the different poems and languages. These ideas became the foundation of the theoretical concept of *phonemic likeness* (i.e. similarly sounding syllables or phonemic units, regardless of language) in my research.

Leading up to the performance, this work presented a challenge which was an inevitable byproduct of this artistic research: diction across the languages for the singer. In seeking out solutions to teach the performer the utterances of the different languages, I interviewed the

performer of the premiere on her experience with learning the work. It was illuminating to understand how performers sing in languages which they do not speak even though this was not a subject that had direct impact on the compositional focus of the research. This allowed me to consider metaphors, performance techniques, notation and compositional tools integrated in the music which would help singers come closer to the intended pronunciation.

The work was performed by Elisabeth Hetherington (soprano), Matti Pulkki (accordion) and Liis Viira (harp) on May 31, 2018 at the Great Hall of *Eesti Ajaloomuuseum* (Museum of Estonian History) in Tallinn.

Chapters of Reference:

- Musical Interpretation of Articulative Properties in Speech (**Chapter 3.1**)
- Relativity of Pitch in Tonal Languages and Its Impact on Song (**Chapter 4.1**)
- Composing with Norwegian 'Pitch Accents' versus Cantonese 'Tones' (**Chapter 4.2**)
- The Composer and the Performance of Songs Set to Texts of Tonal Languages (**Chapter 4.4**)

---

***isjoriske forsteder*** (2018) — an electroacoustic tone poem for solo soprano voice

Languages: Estonian, Finnish, Livonian, Izhorian, Karelian and Northern Sami

I was interested to work with the smallest phonetic segments in this work: phonemes. I was particularly focused on the examination of phonemic structures between related languages, as the transformation of single words — for example, from Estonian to Finnish — could be comparable to musical forms. The Uralic languages form a geographical continuum, with the nearly extinct language of Livonian in Latvia being the southernmost point and Northern Sami in Norway's Finnmark county as the northernmost point. This geographical concept became the foundation of the piece *isjoriske forsteder* (or *Izhorian Suburbs,* in English). The work is originally written for solo voice and large ensemble, in three movements, but is adapted to an electroacoustic version with various field recordings made in suburbs of Tallinn and Helsinki, reflecting the suburban Finno-Ugric landscapes. Only two movements were completed and made available for the public thus far.

I experienced a similar challenge from *natt-öö-夜* in the production of this piece: the soloist performing the work spoke none of the Finno-Ugric languages. To make matters more complicated, endangered languages such as Izhorian and Livonian are even more difficult to realise, as there were few recordings or speakers available. I was fortunate enough to have come in contact with a researcher in Livonian language who was able to assist me with the pronunciation. Although, as the focus of the piece is to isolate and play with the phonetic similarities between these related languages, none of the texts were used in full or put in focus in the piece. The following text is taken from the programme notes of the concert with Elisabeth Hetherington (soprano) and Hans-Gunter Lock (sound diffuser) held on May 2, 2019 at the Estonian National Library's Large Conference Hall:

The initial inspiration came from an odd epiphany: growing up in Finnmark, the northernmost county of Norway, the Sami news programme, *Ođđasat,* was often seen on television. Thus, as I embarked on studying Estonian and learned that the similarly sounding word *'uudised'* bears the same meaning, I became fascinated about the sound transformations of southern Finnic languages extending up to the north. I understood these many similarly sounding — and meaning — words as some sort of musical development. *Isjoriske forsteder* takes, then, its formal structure from sounds of single, simple words with subtle change in sounds as one journeys geographically northwards (e.g. the word 'bay' is *lop* in Livonian, *laht* in Estonian, *lahti* in Finnish and *luovhta* in Northern Sami). These words are developed into motivic elements from their rhythmic structures, since quantity — or duration of phonemes — plays an important role in Finno-Ugric languages. In addition, as the single words transform and showcase the sojourn through the Baltic-Finnic regions, supplementary texts from older times are also used. Among which, Juhan Liiv's poem *Üle vee* indicates the traversing of the Gulf of Finland. Other texts of more abstract imageries include an Izhorian lament collected by Estonian philologist Arvo Laanest from the middle of the twentieth-century and a Livonian story about blue cows which grazed on the Kolka peninsula — *Sinnizt nīemod* (used in courtesy of linguist Tuuli Tuisk and the University of Tartu).

As Izhoria (*Isur*) is somewhat the midpoint of this journey, one could naively say that Finland, Estonia and the many Baltic-Finnic tribes are its 'suburbs'. In this electroacoustic version, sounds from Tallinn's and Helsinki's suburban landscapes reflects this secondary theme of the composition.

This work was generously supported by the Norwegian Arts Council's commission grant.

Chapters of Reference:

- Quantity Distinction, Rhythm and Formal Construction (**Chapter 3.2**)
- On the Pitch-Quantity Phenomenon (**Chapter 3.3**)

***vihik (a)*** (2018) — for string quartet

Languages: Estonian, Norwegian, Livonian, Finnish

The *vihik* series of compositions was realised when I began to gather fragmented text materials in sketchbooks intended for integrating them into larger works. These foundational materials were scattered jot notes frequently in unrelated languages, written down spontaneously. I realised that the fragmentary brainstorm materials could be juxtaposed and integrated into an independent piece, for the sake of finding out what musical contexts they could offer. *vihik (a)* was the first of such pieces from the period when I was looking at many Finno-Ugric text sources working on *isjoriske forsteder*. The Finno-Ugric fragments of interest in *vihik (a)* were originally intended for other works. In addition to the phrases in my sketches, I also had written lines of texts which were complementary to the found materials. When I was offered an opportunity to work with Rolston Quartet in Canada, I challenged myself to compose with the phonetic theories derived from other works in this very instrumental context. Working with phonetic concepts in an instrumental context had been an artistic question posed by many of my colleagues previously and I took this opportunity to explore this possibility.

The gathered and original text materials were first read aloud for their prosodic characteristics. I paid attention to the timbral, pitch and durational qualities of the utterances and 'translated' these interpretations into playing techniques of the string quartet. I decided to restrict myself to parameters related to quantity and pitch. Short speech segments from the texts would remain short, while long ones would remain long. Voiced consonants would be pitched and articulated, unvoiced would be more percussive. For example, [t] sounds are represented by short staccatos, diphthongs such as [æi] are accentuated by glissandi to show the metaphor of shift, and so on. Forms of the short movements were drawn from the temporal qualities of the text fragments. The work was premiered on February 1st, 2019 in *Theatre d'Alliance Francaise* in Toronto, Canada by the Rolston String Quartet.

Chapters of Reference:

- Musical Interpretation of Articulative Properties in Speech (**Chapter 3.1**)
- Quantity Distinction, Rhythm and Formal Construction (**Chapter 3.2**)

***kj/ærlige ord*** (2018) — for two soprano and mezzo-soprano voices and large ensemble

Languages: Estonian and Norwegian

I was invited to compose a short work for three voices and *Orkest de ereprijs* in the Netherlands in the autumn of 2018. Following the method of the *vihik* pieces, I worked with short poetry fragments by Estonian poet Juhan Liiv (1864 - 1913) and my long-time collaborator Linda Gabrielsen. The title of the work bears two meanings, separated by the slash dividing the Norwegian adjective 'loving' — *kjærlige* — in plural form. Removing the *kj-,* the remaining word — *ærlige* — means honest, in its plural form. *Ord* means 'words' in this instance. The title was given to the piece as the two contrasting poems approach the themes of closeness and intimacy in contradictory ways. This work challenged my creative process of working with not only one, but several singers who did not speak either languages natively. I was mostly focused on the timbral qualities of the texts, particularly with consonant sounds which could be extended over time. I later discovered that the mezzo-soprano soloist was a native Finnish speaker with some degree of fluency in Swedish, which solidified her role as the diction coach for the other two singers. The large wind ensemble with electric bass was designed as a static canvas for the texts. As if it was a miniature concerto, the singers' lines were the driving forces of the three-minute work.

The work was premiered on February 2nd, 2019 in *Gigant* concert hall in Apeldoorn, the Netherlands by Laura Ginström, Ana Maria Lopes and Francisca Branco together with the *Orkest de ereprijs* under the baton of Rob Vermeulen.

Chapters of Reference:

- Musical Interpretation of Articulative Properties in Speech (**Chapter 3.1**)
- On the Pitch-Quantity Phenomenon (**Chapter 3.3**)
- The Composer and the Performance of Songs Set to Texts of Tonal Languages (**Chapter 4.4**)

---

***messe norvégienne profane*** (2018) — for countertenor, brass quintet, optional harp, electronics and audience choir

Languages: Norwegian, some English and Estonian

This secular mass was commissioned by the Norwegian brass quintet *Nynorsk messingkvintett* and Scottish-Norwegian countertenor, Sean Bell. The original texts of the work, inspired by the believer's doubt of god, were written by the Norwegian writer Linda Gabrielsen. This work imitates the church ritual of Sunday services I attended when I was an adolescent. Between sermons and songs, I wove Gabrielsen's text with liturgy-like elements involving the audience in the score. An 'audience score' is placed on the seats in the concert venue before the performance begins. The countertenor, who serves as the cantor and the soloist, directs the audience to sing from the score at certain moments. The audience involvement is short in duration and the passage's beginning pitches are not always given, as I wanted to emulate the uncertain quality of singing of churchgoers who attended the sermons in my youth.

Compared to the other works in my artistic research, the song-speech distinction in this piece bears a vital directive function. Although speech takes on an instructive role in this piece (e.g. telling the audience to sing or pleading to god), its quantities, particularly in repeated phrases and utterances have been applied to the creation of rhythmic structures. This can be heard since the very first notes in the brass quintet of the piece, which echoes the speech rhythm of *'Fader vår'* ("Our father"). The Norwegian pitch accent is focused upon in the sung sections of the piece, as the text has largely informed my pitch choices.

The piece was performed in Norway, Estonia and Finland by the *Nynorsk messingkvintett,* Sean Bell and Liis Viira. The premiere took place on June 9th, 2019 in *Jakob kirke* in Oslo. A short discussion with music writer Alf van der Hagen and the writer Linda Gabrielsen followed the concert. The commission was generously supported by the Norwegian Composers' Fund.

Chapters of Reference:

- Quantity Distinction, Rhythm and Formal Construction (**Chapter 3.2**)
- Relativity of Pitch in Tonal Languages and Its Impact on Song (**Chapter 4.1**)
- Pitch in Speech and Its Musical Representation (**Chapter 4.3**)

***vihik (c)*** (2019) — for solo mezzo-soprano and orchestra

Languages: Estonian, Norwegian

*vihik (c)* is the third piece in my *vihik* series. It combines the poems *Passasje* ("Passage") and *På sjukehuset* ("At the Hospital") by Norwegian writer Hanne Bramness and *** by Marie Heiberg in a single composition. The three short texts focus on the theme of desolation, especially with the biographical context of Heiberg, who spent a large part of her life in an asylum where the poem was written. The work was an exercise in transitioning between languages, particularly between Estonian and Norwegian. Similar sounds in the three texts between the two languages were used as 'portals'. The piece was performed on May 15th, 2019 by soloist Nele Erastus and the Estonian Academy of Music and Theatre's orchestra under the baton of Toomas Vavilov.

Chapters of Reference:

- Compositional Topics on Quantity (**Chapter 3**)
- Relativity of Pitch in Tonal Languages and Its Impact on Song (**Chapter 4.1**)
- The Composer and the Performance of Songs Set to Texts of Tonal Languages (**Chapter 4.4**)

---

***hvorfor pusen?*** (2019) — a chamber opera for three voices and chamber ensemble

Languages: Norwegian

Yet another collaboration with Norwegian writer Linda Gabrielsen, *hvorfor pusen?* was an anomaly within the bulk of artistic outputs in the course of my research. It was initially a submission for the Shanghai New Music Week's Annual Composition Competition in 2019 with little to do with the research. The category just so happened to be chamber operas and I was then already in the early planning phases with Linda on my next chamber opera work, *minn(i)e*. When I presented Linda with the prospects of creating a short prototype theatre piece for this competition, she responded positively with some draft materials which then became the libretto of this work. The competition did not restrict the language choice of the libretto and so it became a natural choice that I would submit this ongoing collaboration in its original language.

The work received the second prize of the competition and found its relevance to my artistic research just a month before its premiere when I had to derive methods to assist the Chinese singers with diction, as none of the performers had any prior knowledge of Scandinavian languages. This necessitated a method which is quite integrated in the practice of my artistic research, as I had to invent metaphors and instructions to convey how the enunciations are achieved with our human voice apparatus. The level of 'success' — here defined as the ability for the singers to recreate the speech sounds in a recognisable manner for native speakers — varied from performer to performer, as they had little time to memorise the score. However, this pressing situation gave way to some useful tools and methods adaptable for future projects.

The chamber opera was premiered on September 18th, 2019 at the Shanghai Conservatory of Music by faculty members of the music school and members of the Shanghai Opera. The conductor was Chengjie Zhang and stage director was Shiru Wang.

Chapters of Reference:

- Composing with Norwegian 'Pitch Accents' versus Cantonese 'Tones' (**Chapter 4.2**)
- Pitch in Speech and Its Musical Representation (**Chapter 4.3**)
- The Composer and the Performance of Songs Set to Texts of Tonal Languages (**Chapter 4.4**)

---

***vihik (d)*** (2019) — for SSSSAAAA choir

Languages: Estonian, Norwegian, Cantonese

Since *natt-öö-夜*, the pilot project of this artistic research, *vihik (d)* was the second work which incorporated all of the focus languages outlined in this dissertation. Projects prior to this work often had combinations of two of the three languages or were studies on a single language. With this work, I participated in a composition competition for SSAA choir with the Swedish-speaking Finnish ensembles Eviva and Exaudio and the work was awarded first prize in the competition. The work drew excerpts from the ancient Chinese text *Shijing* ('*Book of Poetry*'), poetry by the late Norwegian poet Gunvor Hofmo and texts collected by Jakob Hurt *Vana Kannel*. Similarly to *natt-öö-夜*, my process involved seeking out similarly sounding elements of the texts.

As well, the question of semantic compatibility was central, despite the likelihood where the audience would not be able to understand the nuances between the languages. Nonetheless, the texts *"Mis seäl meie õue alla?"* (*'What lies below our yard?'*) from *Vana kannel*, *"Du flytter"* ("You are moving") and *"Ingen har ropt"* ("No one has yelled") by Gunvor Hofmo and **"牆有茨"** ("The walls have thorn") from *Shijing* were selected for creating a narrative of physical enclosement and isolation. Hushing and quiet fricatives from the texts (e.g. [s] or [ʃ] sounds) are central to both the narrative and the sound world of the piece.

Selected by a jury panel consisting of Hanna Kronqvist, Jutta Seppinen, Elisa Huovinen and Mia Makaroff, the work was planned to receive two Finnish performances in late April 2020. However, due to the Covid-19 pandemic, the choir had postponed the performances to 2022. Hanna Kronqvist was the conductor for both performances.

Chapters of Reference:

- Quantity Distinction, Rhythm and Formal Construction (**Chapter 3.2**)
- Relativity of Pitch in Tonal Languages and Its Impact on Song (**Chapter 4.1**)
- The Composer and the Performance of Songs Set to Texts of Tonal Languages (**Chapter 4.4**)

---

**minn(i)e** (2020) — a chamber opera for 5 voices, accordions and electric guitar

Languages: Estonian, Norwegian, Cantonese

Lasting over 70-minutes in length, *minn(i)e* is a work which combines texts by Norwegian novelist Linda Gabrielsen, Estonian writer Maarja Kangro and Hong Kong linguist Chapman Chen. At the conception of the project, I tasked each writer to reflect on the diary kept by Wilhelmina 'Minnie' Vautrin, an American missionary in China during the Second World War. The original entries written by Vautrin detail her own daily life as a schoolteacher at the Ginling College and her effort in harbouring refugees on the school premise at the dawn of and during the Imperial Japanese Army's occupation of Nanking. The brutal but clinically documented events in the diary has served

as an inspiration to films, writings and research on the Second World War in China in the past decades.[51]

The production in Tallinn in the spring of 2021 was a joint collaboration with Finnish opera director Anselmi Hirvonen, stage designer Magnus Pind, conductor Lodewijk van der Ree, the Estonian company Opera Veto and instrumentalists hailing from Norway, Finland, Serbia, the Netherlands, the United States and Estonia.

This work puts the dissertation's theories between speech sounds in composition I have explored in the past years into practice. The focus on using multiple languages in a single work is to be noted. The concept that spoken or sung texts are treated as sounds rather than its linguistic function is the foundation of I derive compositional materials. Furthermore, one challenge which distinguished this work from past experiments is the use of three new libretti. In past works such as *vihik (d)* and *natt-öö-夜,* I selected poems and texts which suited my compositional purposes. This freedom was lost in *minn(i)e* as the librettists' style of writing and choice of words are beyond my control. Although, the previous works' curatorial angle was replaced by the advantage of working with living writers where real-time input and feedback were possible. The compositional process relied mainly on my flexibility in manipulating sounds of the uttered texts.

The production and commissioning of the work was generously supported by the Norwegian Arts Council, the Nordic Culture Fund, Nordic Culture Contact, the Norwegian Composers' Society, the Estonian Cultural Endowment, Taike, among many others.

Chapters of Reference:

- Quantity Distinction, Rhythm and Formal Construction (**Chapter 3.2**)
- Composing with Norwegian 'Pitch Accents' versus Cantonese 'Tones' (**Chapter 4.2**)
- Pitch in Speech and Its Musical Representation (**Chapter 4.3**)

[51] Works such as Iris Chang's book *The Rape of Nanking,* the documentary film *Nanking* directed by Bill Guttentag and so forth.