Phonetics-based Techniques in My Compositional Methodology

and

Two Compositions:

*ŠÀ {karāz}* for large ensemble

and

*eschaton according to bēl-rē'u-šu* for percussion trio

Onur Yildirim

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Musical Arts
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

# Abstract

Phonetics-based Techniques in My Compositional Methodology

and

Two Compositions:

*ŠÀ {karāz}* for large ensemble

and

*eschaton according to bēl-rē'u-šu* for percussion trio

Onur Yildirim

This dissertation explores various ways of working with acoustic analyses of speech in music composition. The first chapter presents an overview of whistled languages and discusses their potential to act as blueprints for optimizing phonetic data for compositional use. The second chapter details my workflow for incorporating formant and fundamental frequency analysis data from the phonetics software Praat into my compositional methodology. Broadly inspired by the ways in which whistled utterances transform spoken language, the workflow consists of an analysis phase in Praat followed by the conversion, optimization and orchestration of the extracted phonetic data in the computer-assisted composition environments OpenMusic and bach.

Also included in the dissertation are two compositions that are both informed by phonetics. The first composition, *ŠÀ {karāz}* for large ensemble, contains, among the various ways it attempts to instrumentally imitate speech, a section that is constructed with the help of the workflow described in the second chapter. The second composition, *eschaton according to bēl-rē'u-šu* for percussion trio, engages in a deconstruction of the established roles of speech and

instruments in my music, in which the performers are, at times, asked to imitate the sounds of percussion instruments with their voice, in an attempt to blur the line between speech as "the imitated" and instruments as "the imitators."

# Table of Contents

# List of Figures

# Acknowledgments

I would like to express my heartfelt gratitude to my dissertation sponsor, Brad Garton, for his mentorship, guidance and continuous support throughout the dissertation process as well as during my time at Columbia. I also sincerely thank the members of my dissertation committee, Zosha Di Castri, Mahir Cetiz, Ben Holtzman and George Lewis, for their critical engagement with my work and their insightful feedback.

The acknowledgments would not be complete without mentioning Fred Lerdahl, whose influence on my decision to pursue the intersection of language and music as a major thread in my compositional work would be impossible to overstate. His classes *Composition & Cognition* and *Language & Music* provided a solid foundation for the chapter on whistled languages. Professor Lerdahl was also directly responsible for convincing me to attend the Columbia University Seminar on Language and Cognition chaired by Robert Remez. The work of Professor Remez and other scholars I had the chance to become acquainted with during the meetings of the seminar inspired my compositional thinking in profound ways.

Lastly, I am forever grateful to my family, and above all, to my wife Vlatka, for the love, support and continued encouragement.

# Introduction

Since my introduction to phonetics and phonology as a junior undergraduate student, these two subfields of linguistics have gradually become the primary source from which I derive my compositional ideas. While my initial fascination had more to do with the incredible timbral variety one would encounter in the sounds of speech and how that variety could be mirrored in musical textures, over time, speech as a cognitive phenomenon has become a crucial element of my compositional thinking.

One area that occupied me as a composer has been the differences and similarities in how we perceive music and speech. In other words, where does speech end and music begin? Can a piece of acoustic music seamlessly transition into a state that would be perceived as (or mistaken for) speech? Can it transition back into being perceived as music again? Is it even possible to speak of a natural gradient between the two, or was the gradient I presumed to exist merely a construct in my mind?

Such questions led me to listen to and examine speech recordings obsessively. This examination initially took the form of straightforward spectral analyses. My pieces from this early phase were essentially exercises in one-to-one conversion of the overtone content of speech recordings into pitches and rhythms, without much regard for the underlying phonetic structures. While I was pleased with these compositions as standalone pieces of music, I found it hard to justify the effort. I quickly became aware that the resultant music was capturing little more than the general prosodic contour and pacing of the speech recordings I analyzed. In a way, the analysis data acted more like a reservoir of pitches and rhythms without a clear internal hierarchy. I wanted to compose a more speech-like music—not only in abstraction, but sonically as well. To my surprise, worrying less about what is in speech signals and focusing instead on how speech is produced in the human vocal tract helped me transition into the next phase of my speech-inspired compositional work.

It occurred to me that the vocal tract itself resembles a musical ensemble in the sense that its constituent parts responsible for specific types of articulations work in tandem to produce different speech sounds. Starting with this premise, which, in some ways, alludes to articulatory phonetics as well as the source-filter model of speech production, I assumed that assigning specialized roles to specific instruments in which they would try to imitate timbral profiles of phonetic segments could help me come up with speech-like sonorities in my music.

My earlier exposure to Peter Ablinger's player piano piece *A Letter From Schoenberg* had already demonstrated to me that the cognitive mechanism for perceiving a music passage based on speech analysis as speech instead of music was akin to flicking a speech mode switch in the brain.[1] With the right cue, which, in the case of Ablinger's piece, happened to be displaying the text of the analyzed speech recording along with the performance, the listener seemed to immediately perceive the piece of music as intelligible speech. I was thoroughly fascinated by this phenomenon, yet I also wondered whether a similar effect could be attained with a more heterogeneous selection of acoustic instruments played by human performers. Furthermore, since I was not interested in the intelligibility of speech but rather the *recognition* of the resemblance of a musical passage to some form of speech, I wondered if this recognition mechanism could be engaged without an explicit visual cue to prompt the listener. In some ways, I wanted my music to appeal to a deeper level of cognitive familiarity such that, in an ideal scenario, the listener would be left wondering whether a performer from the ensemble had just spoken.

[1] Peter Ablinger, "A Letter From Schoenberg," 2008, https://ablinger.mur.at/txt_qu3schoenberg.html.

The vocal-tract-as-ensemble (or ensemble-as-vocal-tract) idea allowed me to come surprisingly close to these two goals. While I still frequently referred to spectrograms of speech recordings during the composition phase, my focus shifted from solely orchestrating individual sine-wave components of speech to finding single and layered instrumental sounds that best imitated specific phonetic segments. Within this context, I also started using software specifically built for phonetics work such as Praat,[2] which has been a mainstay in my compositional work ever since.[3]

The answer to my earlier question about the existence of a natural gradient between music and speech came to me through my discovery of whistled languages, which will be discussed extensively in Chapter 1 of this dissertation. These auxiliary languages, which have evolved to rely exclusively on whistling to overcome certain acoustic challenges pertaining to their environment, may sound like music performances to those unfamiliar with them. Yet, even while reducing the wealth of phonetic information otherwise present in ordinary speech into whistled utterances, they maintain intelligibility and fulfill their function as languages. As such, it has been argued that they exist as intermediary cognitive phenomena between music and speech,[4] and this premise has also been corroborated by recent research in neuroscience that

---

[2] Paul Boersma, "Praat, a System for Doing Phonetics by Computer," *Glot International* 5, no. 9/10 (2001): 341–45.

[3] There are several reasons why specialized tools such as Praat outperform general purpose acoustic analysis software when dealing with speech signals. Firstly, because Praat contains a toolset, workflow and analysis algorithms tailored for speech signals, it is significantly easier to obtain accurate fundamental frequency and formant analysis results from it without the need to spend hours adjusting analysis parameters. Anyone who has tried to extract the fundamental frequency contour of a speech recording using some of the widely used audio analysis tools knows that even such a supposedly simple task is far from trivial. Secondly, Praat offers a very streamlined user interface for annotating speech recordings with phoneme labels. This feature makes working on and extracting specific portions of speech recordings a breeze. Furthermore, the annotation files can be exported for later processing in other software environments—a process I will be talking about in more detail in Chapter 2.

[4] Susan Glaser, "The Missing Link: Connections Between Musical and Linguistic Prosody," *Contemporary Music Review* 19, no. 3 (2000): 137–38, https://doi.org/10.1080/07494467.2000.11689734.

demonstrates simultaneous activity in brain regions previously thought be exclusive to either language or music when native whistlers listen to whistled utterances.[5]

Whistled languages have been an important source of inspiration in my recent compositional methodology where I strive to find ways to work with a reduced but strategically chosen set of phonetic features while still trying to maintain a certain degree of speech-like quality. In particular, the tendency of certain types of whistled languages to closely follow the second formant contour of ordinary speech in their utterances turned my attention to individual formant contours as sources of compositional material. The work of Robert Remez on sine-wave speech, where a fair degree of intelligibility is achieved only with sinusoidal signals tracking the frequency and amplitude contours of the first three or four formants of speech, also showed me how a significantly reduced set of acoustic signals can successfully convey phonetic information.[6] It was remarkable to observe that the formant information I extracted from Praat, when sequenced using piano sounds instead of sine-wave signals, even in semitone resolution, produced similarly intelligible results.

My recent interest in working with a reduced set of phonetic data motivated me to develop a specific compositional workflow which I first used in the last section of my large ensemble piece *ŠÀ {karāz}*, which is one of the two pieces that are included in this dissertation. Chapter 2 presents a detailed description of the said workflow. It includes a walkthrough of the patches I developed for the computer-assisted composition programming environments

---

[5] Onur Güntürkün, Monika Güntürkün, and Constanze Hahn, "Whistled Turkish Alters Language Asymmetries," *Current Biology* 25, no. 16 (August 17, 2015): R706-708, https://doi.org/10.1016/j.cub.2015.06.067.

[6] RE Remez et al., "Speech Perception without Traditional Speech Cues," *Science* 212, no. 4497 (May 22, 1981): 947–50, https://doi.org/10.1126/science.7233191.

OpenMusic[7] and bach[8] (which runs in Max[9]). These patches allow me to import formant and fundamental frequency analysis data from Praat and optimize the imported content for compositional use. The Appendix section contains the screenshots of these patches. It is crucial to stress that other sections of *ŠÀ {karāz}*, while still based on analyses of speech, were composed using methodologies that are beyond the scope of this dissertation.

The second piece included in the dissertation, *eschaton according to bēl-rē'u-šu* for percussion trio occupies a peculiar spot among my compositional work informed by phonetics. Even though the percussionists are frequently asked to perform phonetic actions, the piece does not contain any material directly based on computer analyses of speech. Instead, I rely on my memory and imagination in trying to choose percussion sounds that may resemble certain phonemes. However, perhaps an equally important aspect of the piece is that it aims to overturn the previously established roles of speech as "the imitated" and the instruments as "the imitator" in my music. As such, certain percussion sounds in the piece, instead of merely fulfilling a role as imitations of conventional speech sounds, are primarily there to be imitated by a less conventional phonetic action—perhaps not unlike beatboxing. An example of such imitation would be the frequently employed long utterances in the vocal fry register that attempt to imitate the peculiar sound of a rasping stick (also called Reibestock) scraping the edges of various instruments, such as temple blocks.

---

[7] Jean Bresson, Carlos Agon, and Gérard Assayag, "OpenMusic: Visual Programming Environment for Music Composition, Analysis and Research," in *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11 (New York, NY, USA: Association for Computing Machinery, 2011), 743–46, https://doi.org/10.1145/2072298.2072434.

[8] Andrea Agostini and Daniele Ghisi, "A Max Library for Musical Notation and Computer-Aided Composition," *Computer Music Journal* 39, no. 2 (2015): 11–27.

[9] *Max*, version 8.1.11, Visual programming environment for multimedia (Walnut, CA: Cycling '74, 2021), https://cycling74.com/products/max.

# Chapter 1: Whistled Languages as Intermediary Phenomena Between Music and Speech

The first section of this chapter presents an overview of whistled languages, which are auxiliary forms of spoken languages that evolved for communication in acoustically challenging natural environments. As I touched upon in the introduction section, the seemingly intermediary position of whistled languages within the music-language cognitive continuum came as a solution to my quest to come up with transitions into and out of musical material based on speech analysis. Furthermore, finding out about the typological consistency of the acoustic transformation (or simplification) processes at play in determining the frequency contours (i.e., melodic structure) of whistled utterances was noteworthy for me from a compositional perspective in that these processes have the potential to function as natural blueprints for selectively simplifying phonetic data[10] for compositional use.

My spectral analyses of a small number of whistled utterances from the whistled language of Kuşköy, Turkey, which I discuss in detail in the second section of this chapter, largely match the descriptions I have previously encountered in the research literature. In particular, the expected tendency of the frequency contours of the whistled utterances to generally match (or resemble) the second formant (F2) contour of their spoken counterparts can be observed in my analyses. However, the whistled language utterances I have analyzed do also

[10] What I mean here with "selectively simplifying" is picking only the most salient portions of the phonetic data in a given acoustic or musical context. Due to the large amount of frequency information that can be obtained from Fast Fourier Transform-based spectral analyses of speech, especially when the ultimate goal is to compose (traditionally) notated music for acoustic instruments played by human performers, simplification of the acoustic data in both frequency and temporal domains is a practical necessity. While this simplification requirement could also apply to acoustic music based on spectral analyses of sounds other than speech signals, the complexity of the cognitive mechanisms and acoustic cues at play that allow us to recognize a given acoustic signal as (intelligible) speech presents extra challenges when figuring out what portion of the phonetic/acoustic information is particularly indispensable, under what type of auditory conditions and why.

diverge from a F2-like frequency contour structure in certain ways. While possible causes of some of those divergences are already discussed in the existing research literature, where possible, I also try to offer some of my own explanations.

## 1.1 An Overview of Whistled Languages

In her article *The Missing Link: Connections Between Musical and Linguistic Prosody*, Susan Glaser defines the relationship between speech prosody and melody as a continuum, and, in the linguistic domain, mentions whistled languages as phenomena that occupy a middle position between speech prosody and melody.[11] Indeed, to those untrained in this obscure language practice found in many parts of the globe, whistling may sound more like a poor musical performance or an imitation of an exotic bird rather than a language. While imitation—or maybe more accurately expressed as emulation or transformation—is an undeniable element of whistled languages, here the act of imitation pertains to the sounds of natural spoken languages. Therefore, though widely used in the literature, the term "whistled language" can be considered a misnomer since it may lead one to think that whistled languages are independent entities unrelated to natural languages, while the case is quite the contrary.

So, what exactly is whistling in a linguistic context? And why is it used at all? Simply put, whistling or whistled speech is a natural mode of speech that emulates certain salient phonetic features of regular spoken language within the articulatory constraints of whistling. This means that, while whistled forms of languages are always directly based on spoken languages and employ the very same grammar, morphology and syntax of their source languages, they greatly simplify the segmental features of spoken utterances by transforming one or more of these features, such as specific formants associated with certain phonemes, into

[11] Glaser, "The Missing Link: Connections Between Musical and Linguistic Prosody," 137–38.

whistled melodies while discarding the remaining features. Despite this dramatic transformation of a complex phonetic system into a simple stream of modulated frequencies, whistled languages still remain highly intelligible to trained speakers. Building upon the fact that the sounds of whistled languages are completely determined by the phonemic inventories of the source spoken languages, Julien Meyer argues that whistled speech recognition is not radically dissimilar to the recognition of whispered speech,[12] which, while requiring much less training, also dispenses with an important phonetic feature of ordinary speech—namely, voicing.[13] Furthermore, like whispered speech, whistling does not replace but rather complements regular speech under specific conditions.[14]

The main benefit whistling provides to its user is the almost effortless attainment of a loud acoustic signal which dramatically expands the range of communication without the need to put excessive strain on the vocal folds as would be the case with shouting.[15] Whistles can easily get as loud as 120 dB within their typical fundamental frequency band of 1-4 kHz and have an average propagation range of 2-3 km which can extend up to 10 km in ideal acoustic environments.[16] By contrast, the exhaustion threshold of shouting seems to be at around 90-100 dB, which is barely enough to cover a distance of 200 meters.[17] Another important advantage that whistling offers over shouting is the distinctness of its sonic signature which enables it to be

---

[12] Julien Meyer, *Whistled Languages: A Worldwide Inquiry on Human Whistled Speech* (Berlin Heidelberg: Springer-Verlag, 2015), 105, https://doi.org/10.1007/978-3-662-45837-2.

[13] René Guy Busnel and André Classe, *Whistled Languages* (Berlin: Springer-Verlag, 1976), 3.

[14] Meyer, *Whistled Languages*, 3.

[15] Meyer, 2–3.

[16] Busnel and Classe, *Whistled Languages*, 34.

[17] Meyer, *Whistled Languages*, 83.

the salient stream in complex natural acoustic environments with wide-band ambient noise. Because of the almost sine-wave-like pure quality of the whistle signal with only a few (usually quiet) harmonics (Figure 1), much of the acoustic energy is focused on an extremely narrow frequency band, helping the signal to cut through various types of natural background noise, even in the presence of sound dampening or diffusing physical obstructions in the landscape such as trees, scrub and rough terrain.[18]

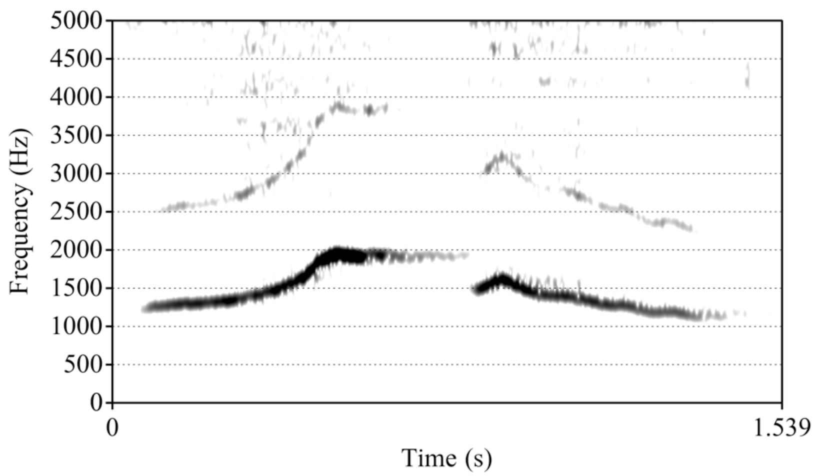



**Figure 1: Spectrogram of a whistle**

Though probably a secondary adaptation, secrecy is yet another aspect associated with whistled languages. Not only is whistled speech largely unintelligible to the speakers of the source language who are untrained in whistling, but many times, as implied in the introduction earlier, it is not even registered as a speech act by unaware strangers. This peculiarity of whistled speech has proven to be a valuable asset for certain populations. A perhaps extreme example of such function is reported from the French village of Aas, where the whistlers engaging in food fraud amidst the food shortages of World War II, successfully managed to inform each other of

[18] Busnel and Classe, *Whistled Languages*, 39.

what would otherwise be unannounced police inspections.[19] Finally, the exclusivity whistled speech grants to the inhabitants of a region inevitably contributes to a sense of community among its practitioners, and whistling may easily become a cherished part of the local identity. In Kuşköy, where the local whistled speech colloquially referred to as *kuş dili* "(bird language") has given the village its name, one of the highlights of an annual festival is a widely attended whistling competition where the criteria are the intelligibility and syntactic complexity of a whistled sentence.[20]

Given the features discussed above, it shouldn't come as a surprise that whistled forms of languages are most commonly encountered in sparsely populated mountainous landscapes, regions with dense vegetation, and generally in locations where inhabitants conduct their daily activities in solitary settings far away from one another. Hunting, shepherding and hill agriculture are some examples of such activities.[21] Lack of an efficient transportation infrastructure has also typically acted as a further contributing factor for the evolution of a whistled language, making it a valuable, if not indispensable, long distance communication method for its users. However, rapid modernization of rural areas and increased coverage of GSM networks around the world render most of the previously vital functions of whistled speech obsolete, making whistling an endangered practice.[22]

In contrast to the geographical and ecological factors that are known to correlate with the global distribution of whistled languages, neither the family a language belongs to, nor its typological profile seem to have an effect on how likely it is for a language to develop a whistled

---

[19] Busnel and Classe, 15.

[20] Meyer, *Whistled Languages*, 35.

[21] Meyer, 1.

[22] Meyer, v.

mode.[23] Languages with attested whistled modes come from many different language families, such as Afro-Asiatic (Tamazight), Indo-European (Greek, Spanish), Niger–Congo (Ewe), Oto-Manguean (Mazatec) and Turkic (Turkish).[24] Among these languages, some are agglutinative, while others are fusional; some have extensive case systems, while others have lost the grammatical case completely. There is however, one inherent feature of a language that influences the *type* of whistled mode it is capable of developing, and that has to do with whether a spoken language is tonal or not.

This brings us to the two main groups of whistled languages: whistled languages based on tonal languages and non-tonal (stress or pitch accent) languages. The main distinction between the spoken forms of these two categories is that, in tonal languages, pitch (i.e., fundamental frequency of the vocal folds) is a salient feature that carries lexical or grammatical information comparable to other phonological units, whereas in non-tonal languages, prosodic features are subordinate to other types phonemic distinctions and mainly used for a limited set of functions such as accentuation and conveying emotions.

Whistled modes of both types of spoken languages leave out or transform segmental features of speech; however, they do so in different ways. When whistled, tonal languages transpose up the fundamental frequency (F0) of the glottal pulse to the whistle range and directly retain the contour. They discard all other features of speech that pertain to segments of the spoken language, such as formants. Surprisingly, since pitch is such an integral part of lexical units in tonal languages, the whistled varieties don't suffer from a significant reduction in

[23] Meyer, 29.

[24] Meyer, 29–30.

intelligibility.[25] An interesting study shows that the discarding of formant information is such a typical feature of the whistled forms of tonal languages that it occurs even in cases where tone carries less phonetic information than the vowel quality.[26]

The whistled register of non-tonal languages, on the other hand, discard the F0 frequency of speech, and the modulated frequency of the whistle usually follows the contour of selected formants of the segmental features.[27] The second formant (F2) is widely cited in the former studies of whistled non-tonal languages as the dominant source of the frequency content of the whistle signals, and the next section of this chapter will discuss this feature in the context of the whistled speech of Kuşköy, which is based on Turkish, a non-tonal, stress accent language.

## 1.2 Spectral Analyses of the Whistled Turkish of Kuşköy

The whistled Turkish of Kuşköy is, in many ways, a textbook example of a whistled language. As mentioned earlier, Kuşköy is a village in the densely vegetated mountainous region of Northeastern Turkey. The area is sparsely populated, and the inhabitants mostly live in houses that are part of small groupings spread apart from each other on the hills. To this day, the road infrastructure remains less than ideal. A river that flows next to the most densely populated part of the village acts as a strong source of ambient noise. Whistled language is still used actively by the local population and plays a vital role especially for shepherds who spend their summers in isolation in the high plateau above the village.[28]

---

[25] Busnel and Classe, *Whistled Languages*, v.

[26] Meyer, *Whistled Languages*, 31.

[27] Busnel and Classe, *Whistled Languages*, v.

[28] Meyer, *Whistled Languages*, 33–35.

For my own whistled utterance analyses, the “corpus” I had access to comprised of short recordings of three whistled Turkish sentences that accompanied a *New Yorker* article[29] about the research conducted by the Turkish-German neuroscientist Onur Güntürkün and his colleagues on the neural processing of the whistled language of Kuşköy.[30] Since I didn’t have access to the spoken versions of the sentences recorded by the original whistler, for the purpose of analyzing the speech formants and comparing them to the whistled utterances, I had to record myself pronouncing the annotated Turkish sentences. This method is problematic for a number of reasons. Firstly, differences between the vocal tract dimensions of individuals create some variations in the formant frequencies. While this variation doesn’t present a problem for speech perception, not having a control data set comprised of the spoken formant values of the whistler makes is difficult to authoritatively conclude whether whistling causes any shifts in formant frequencies. Secondly, the Turkish dialect spoken in Kuşköy is markedly different from the standard Turkish (commonly called “Istanbul Turkish”) I speak, and there are differences in the phonologies of these two dialects which may potentially extend to the formant values of vowels. An example of such difference may be the phoneme /u/ used in Kuşköy, which is reported by Julien Meyer to be the unrounded variety [ʊ] with a much higher F2 value (1200-1500 Hz) than the standard rounded [u].[31] While the F2 central frequency of my /u/ phoneme (around 750 Hz) is slightly higher than the average value (595 Hz) shown in Catford’s book *A Practical*

[29] Michelle Nijhuis, “The Whistled Language of Northern Turkey,” The New Yorker, August 17, 2015, https://www.newyorker.com/tech/annals-of-technology/the-whistled-language-of-northern-turkey.

[30] Güntürkün, Güntürkün, and Hahn, “Whistled Turkish Alters Language Asymmetries.”

[31] Meyer, *Whistled Languages*, 111.

*Introduction to Phonetics*, it is nowhere near the value reported by Meyer and perceptually more similar to the rounded form [u].[32]

The software I have used for my analyses is Praat, and it was created by Paul Boersma and David Weenink of the University of Amsterdam.[33] I have additionally used a Praat script made by Chris Darwin of the University of Sussex which automatically generates sine-wave speech[34] files by creating three individual sine wave streams that track the frequency and amplitude contours of the first three formants of a given speech recording.[35] The sine wave renditions of the formant contours of my voice that I will be comparing to the recordings of the whistled utterances from Kuşköy were created with the help of the said Praat script.

The first recording I analyzed contains the whistled sentence 'Karadeniz çok güzel' ('The Black Sea is very beautiful') (Figure 2).[36] Fitting the descriptions in the literature, the fundamental tone of the whistle is pure and narrow-band, and the harmonics are almost perfect integer multiples of it. Especially in the light of the traditional descriptions of the typical whistle as having weak harmonics (such as the whistle signal in Figure 1), the observed harmonics here are somewhat louder than expected. While the different recording chain may also contribute to such differences, I believe that the real reason is the possible use of the two-finger whistling technique in place of the more common bilabial technique. The two-finger technique produces a very loud whistle intended for longer-distance communication.[37] It is known that a type of

---

[32] J. C. Catford, *A Practical Introduction to Phonetics* (New York: Oxford University Press, 1988), 161.

[33] Boersma, "Praat, a System for Doing Phonetics by Computer."

[34] Remez et al., "Speech Perception without Traditional Speech Cues."

[35] Chris Darwin, *SWS*, Praat script, 2003, http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/SWS.

[36] Nijhuis, "The Whistled Language of Northern Turkey."

[37] Meyer, *Whistled Languages*, 71.

whistling technique employed by the Akha language speakers, which makes use of a leaf, gives more acoustic energy to the harmonics,[38] so the two-finger technique may also have a similar effect on the signal.

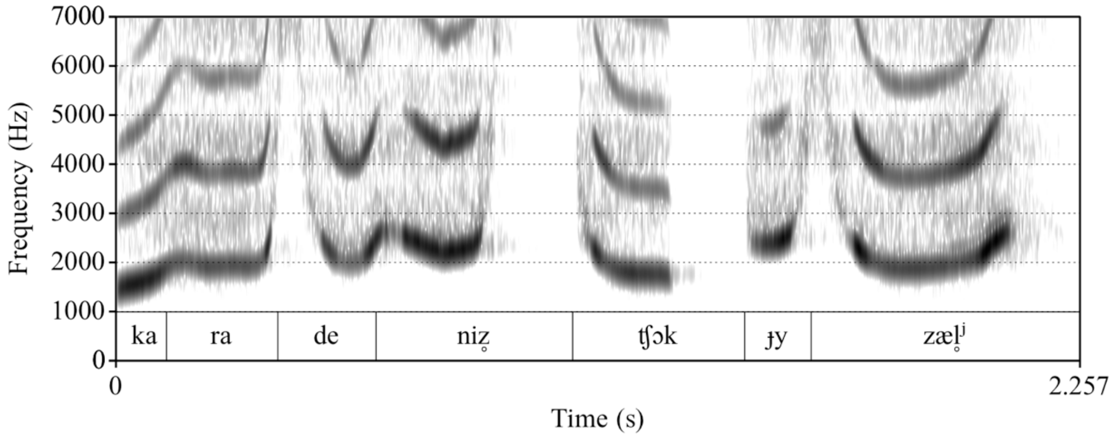



**Figure 2: Annotated spectrogram of the whistled Turkish sentence "Karadeniz çok güzel."**

Upon an aural comparison of the whistle signal to the sine wave rendered version of the second formant values of my own pronunciation of the sentence (Figure 3), the resemblance is striking—especially in the light of the possible sources of diversion I have mentioned previously. A quick look at the spectrogram shows that the overall contours of the two signals are indeed similar. However, some obvious differences also exist. Firstly, the overall range of frequencies is generally higher in the whistled signal. The median frequency of the fundamental of the whistled signal is around 2000 Hz, whereas the median of the F2 signal is around 1600 Hz. While the different vocal tract dimensions or the dialect of the whistler may account for at least some of the differences, according to a research conducted by Meyer and his team, the further whistlers must communicate—or in other words, the louder they need to whistle—the higher the entire

[38] Meyer, 87.

frequency range of the whistle signal.[39] Given the high possibility that the two-finger whistling technique intended for loud whistling is being used in this recording, Meyer's findings may explain the observed difference. Another conclusion we can probably reach from the observed difference in range is that the relative contour may be a more important factor for the intelligibility of whistled speech than matching the absolute frequency values of F2. It may be that the comfortable/idiomatic range of whistling imposes some kind of natural limit on how much the central frequency of F2 may be transposed, ensuring the intelligibility of the signal.

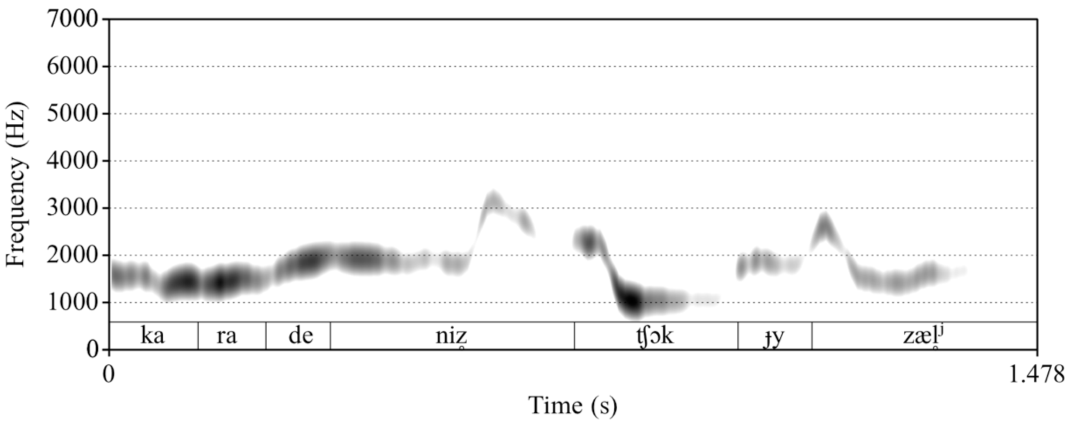



**Figure 3: Sine wave rendering of the 2nd formant (F2) of the sentence "Karadeniz çok güzel."**

My analyses of the two other whistled speech sentences also closely matched my observations from the example above. However, there was a single instance where the deviation from F2 contour was not straightforward enough to explain via vocal tract size or dialectal difference, and that occurred on the 'var mı' [ˈʋar mɯ] syllables in the question sentence 'Taze ekmek var mı?' [taːzɛ ɛcˌmɛc ˈʋar mɯ] ('Is there fresh bread?') (Figure 4).[40] If the whistle was to follow the F2 contour of the spoken sentence (Figure 5), there would be a much gentler rising

[39] Meyer, 81, 109.

[40] Nijhuis, "The Whistled Language of Northern Turkey."

contour on the word 'var' (950 to 1250 Hz). However, the whistle signal quickly goes up in frequency (1250 to 2235 Hz) within the duration of the word 'var' and goes down abruptly at the onset of 'mı' (to 1925 Hz). The contour here is somewhat closer to the F3 rather than F2, and my explanation for the divergence is that it is facilitated by the typical pronunciation of the syllable before the interrogative particle 'mı' (in this case, 'var') with a rapidly rising intonation (about a perfect fifth up in) in Turkish. Even though the exact mechanism at play is not clear to me, the abrupt switch to the F3-like contour may be triggered by a complex interaction of the (discarded) F0 and F2. While Meyer doesn't specifically deal with interrogative intonation in his book, he does mention that stress slightly increases the frequencies of whistled vowels.[41]

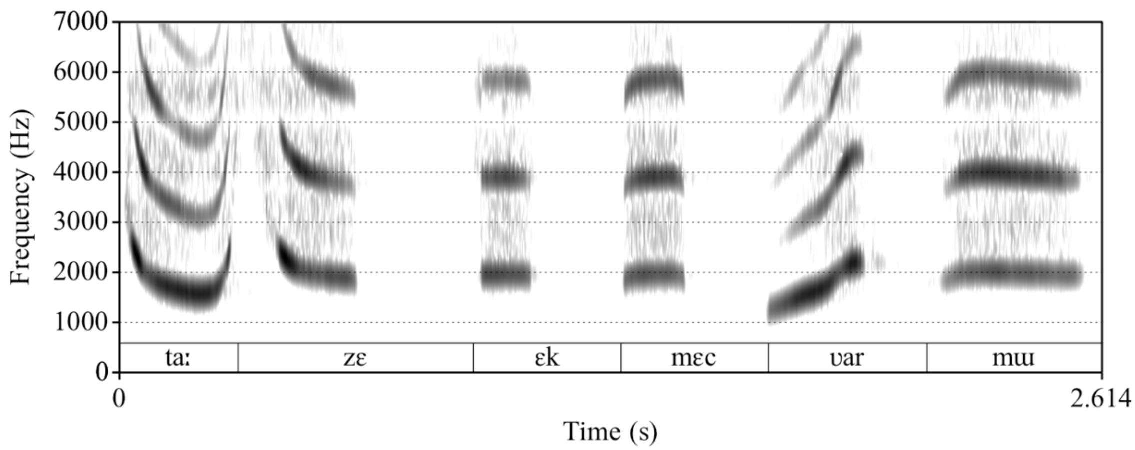



**Figure 4: Annotated spectrogram of the whistled Turkish sentence "Taze ekmek var mı?"**

[41] Meyer, *Whistled Languages*, 113.

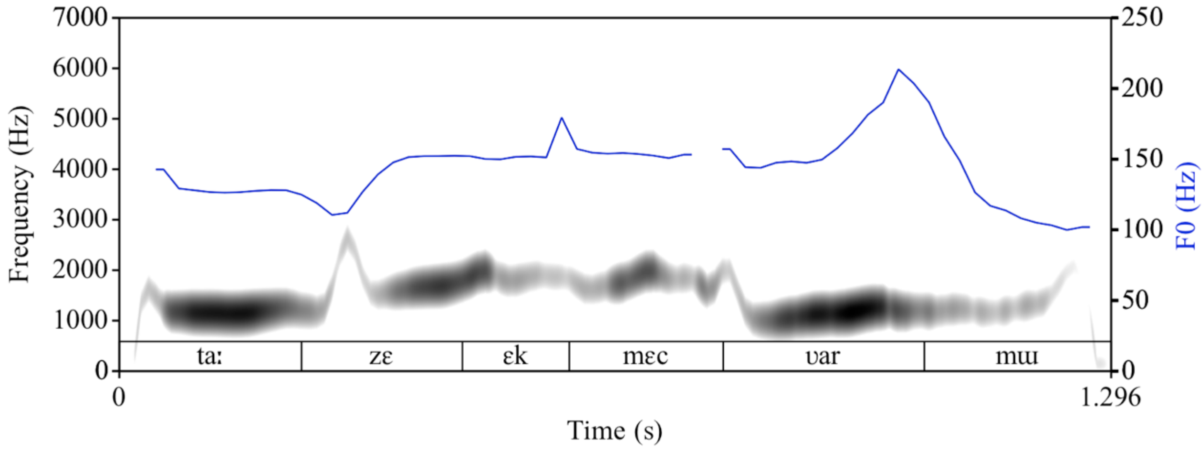



**Figure 5: Sine wave rendering of the 2nd formant (F2) of the spoken Turkish sentence "Taze ekmek var mı?" and the F0 contour (in blue; note the different Hz scale on the right)**

Although my analyses largely confirmed the salience of the second speech formant as the model for the whistled mode of a non-tonal language like Turkish, a more important question remains: What makes the second formant so special in a whistled language context? Before we try to answer this question, a good point to remind ourselves is that traditional whistlers (of non-tonal languages) all report that they do not consciously try to construct a melodic line based on F2 but rather think about the spoken language while they are articulating whistled words. Therefore, the answer may be related to the articulatory constraints of whistling. A team of engineers from the Center for Robust Speech Systems at UT Dallas recorded a whistle corpus by asking 30 subjects (17 males and 13 females) "to capture maximum variability in their whistling style" and "[not] to imitate any particular song or melody."[42] A statistical analysis of this corpus revealed the spectral center of gravity of included whistle signals to be 1600.3 Hz, with a

[42] Mahesh Kumar Nandwana, Hynek Boril, and John H. L. Hansen, "A New Front-End for Classification of Non-Speech Sounds: A Study on Human Whistle.," in *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association* (Dresden, Germany, 2015), 1983, http://www.isca-speech.org/archive/interspeech_2015/i15_1982.html.

secondary smaller peak at around 2500 Hz.[43] After seeing these numbers, I calculated the mean F2 values of the phonemic vowels of Turkish and Greek (using values from *A Practical Introduction to Phonetics*[44]). I ended up with a value of 1616 Hz for Turkish and 1509 Hz for Greek (min./max. F2 values for both languages: 595/2400Hz). Needless to say, both mean values are extremely close to the spectral center of gravity found by the UT Dallas team, which may again suggest that at least part of the answer to the F2 question lies in the articulatory constraints set by whistling.

The other part of the answer may be related to the crucial importance of F2 for vowel recognition. Findings from a study on vowel recognition conducted by researchers from University of Pretoria conclude that, for vowel recognition in quiet or low noise (i.e., typical) conditions, F2 provides a more important cue than F1, and only in severely noisy conditions the auditory system demands a more complete spectral picture and the relative importance of the two formats is equalized.[45] Since whistled languages are much louder than normal speech and, therefore, unlikely to be severely masked by ambient noise, it is probably safe to assume that they are typically in a situation where F2 remains the salient perceptual cue for phoneme recognition.

### 1.3 Additional Remarks

Whistled register of speech is a truly fascinating and elegant adaptation that facilitates social interaction and cooperation in environments that are hostile to regular methods

[43] Nandwana, Boril, and Hansen, 1983–84.

[44] Catford, *A Practical Introduction to Phonetics*, 163.

[45] Rikus Swanepoel, Dirk J. J. Oosthuizen, and Johan J. Hanekom, “The Relative Importance of Spectral Cues for Vowel Recognition in Severe Noise,” *The Journal of the Acoustical Society of America* 132, no. 4 (October 1, 2012): 2652–62, https://doi.org/10.1121/1.4751543.

communication, and it fulfills these duties by strategically (and economically) deploying our cognitive and physical resources. It holds a unique place as a phenomenon that sits halfway between music and language, and recent studies, such as the one conducted by Güntürkün and his team, provide a strong argument for this classification based on neuroscience.[46]

An important area that I haven't explored in this chapter is how consonantal systems of spoken languages are specifically represented in whistled speech. While F2 still seems to be an important element for encoding consonants in whistled melodies, exact processes that take place in their transformations are much less straightforward compared to vowels. Judging from the higher frequency range of certain consonants in spectrograms, F3 may also be systematically relevant in their transformation. However, the limited number of recordings I currently have access to prevent me from generalizing their exact behavior.

Lastly, as I briefly touched upon in the introduction to the chapter, the specific ways in which whistled languages transform regular speech strike me as potential blueprints for novel ways of incorporating phonetic data into music composition. In particular, possible musical analogs for the language-specific, environment-dependent and cognitively pragmatic simplification processes that underlie whistled communication may remedy the sometimes too homogeneous and invariably busy musical textures resulting from a spectrally maximalist and less selective use of phonetic data. While the functionality of the compositional workflow I will be describing in the next chapter goes beyond creating whistle language-like musical passages and is currently more geared towards facilitating the simultaneous use of different combinations

[46] Güntürkün, Güntürkün, and Hahn, "Whistled Turkish Alters Language Asymmetries."

of formant (and fundamental frequency) contours,[47] the initial inspiration most definitely came through my immersion in whistled languages.

[47] In its current form, the output of the workflow can be considered more akin to sine-wave speech, which itself shares similarities with whistled languages.

# Chapter 2: Optimization of Phonetic Data for Compositional Use

This chapter presents a walkthrough of my procedure for employing formant and pitch contour data in my compositional work with relevant musical examples and explains in detail the thinking behind the analysis methods and the data filtering/simplification algorithms I employ. The process can roughly be divided into the below four stages:

1) **Phonetic Analysis:** After cleaning up a speech recording as needed, I analyze the formant (F1-3) and the fundamental frequency (F0) contours of a speech recording in Praat.[48] The analysis stage also includes a phonemic (or phonetic) transcription step where I manually mark phoneme and syllable boundaries and annotate the spectrogram analyses using the International Phonetic Alphabet (IPA) notation.
2) **Data Conversion:** I export the formant and fundamental frequency (pitch) analysis data from Praat and convert it into the OpenMusic[49] syntax using patches I built for the purpose. The conversion phase also involves making important decisions with regards to the optimal rhythmic resolution and dynamic range required for the specific section of music.
3) **Filtering:** This is a very important step where I use a combination of three complementary algorithms that help me reveal and reinforce the inherent rhythmic patterns in the formant and fundamental frequency contours, increase their playability and optimize them for automatic or semi-automatic rhythmic quantization.

---

[48] Boersma, "Praat, a System for Doing Phonetics by Computer."

[49] Bresson, Agon, and Assayag, "OpenMusic: Visual Programming Environment for Music Composition, Analysis and Research."

4) **Quantization and Orchestration:** I use the built-in quantization object in the computer-assisted composition environment bach[50] to rhythmically quantize the filtered contour data imported from OpenMusic. The orchestration of the quantized contours concludes the workflow.

## 2.1 Phonetic Analysis

A good example of the process may be illustrated by taking a closer look at the phonetic material used in the last section of my piece *ŠÀ {karāz}* (see Music Score A, mm. 94-99). As the source material for the section, I picked the sentence “The Invention of Printing, though ingenious, compared with the invention of Letters, is no great matter” which opens the chapter “Of Speech” in Thomas Hobbes’s *Leviathan*,[51] and I recorded myself reciting it. Afterwards, I used Praat to extract formant information from the recording. Figure 6 shows the resultant IPA-annotated Praat spectrogram of my pronunciation of the phrase “The Invention of Printing […]”.

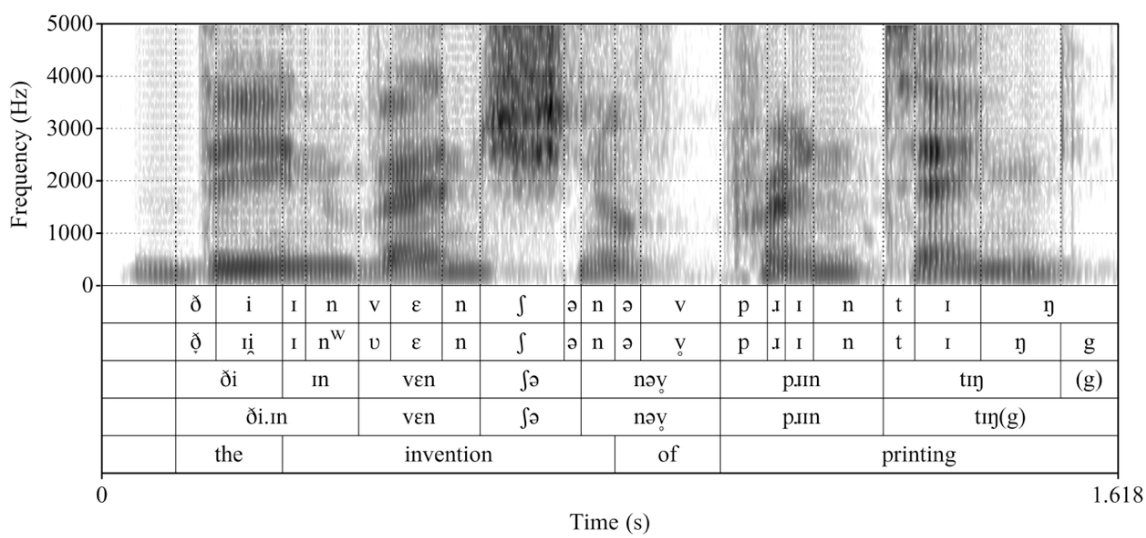



**Figure 6: Annotated spectrogram of the spoken phrase “The Invention of Printing […]” from *Leviathan***

---

[50] Agostini and Ghisi, “A Max Library for Musical Notation and Computer-Aided Composition.”

[51] Thomas Hobbes, *Hobbes’s Leviathan*, ed. W. G. Pogson Smith (London: Oxford University Press, 1909), 23.

My annotations frequently consist of two or more layers (or tiers, as they are called in Praat), and Figure 6 is an example of what those annotations may look like. I usually start by placing the analyzed words in the lowermost tier and carefully aligning their onsets with the spectrogram. The uppermost tier contains a broad (i.e., phonemic) transcription of the speech recording. In other words, it marks phoneme boundaries without providing information about any allophonic variations, or, perhaps more importantly in this case, larger deviations from the expected phonemes resulting from my nonnative pronunciation of English. This tier is immensely helpful for quickly locating specific parts of a recording later in the workflow and extracting any type of Praat analysis data from them.

The second tier, while not a complete narrow transcription in the true sense of the word, provides phonetic transcriptions of only the phonemes that—to my ears—significantly deviate from their expected (i.e., native) pronunciation. In this case, some the peculiarities in my pronunciation include the labialization of the first /n/ in the word "invention" as [$n^w$] and the audible release of the voice velar plosive [g] after the final /ŋ/ in the word "printing". Identifying these types of variations may prove useful for two reasons. Firstly, they clarify any unexpected formant contours I may get later in the analysis process that could potentially confuse me if I were to just refer to the broad transcription; and secondly, I occasionally choose to emphasize these types of pronunciation peculiarities (via louder dynamics, extended techniques, etc.) in the later stages of the compositional process in an effort to better reflect more of the characteristics of the speaker's voice in the final music.

The 3rd and 4th tiers present two alternative ways of syllabifying my specific pronunciation of the sentence. While I casually call this process "syllabification," my real intention is to treat it as a preliminary step for determining the large-scale rhythmic structure of the phrase as I perceive it. These tiers may also continue to separately mark some of the important phonetic features indicated in the narrow transcription tier. As such, the boundaries in these tiers do not always correspond to the conventional syllable boundaries of words or the rhythmic structure that can be algorithmically deduced from the individual formant or fundamental frequency contours. In addition to functioning as a general guide for musical phrasing, the onset times extracted from these tiers often serve as a good reference for making decisions such as picking optimal tempi for rhythmic quantization.[52] A simple patch I built in the computer-assisted composition environment bach uses the onset values and labels extracted from these manual annotations (saved as Praat TextTier files) to generate standalone rhythms or to create phoneme/syllable markers that can be overlaid on musical notation derived from pitch and formant contour data later in the workflow. A rhythmically quantized rendition of the onset times from the 4th tier (along with the corresponding unquantized syllable markers for timing comparison) can be seen in Figure 7.

Praat's default settings rarely give me the optimal analysis results that I can easily import into OpenMusic. For this reason, most of the time, some tinkering with settings such as number of formants, upper/lower boundaries of analysis as well as the length of the analysis window is needed. Once I am happy with what I see in the analysis window (Figure 8), I save the formant, pitch and the annotation analyses as individual files in the Praat "short text" format. Since this file format does away with redundant labeling and saves all values as single items in separate

[52] It is important to stress that determining the optimal rhythmic structure (and, by extension, the musical phrasing) of a spoken phrase is not straightforward and frequently requires me to go back and forth between onset values extracted from different annotation tiers throughout the composition process. While onset times corresponding to manually marked syllables (and other larger groupings) are very helpful in the early stages of composition, I still occasionally go back to the broad and narrow transcriptions to selectively add more inner detail to larger musical phrases.

lines (rather attribute–value pairs), the resultant data files are relatively easy to parse as lists within the Common Lisp-based syntax used in OpenMusic.

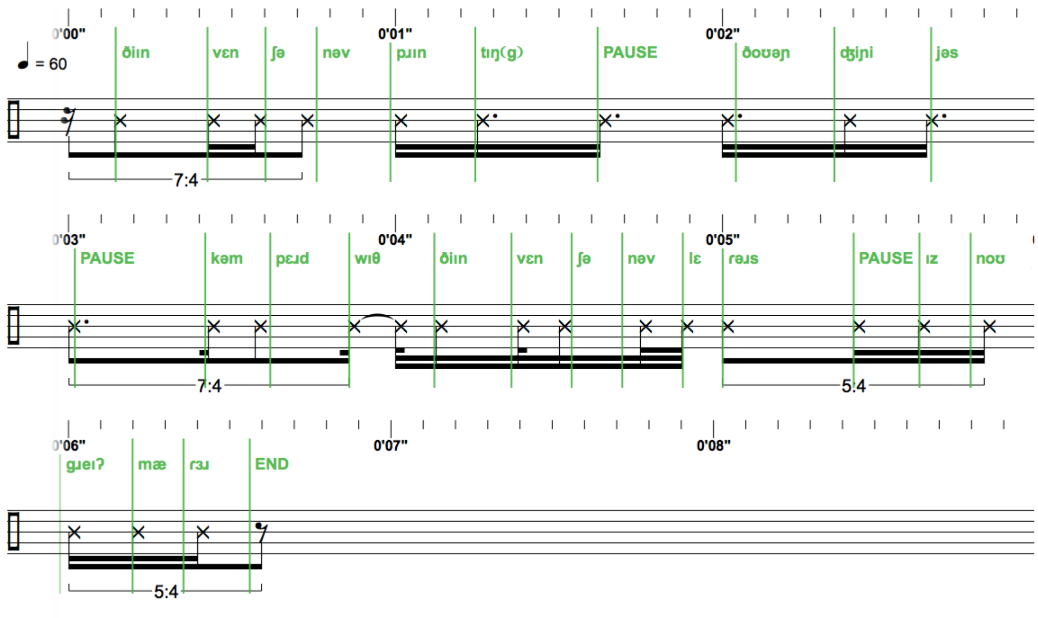



**Figure 7: *bach.score* object displaying quantized rhythms and markers generated from the onset times and labels extracted from the 4th annotation tier of the whole sentence from *Leviathan***

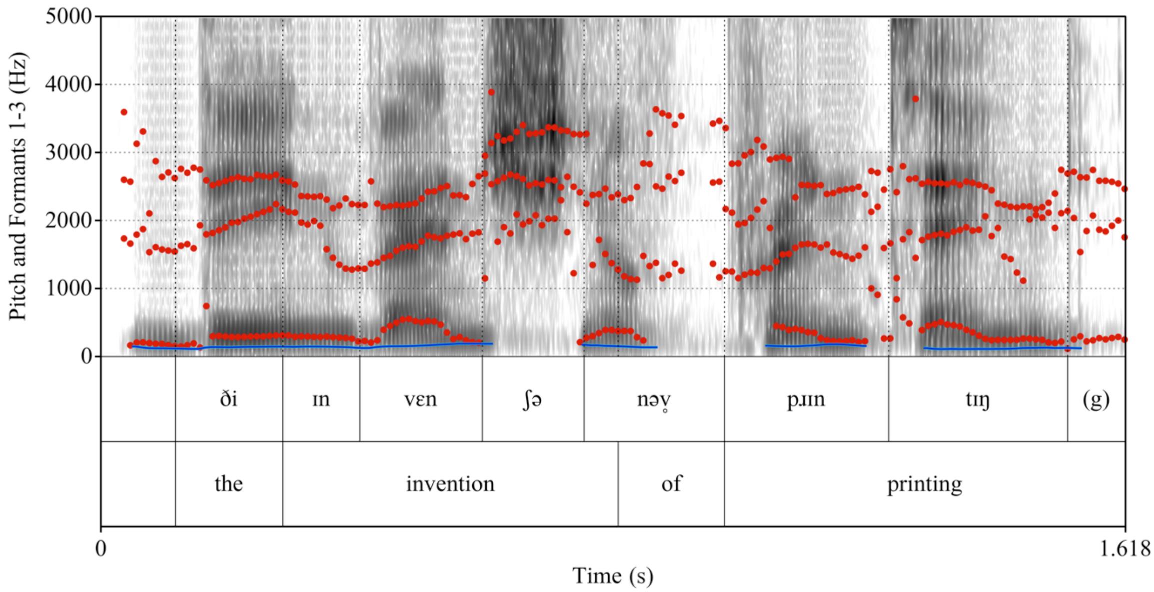



**Figure 8: Annotated spectrogram with pitch/F0 (blue) and formant/F1-3 (red) contours of the spoken phrase "The Invention of Printing […]" from *Leviathan***

## 2.2 Data Conversion

I built two separate Praat data import patches in OpenMusic: one for extracting formant (F1-3) contours (see Appendix B) and another for the fundamental frequency (F0) contour (see Appendix C). Praat offers several ways of exporting the pitch contour data, and the format I chose to use in OpenMusic is PitchTier, which omits amplitude (or intensity) information and simply records the absolute time and frequency of voiced phonetic content whenever the last detected frequency value changes. Since this method does not explicitly specify where voicing ends, it may produce less than ideal results in contexts where very fast changes in voiced and unvoiced phonetic components need to be represented with high accuracy. However, I found that it works perfectly for transcribing pitch contours of speech for acoustic instruments where it would anyway be nearly impossible to accurately reproduce the minute pitch and voicing changes that occur at millisecond intervals.

The OpenMusic patch I built directly uses the time values in the file and converts frequency values to midicents. The lack of amplitude information in the PitchTier data format does not pose a problem as I frequently find myself diverging from the original dynamic shape of the F0 contour frequently when I am composing. As the F0 contour of the phrase "The Invention of Printing […]" shown in Figure 9 illustrates, the absence of direct encoding of where voicing ends in the data also presents little issue here as that information can already be deduced from the length of time where there is no frequency change, which, in a speech context, almost always means no voiced phonemes or, simply, silence. The only user-modifiable parameter of the patch is the step size, which can be increased to integer multiples of the frame size specified in the Praat PitchTier file to reduce the rhythmic resolution further and filter out a large number of frames.

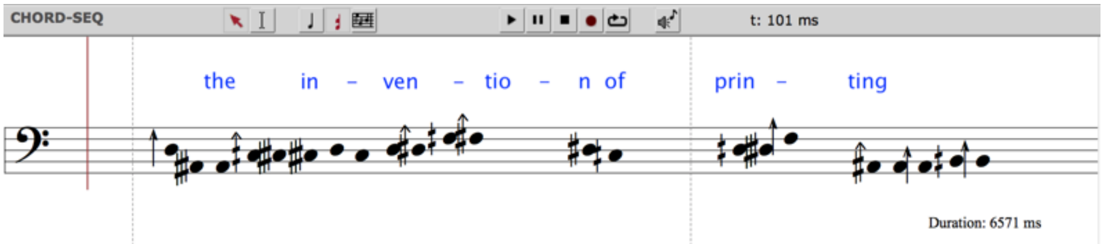



**Figure 9: OpenMusic CHORD-SEQ object with pitch contour (F0) analysis of the spoken phrase "The Invention of Printing […]" imported from Praat (8th-tone approx., 50 ms step size)**

The formant file import patch works similarly for the most part. However, apart from the step size, it has two additional parameters: the formant number, which selects the specific formant stream (F1, F2 or F3), and amplitude (intensity) threshold, which can be set at a specific value (in Pascal) to discard very quiet formant values. The amplitude threshold parameter is crucial because, as opposed to the PitchTier format I use for F0 tracking, the formant analysis files contain regularly sampled frames that also include amplitude information, and one frequently needs to filter out noise that may have accidentally been identified as quiet formants in the Praat analysis. The threshold setting, used in moderation, allows me to have cleaner, more rhythmically defined raw formant material to be used in the later stages of the composition process. The *bach.roll* object in Figure 10 illustrates how gradually increasing the threshold value affects the F2 contour of the spoken phrase "The Invention of Printing […]" (annotated with IPA symbols). The remaining amplitude values can be scaled between a desired dynamic range in OpenMusic as needed. Currently, the patch is only able to process the first three formants. This is usually enough for estimating the phonetic content of the speech recordings, and the very high frequency range of the formants beyond F3 present an increasing challenge for the transcription for acoustic instruments.[53]

[53] However, for the sake of added flexibility, I still plan to introduce the feature in a future iteration of the patch.

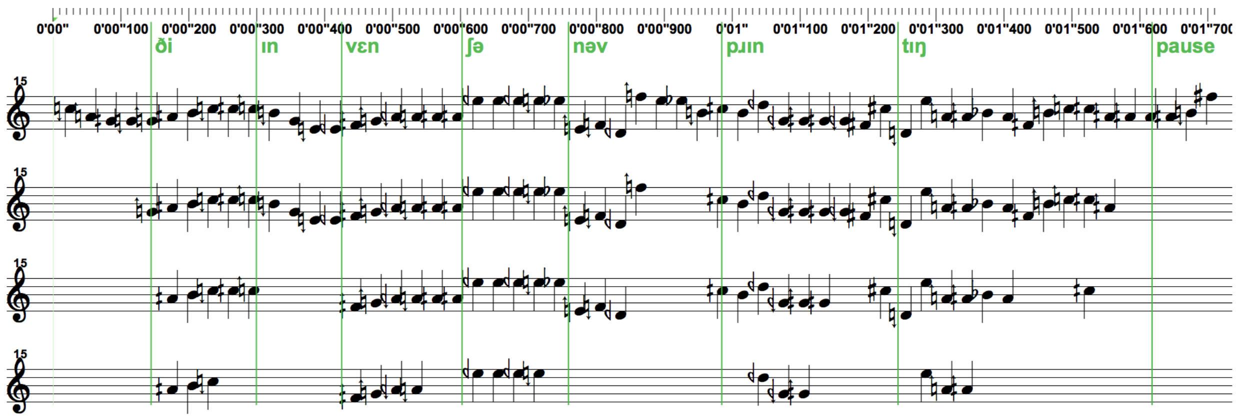



**Figure 10: *bach.roll* object showing the effect of gradually increasing the amplitude threshold value on the F2 contour of the spoken phrase "The Invention of Printing […]" from *Leviathan* (8th-tone approx., 50 ms step size)**

Carefully listening to MIDI sequences of the analysis data rendered with piano sounds remains an important part of my workflow. While increasing the analysis step size (i.e., lowering the resolution) inevitably decreases the intelligibility of the data, it is a compromise one usually needs to make to achieve better (i.e., more playable) rhythmic quantization results later in the compositional process. The real challenge here is finding the "sweet spot" which still preserves an optimal amount of phonetic information while giving me a rhythmic grid that can realistically (and this does not mean easily) be reproducible by human players. Upon listening to the MIDI sequences, I observed that the addition of the fundamental frequency contour to the first three formants increases the intelligibility even at the same step size settings. This means that formant data imported at higher resolution may sound less speech-like than the same data imported at lower resolution but with the addition of the pitch contour. I occasionally exploit this effect as a compositional parameter to control how speech-like I want a certain passage to sound.

## 2.3 Filtering

While some filtering already takes place in the import phase by adjusting the step size and the amplitude threshold, a more "content-aware" filtering is usually needed to optimize the rhythmic structure of the formant and fundamental frequency contours. The reason why some type of rhythmic optimization would make sense is apparent in both spectrograms and in staff notation representations: most of the activity in both formant and fundamental frequency contours involves continuous/interpolated changes in frequency, in other words, a lot of short glides. Some of these glides are sometimes too short or nonlinear to be perceived as such, or they may be frequently interrupted by fast consonants, but the basic observation holds. As such, glissandi or glissando-like techniques would be appropriate musical analogs for the imitation of formant and pitch contours. With this in mind, I decided to come up with a method that filters out the long sequences of redundant pitches extracted from the phonetic data and only gives me some musically relevant ones, including the notes that would aid me in constructing glissandi that approximate the formant and fundamental frequency glides, whenever they occur. The separate patch I built for this purpose relies on three strategies which complement one another: temporally delineating pockets of continuous activity in the data, determining the local frequency extrema of the contours, and detecting abrupt frequency changes that otherwise escape the criteria used by the first two strategies (see Appendix D).

Detecting the locations of the attack and release points of pockets of continuous activity (i.e., sequences of notes) in the fundamental frequency and formant contour arrays is usually a logical first step for the filtering process. The algorithm I employ in the patch takes care of this delineation task relatively easily by picking every element in a fundamental frequency or formant contour array that is separated from the adjacent two elements by an onset time delta larger than the step size (i.e., minimum allowed distance between two notes) specified during the import phase.[54] However, whether a "pocket of continuous activity" is a single phoneme, word,

[54] Another custom duration may also be chosen during filtering for further fine tuning or for altering the intended core functionality of the algorithm.

phrase or sentence is context dependent. As shown previously in Figure 10, a high amplitude threshold value in the import stage (especially in combination with a high step size) is likely to make the formant or pitch contour more disjunct with shorter sequences of continuous notes and more silences longer than the minimum step value. Therefore, in such case, the attack and release filter algorithm would let a higher number of notes pass, and the onsets of these notes may correspond to word, syllable or even phoneme boundaries. Figure 11 and Figure 12 show a comparison of the notes algorithm picks from the F3 contour of the phrase spoken phrase “The Invention of Printing […]” imported into OpenMusic with different threshold values.

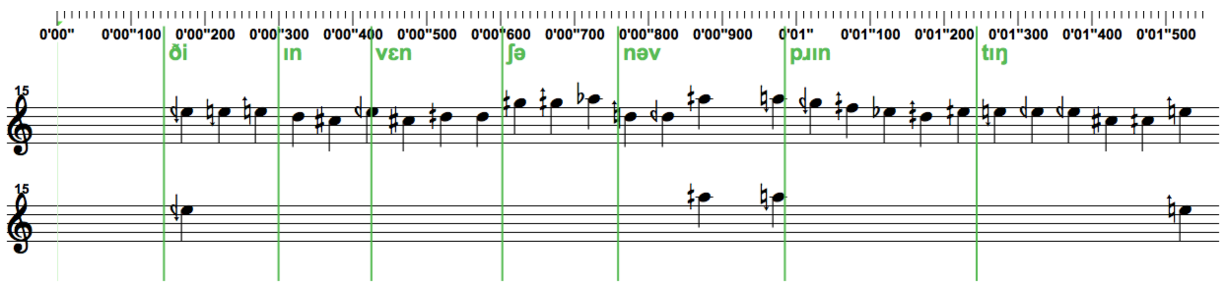



**Figure 11: F3 contour data imported with a step size of 50 ms and an amplitude threshold value of 0.0001 (upper staff) and the attack/release points picked by the algorithm (lower staff)**

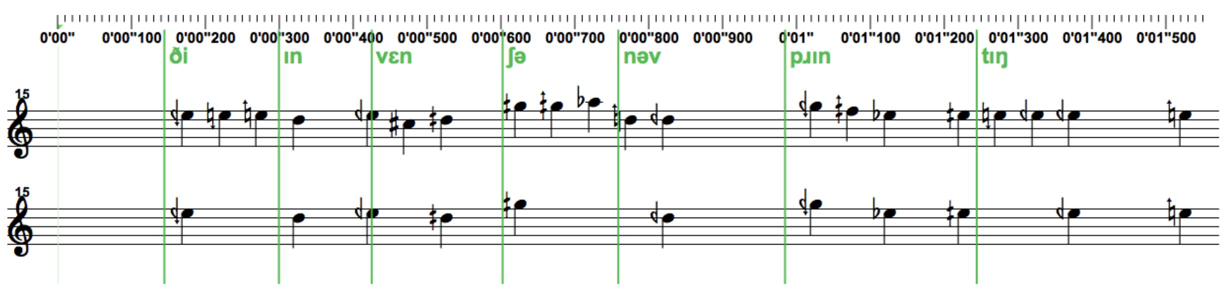



**Figure 12: F3 contour data imported with a step size of 50 ms and an amplitude threshold value of 0.0008 (upper staff) and the attack/release points picked by the algorithm (lower staff)**

As Figure 12 illustrates, the notes that the attack/release filtering algorithm picks from the F3 contour imported with the higher amplitude threshold (0.0008) mostly align with syllable boundaries. However, the resultant sequence still misses some of the anchor notes[55] that can potentially be used for the construction of the rapid glissandi proposed earlier in the chapter as analog for formant or fundamental frequency glides. Examples of such notes are circled in Figure 13.

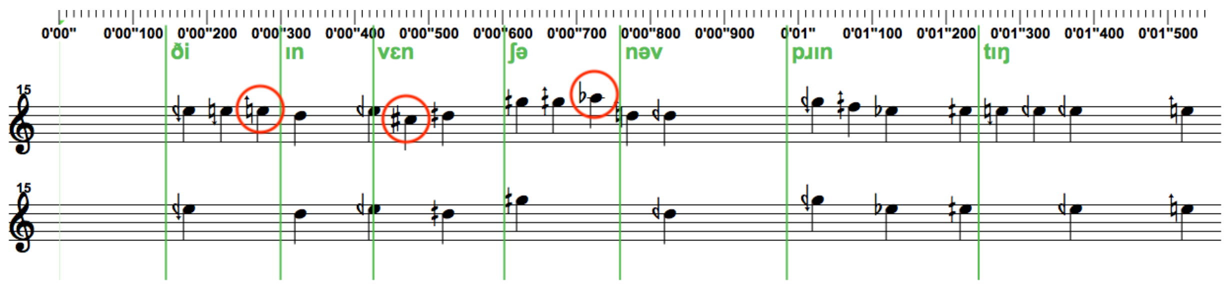



**Figure 13: Potential anchor notes the construction of formant glissandi (circled) not picked by the attack/release filter (lower staff)**

What then is the significance of the notes such as the ones circled in Figure 13? Simply put, they are the anchor notes of the rapid glissandi not picked by the attack/release filter.[56] These notes also happen to usually (but not always) correspond to the local maximum and minimum values (i.e., peaks and valleys) of the imported F3 contour data. As such, the second strategy I employ in my filtering patch is calculating the local extrema (maxima and minima) of a formant or fundamental frequency contour using the following logic: if the pitch value[57] of an element in a contour array is greater than or equal to those of the adjacent two elements, then it is a local maximum. Similarly, the pitch value is a local minimum if it is less than or equal to the

[55] These notes usually mark the beginning or end of a glissando, but as is the case with the circled C#, they can also be the middle point of a bidirectional glissando or even a standalone note (as permitted by the speed at which they occur).

[56] Due to the limitation stemming from the rather rudimentary delineation method that relies solely on a step size threshold (i.e., minimum distance between notes)

[57] Converted from the frequency values in the original Praat data (in Hertz) to midicents in OpenMusic

adjacent two values. Figure 14 shows the notes picked by a combination[58] of the attack/release and local extrema filters and some possible ways to construct glissandi using them as anchor points.[59] Notice how notes such as the $E_{\downarrow 1/8}$ at around the 220 ms mark (under the syllable /ði/) and $G^{\uparrow 3/8}$ at around the 650 ms mark (syllable /ʃə/) are not picked by the local maxima/minima detection algorithm since, in both cases, the preceding note is lower while the following note is higher (i.e. sequences of rising pitches), thereby not meeting the criteria set forth by the algorithm. Going beyond the abstract criteria, if the ultimate goal is to imitate formant glides with rapid glissandi or, in other cases, just to rhythmically simplify the note sequences, then it becomes obvious why those notes are either mostly redundant (i.e., they are within the range of the glissandi) or can be discarded without significant impact on the perceived aural result (the pitch range and the overall contour of the sequence remains the essentially the same).

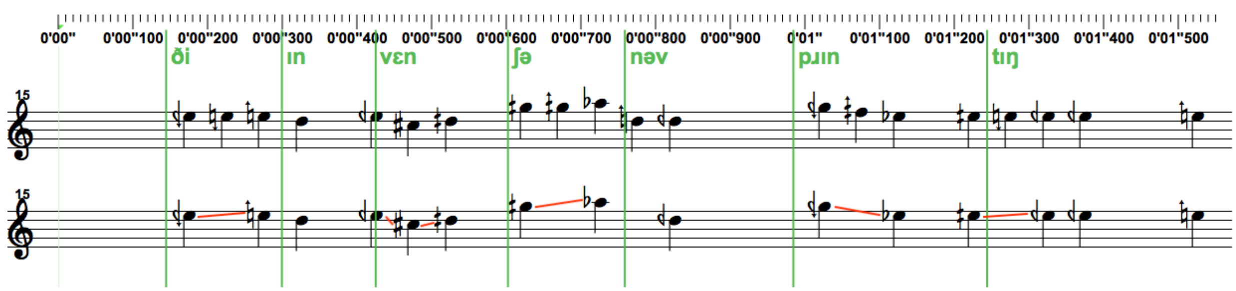



**Figure 14: Notes picked from the F3 contour a combination of attack/release and local extrema filters and some possible glissandi that can be constructed using them as anchor points (lower staff)**

The filtered F3 contour fragment in the lower staff in Figure 14 strikes a good balance of contour simplification and preservation. However, there is still an important note not picked by either algorithm I would pick manually: the $D^{\uparrow 1/8}$ that corresponds to the beginning of the

[58] It is worth mentioning that certain notes may occasionally meet the detection criteria of both filtering algorithms simultaneously (e.g., $E_{\downarrow 1/4}$ at ~420 ms or $E^{\uparrow 1/4}$ at ~1210 ms), but since the OpenMusic patch uses an OR gate to combine the propositional logic of different algorithms, such notes are not picked twice.

[59] The newly added notes by the local minima/maxima detection algorithm are the previously circled ones in the raw F3 contour.

syllable /nəv/ (~760 ms). Although it aurally and visually stands out as a potential attack point (and aligns with the syllable boundary), it doesn't meet the criteria set forth by either algorithm: it is neither separated from any adjacent notes by more than the step size (50 ms) nor is it a local maximum/minimum (Ab > $\mathbf{D}^{\uparrow\mathbf{1/8}}$ > $D_{\downarrow 1/4}$). The last filtering step attempts to detect such notes via an abrupt frequency change detection algorithm using the following logic: if the absolute pitch value difference (Δp) between a note and any of the adjacent notes is greater than the mean Δp of a user-determinable number of notes that come before or after it (a value of 3 or 4 notes per side works well in this instance), then it is considered to be a point of abrupt change. In order to have more control over the detection behavior, I also added an option to specify a permitted deviation amount from the mean Δp value which, in combination with the option to select the number of notes taken into account, allows for a good amount of flexibility. The lower staff in Figure 15 shows the combined output of all three filtering algorithms.

The added flexibility offered by having finer control over some key parameters of the abrupt change detection algorithm prompted me to go back and slightly modify the local extrema algorithm as well. Because the frequency values in the contour data very rarely stay exactly the same between consecutive analysis frames, very small, perceptually insignificant pitch differences between adjacent notes are enough to trigger the local extrema detection. Therefore, it is useful to have a minimum Δp threshold setting below which a Δp value between adjacent elements would not trigger the local extrema detection, in effect getting rid of notes that may be considered noise or simply redundant in the output. A somewhat arbitrarily chosen minimum Δp threshold value of 10 cents was more than enough to get rid of one of the two consecutive notes rounded to $E_{\downarrow 1/4}$ at around the 1300 ms mark in Figure 15 (with midicent values of 9940 and 9944 respectively).[60]

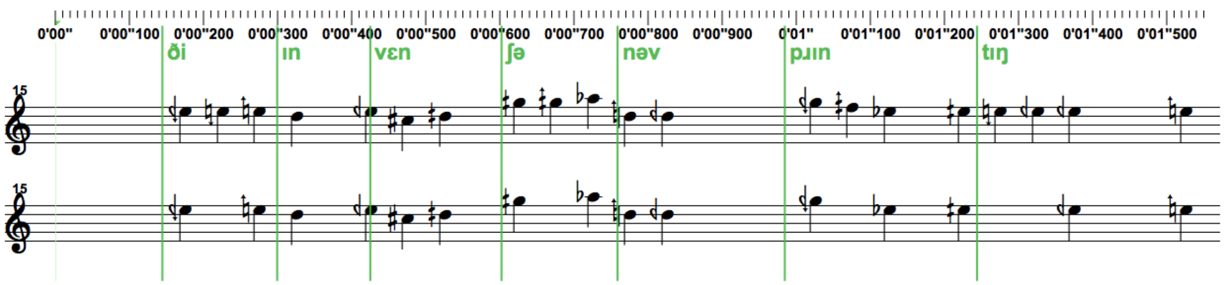



**Figure 15: Notes picked from the F3 contour by a combination of attack/release, local extrema (improved version) and abrupt frequency change filters (lower staff)**

With the last modification to the local extrema detection algorithm, the patch is now ready to process formant and pitch contours for further rhythmic quantization. Figure 16 shows the raw F1-3 and F0 contours of the phrase "The Invention of Printing […]" and their filtered versions. Note that, since I have used a lower amplitude threshold value (0.00014), the raw contours in Figure 16 contain more notes than the previous examples. As demonstrated earlier, this has the effect of capturing more of the quiet formants at the price of decreasing the morpheme delineation performance of the attack/release filter. However, I considered it a necessary compromise and experimented with different minimum Δp threshold (10 to 35 cents) and allowed mean deviation (0 - 50%) settings for each contour to make up for the performance loss. Finally, I have switched from an eighth-tone microtonal grid to a quarter-tone one, which, in this case, offers a good balance of playability and resolution.

[60] OpenMusic and its included libraries contain various functions that can produce similar results, such as *chseq->poly* (a slightly augmented version of which I also employ occasionally), which lets you choose between a semi-, quarter- and eighth-tone microtonal resolution/threshold for merging consecutive notes. However, I find that having finer control over the threshold is extremely useful in the case of phonetic data.

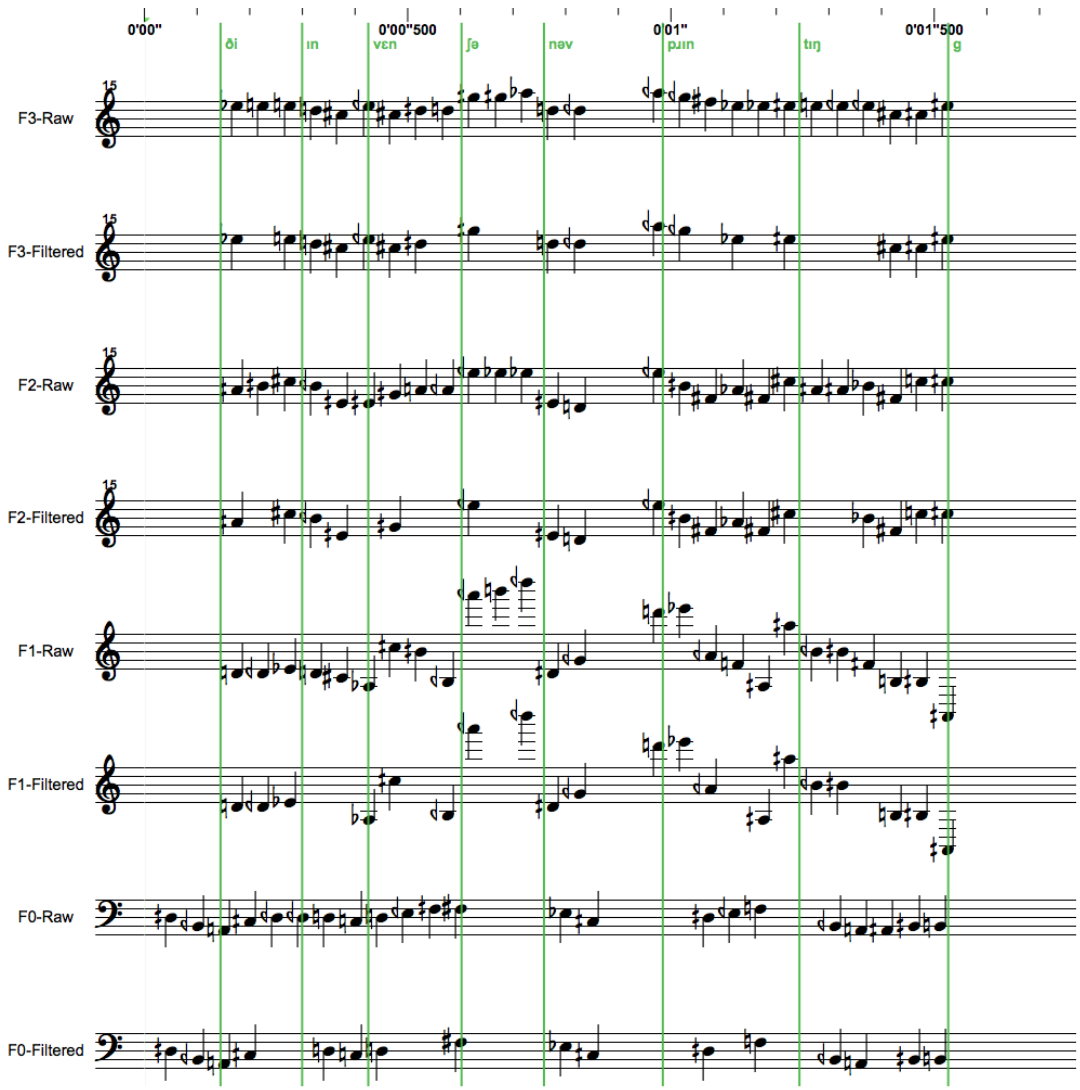



**Figure 16: The raw formant (F1-3) and fundamental frequency (F0) contours and their optimized versions**

## 2.4 Quantization and Orchestration

The resultant rhythmic quantization of the optimized contour data using bach's built-in object *bach.quantize* can be seen in Figure 17. I picked the tempo for the quantization (60 bpm) mainly by trial and error. Even though I tend to change (usually slow down) the original tempi used during rhythmic quantization later in the composition process, I still find it important to find tempi that somewhat preserves (my subjective sense of) the beat structure of the speech recording in question and aligns the boundaries of the larger phrases and/or words with the existing beat structure of a piece as much as possible.

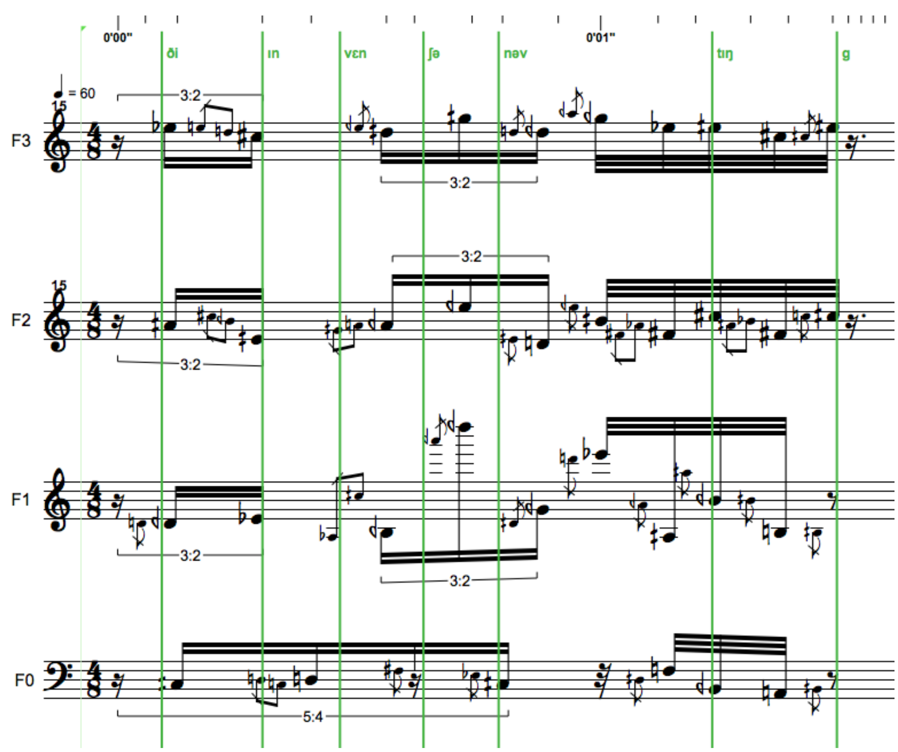
Another important consideration I had in mind was trying to avoid rhythmic units smaller than a 32nd note as much as possible. A tempo of 60 bpm largely fulfilled my criteria. For the sake of avoiding longer sequences of grace notes, I also discarded a few more pitches from some of the contours such as the C# in the optimized F3 contour around the 500 ms mark (within the boundaries of the syllable /vɛn/) in Figure 16. Lastly, I cropped the first three notes of the optimized fundamental frequency contour[61] as bach's quantization algorithm did a better job of approximating the rhythmic structure of the rest of the phrase without them.


**Figure 17: Formant (F1-3) and pitch (F0) contours of the phrase "The Invention of Printing […]" rhythmically quantized in bach**

[61] The notes were later added to the piece in the form of a short glissando in the double bass line in m. 94.

The orchestration[62] of the quantized formant and pitch contours of the phrase "The Invention of Printing […]" can be found in its entirety in measure 94 of *ŠÀ {karāz}* (see Figure 18). Firstly, to facilitate playability and ensure consistency with the preceding section of the piece, I decided to use a slower tempo (♪ = 100) than the one used during quantization. The F2 and F3 contours are exclusively assigned to the two violins, viola and piano. F1 contour is mostly played by the clarinet, but the viola and piano also occasionally join in. While I introduce some variation to how the high strings perform the formant contours later in the piece, the violins and the viola only use touch 3rd and 4th artificial harmonics in this measure. Since most of these artificial harmonics are performed as fairly rapid glissandi, in order to ensure an accurate performance, the overall pitch range of the glissandi and any large leaps need to be kept in check. One outcome of the situation is the occasional need to break a single formant glide gesture into multiple instruments. The first beat of the measure contains two examples of this. The beginning of the F3 contour shown in Figure 17 shared between the 1st violin and the piano, and the accompanying F2 contour is shared between the 2nd violin and viola.

The range of the glissandi and larger leaps remain a constraint for the fundamental frequency contour as well, and, here, that contour is shared by the cello and the double bass. An important technique these instruments both employ is circular bowing, which produces a very speech-like sonority when used in conjunction with rapid, short range glissandi that occurs in a register typical of the human voice.

[62] With the exception of the brief mention of several instrumental sounds (e.g., circular bowing or air sounds) chosen for their overall speech-like quality, the orchestration process described here is mostly limited to preliminary decisions involving instrumental range and playability concerns. A thorough discussion of my orchestration process, which would need to focus on instrumental representations of specific phonetic segments, is beyond the scope of this dissertation.

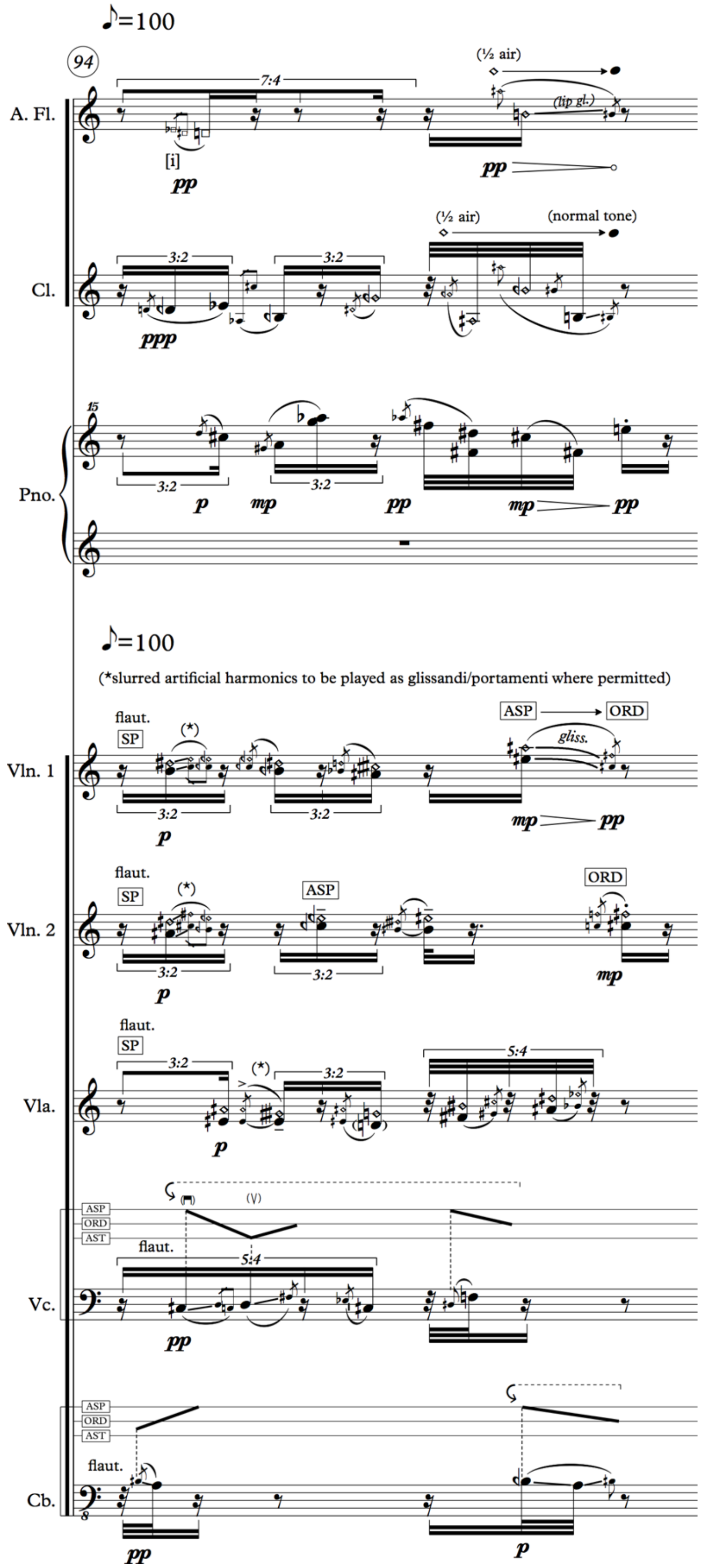
Lastly, the alto flute occasionally joins with air sounds (sometimes articulated with the analyzed phonemes themselves) to reconnect (and hybridize) instrumentally resynthesized phonetic data with sounds reminiscent of real speech.


**Figure 18: Excerpt from *ŠÀ {karāz}* (cropped), *m. 94,* orchestration of the formant and fundamental frequency contours of the phrase "The Invention of Printing […]"**

# Conclusion

My work involving speech analysis has taken different forms over the years, but I have had little trouble extracting new compositional ideas within this paradigm which some composers may consider limited or mundane. The topics in this dissertation offer a glimpse into my most recent compositional methodology. However, even in the context of the two pieces included in this dissertation, the presented thinking should not be considered a "theory of everything" that accounts for all aspects of my methodology that pertain to working with phonetically derived material. Nevertheless, the discussed methods offer ways to overcome the homogeneity problem stemming from constant textural density that plagues most orchestrations of phonetic data as well as a good portion of music based on spectral analysis. Therefore, I expect both the precise transcription of individual formant contours and their usage in combination with other ways of phonetic sonification to be mainstays in my compositional toolkit.

To conclude the dissertation, I would like to present various ways in which the presented research and methodology could be improved:

- My formant filtering methods, while generally serviceable, leave room for improvements. The abrupt frequency change filtering can be made more content-aware by developing a method that selectively takes into account longer or shorter sequences of pitches depending on the context. While the onset times and phonetic labels extracted from manually annotated Praat TextGrid data may provide part of the said context, considering how cumbersome the narrow phonetic transcription process can get when working with a large number of recordings, a separate transient detection algorithm that works in tandem with the existing filtering methods would be a more ideal solution. Ultimately, a black box solution using a deep learning model trained on formant contours and phonetic segments may replace the algorithmic filtering all together. However, I still see some merit in pursuing the algorithmic approach as it provides valuable insight into the structure of formant contours and is easy to customize for different musical scenarios.

- Praat formant data files do not contain separate amplitude information for individual formants. However, there are indirect ways of extracting that information in Praat, and I intend to implement this functionality in the next iteration of the formant import patches as I think having access to that information would increase the rhythmic independence of the different formant contours during the amplitude threshold filtering stage.
- Mainly due to the speed at which individual events occur in speech, as well as the density of salient high frequency content, orchestration of formants (and, in general, phonetic data) remains a challenging task. Developing a workflow that would automate more of this process is one of my long-term goals.
- Currently, I am using three different pieces of software, two of which (OpenMusic and bach) have a good amount of overlapping functionality. The reason for this has less to do with the slight advantage one may offer over the other than me being currently more proficient in OpenMusic. My short-term goal is to move my entire computer-assisted composition workflow to bach, which I am already using for rhythmic quantization and orchestration phases. Among other advantages, bach's integration with Max makes it possible to supplement bach routines with other tools present in Max.

- Praat is open-source, and its functionality has already been ported into other software platforms (e.g., the Parselmouth[63] library for Python and the rPraat[64] package for R). Porting some of Praat's key features directly into Max, either in the form of patches/subpatches implementing some of its analysis algorithms or a more complete Max package with separate objects for different Praat functions, would speed up the workflow significantly.
- An issue I have encountered while researching whistled languages is the scarcity of materials, both in the form of books and a corpus of annotated whistled utterances. For this reason, the next logical step of my research may involve conducting field work in Kuşköy to build a custom corpus of whistled Turkish that would also include ordinary speech recordings of the local whistlers. Such a corpus would help me settle issues related to variations in vocal tract sizes or dialectal features and reach more authoritative conclusions on the sources of frequency deviations between the analyses of formants and whistled signals.

---

[63] Yannick Jadoul, Bill Thompson, and Bart de Boer, "Introducing Parselmouth: A Python Interface to Praat," *Journal of Phonetics* 71 (November 1, 2018): 1–15, https://doi.org/10.1016/j.wocn.2018.07.001.

[64] Tomáš Bořil and Radek Skarnitzl, "Tools RPraat and MPraat," in *Text, Speech, and Dialogue*, ed. Petr Sojka et al., Lecture Notes in Computer Science (Cham: Springer International Publishing, 2016), 367–74, https://doi.org/10.1007/978-3-319-45510-5_42.

# References


Ablinger, Peter. "A Letter From Schoenberg," 2008. https://ablinger.mur.at/txt_qu3schoenberg.html.

Agostini, Andrea, and Daniele Ghisi. "A Max Library for Musical Notation and Computer-Aided Composition." *Computer Music Journal* 39, no. 2 (2015): 11–27.

Boersma, Paul. "Praat, a System for Doing Phonetics by Computer." *Glot International* 5, no. 9/10 (2001): 341–45.

Bořil, Tomáš, and Radek Skarnitzl. "Tools RPraat and MPraat." In *Text, Speech, and Dialogue*, edited by Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, 367–74. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-45510-5_42.

Bresson, Jean, Carlos Agon, and Gérard Assayag. "OpenMusic: Visual Programming Environment for Music Composition, Analysis and Research." In *Proceedings of the 19th ACM International Conference on Multimedia*, 743–46. MM '11. New York, NY, USA: Association for Computing Machinery, 2011. https://doi.org/10.1145/2072298.2072434.

Busnel, René Guy, and André Classe. *Whistled Languages*. Berlin: Springer-Verlag, 1976.

Catford, J. C. *A Practical Introduction to Phonetics*. New York: Oxford University Press, 1988.

Darwin, Chris. *SWS*. Praat script, 2003. http://www.lifesci.sussex.ac.uk/home/Chris_Darwin/Praatscripts/SWS.

Glaser, Susan. "The Missing Link: Connections Between Musical and Linguistic Prosody." *Contemporary Music Review* 19, no. 3 (2000): 129–54. https://doi.org/10.1080/07494467.2000.11689734.

Güntürkün, Onur, Monika Güntürkün, and Constanze Hahn. "Whistled Turkish Alters Language Asymmetries." *Current Biology* 25, no. 16 (August 17, 2015): R693–710. https://doi.org/10.1016/j.cub.2015.06.067.

Hobbes, Thomas. *Hobbes's Leviathan*. Edited by W. G. Pogson Smith. London: Oxford University Press, 1909.

Jadoul, Yannick, Bill Thompson, and Bart de Boer. "Introducing Parselmouth: A Python Interface to Praat." *Journal of Phonetics* 71 (November 1, 2018): 1–15. https://doi.org/10.1016/j.wocn.2018.07.001.

*Max* (version 8.1.11). Visual programming environment for multimedia. Walnut, CA: Cycling '74, 2021. https://cycling74.com/products/max.

Meyer, Julien. *Whistled Languages: A Worldwide Inquiry on Human Whistled Speech*. Berlin Heidelberg: Springer-Verlag, 2015. https://doi.org/10.1007/978-3-662-45837-2.

Nandwana, Mahesh Kumar, Hynek Boril, and John H. L. Hansen. "A New Front-End for Classification of Non-Speech Sounds: A Study on Human Whistle." In *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association*, 1982–86. Dresden, Germany, 2015. http://www.isca-speech.org/archive/interspeech_2015/i15_1982.html.

Nijhuis, Michelle. "The Whistled Language of Northern Turkey." The New Yorker, August 17, 2015. https://www.newyorker.com/tech/annals-of-technology/the-whistled-language-of-northern-turkey.

Remez, RE, PE Rubin, DB Pisoni, and TD Carrell. "Speech Perception without Traditional Speech Cues." *Science* 212, no. 4497 (May 22, 1981): 947–50. https://doi.org/10.1126/science.7233191.

Swanepoel, Rikus, Dirk J. J. Oosthuizen, and Johan J. Hanekom. "The Relative Importance of Spectral Cues for Vowel Recognition in Severe Noise." *The Journal of the Acoustical Society of America* 132, no. 4 (October 1, 2012): 2652–62. https://doi.org/10.1121/1.4751543.

# Appendix A

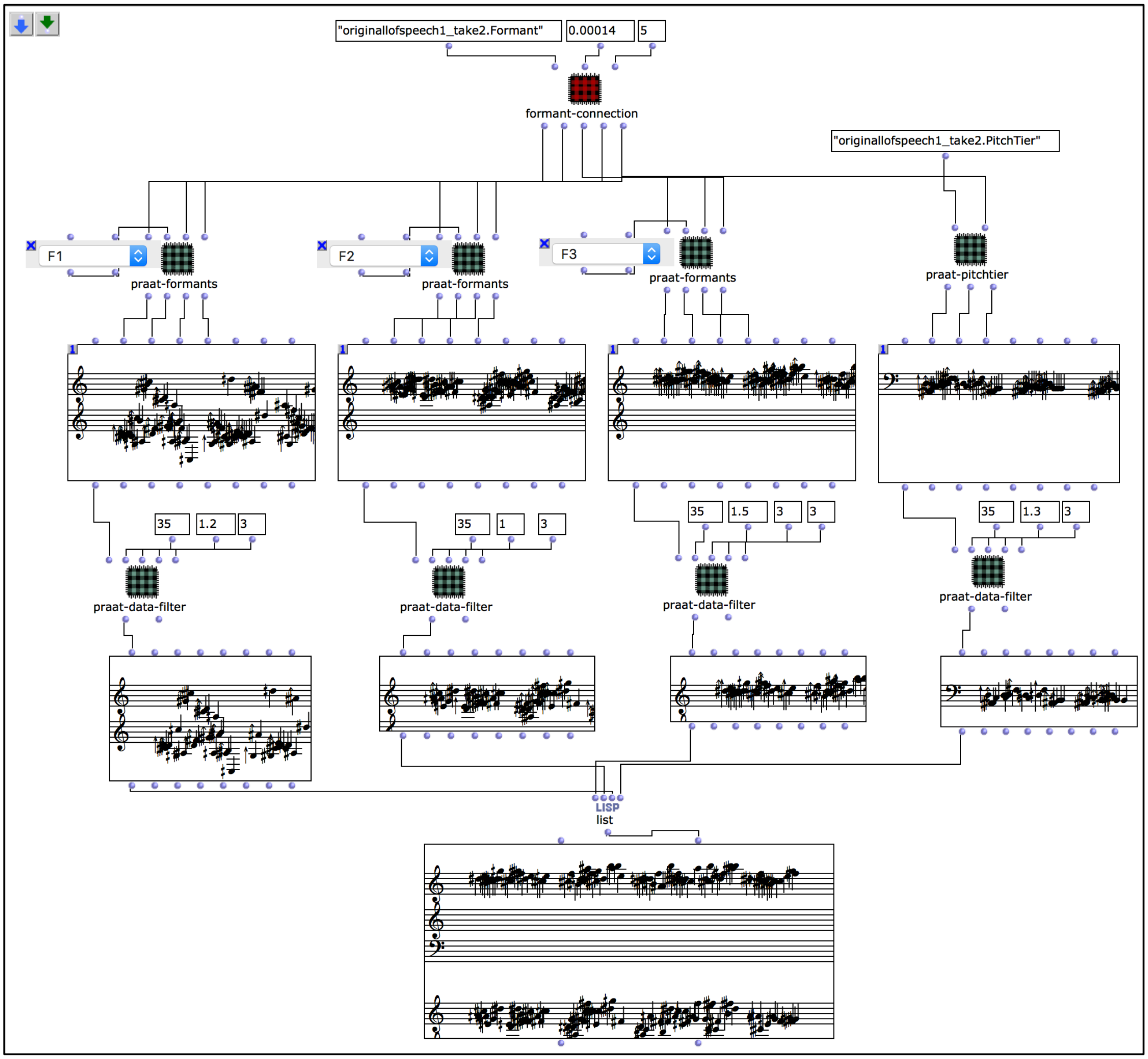



**Figure 19: My main OpenMusic phonetics environment which contains the patches *praat-formants*, *praat-pitchtier* and *praat-data-filter***

# Appendix B

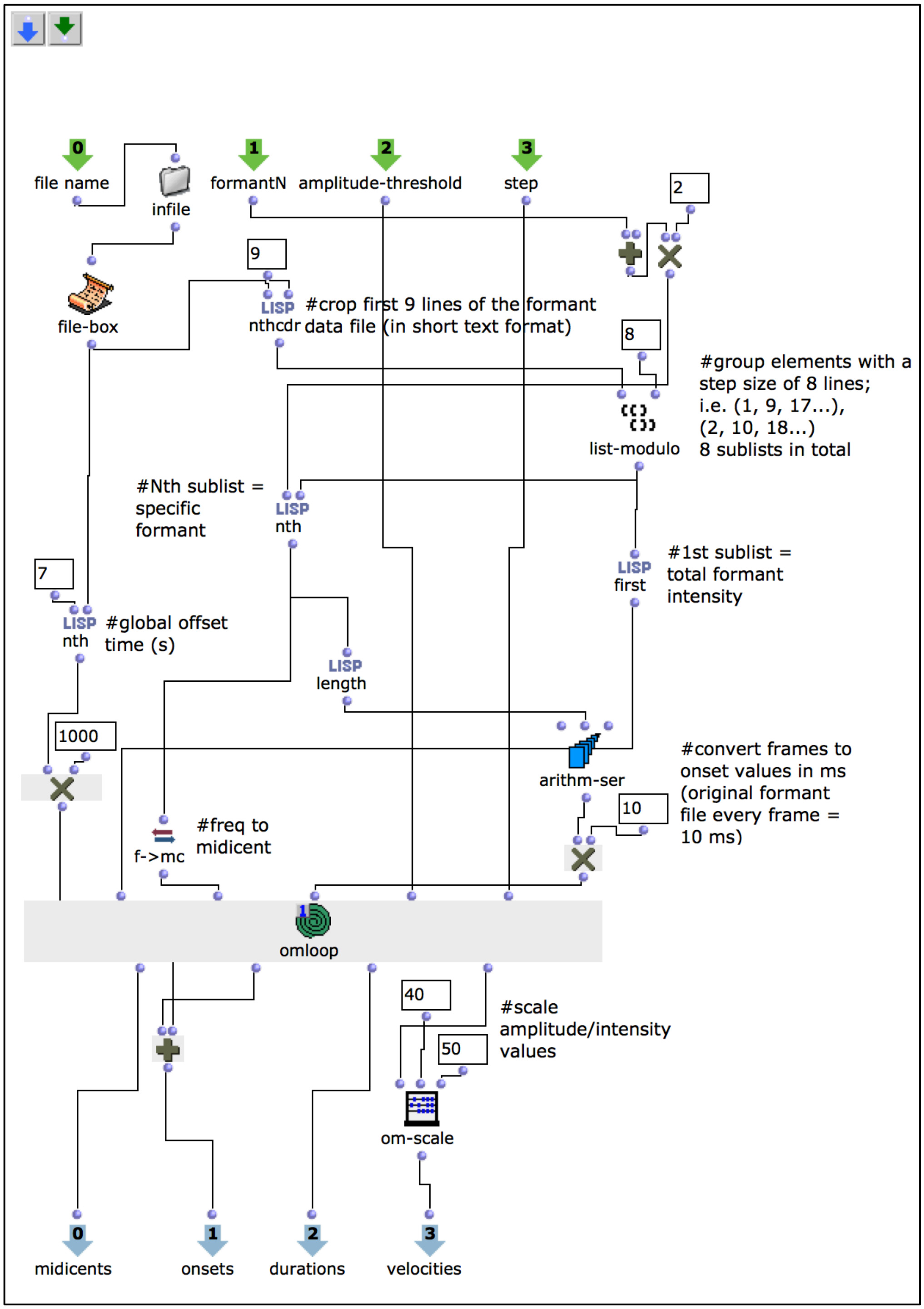



**Figure 20:** ***praat-formants*****, my OpenMusic patch for importing Praat formant data**

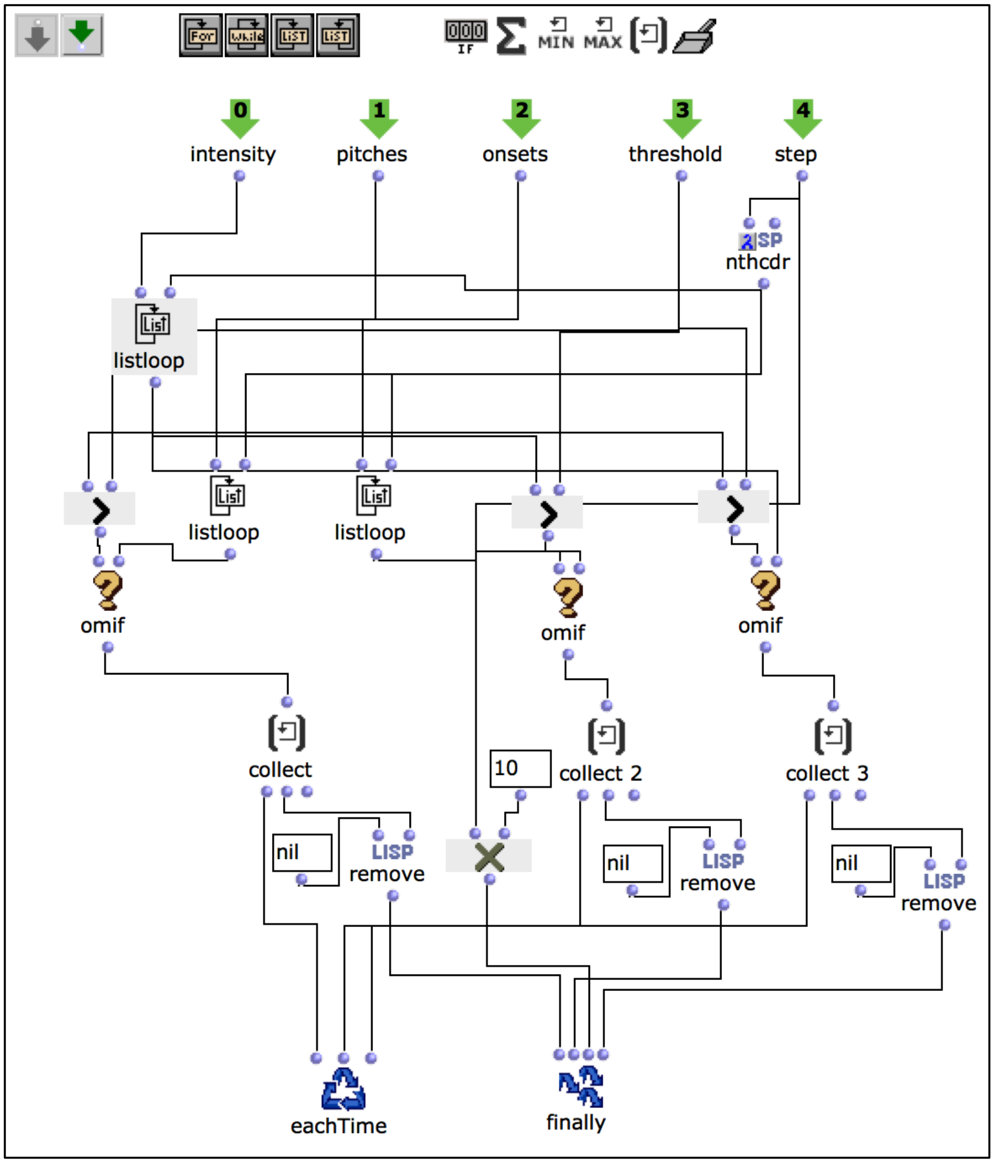



**Figure 21:** ***praat-formants*** **(continued), contents of the omloop inside the patch**

# Appendix C

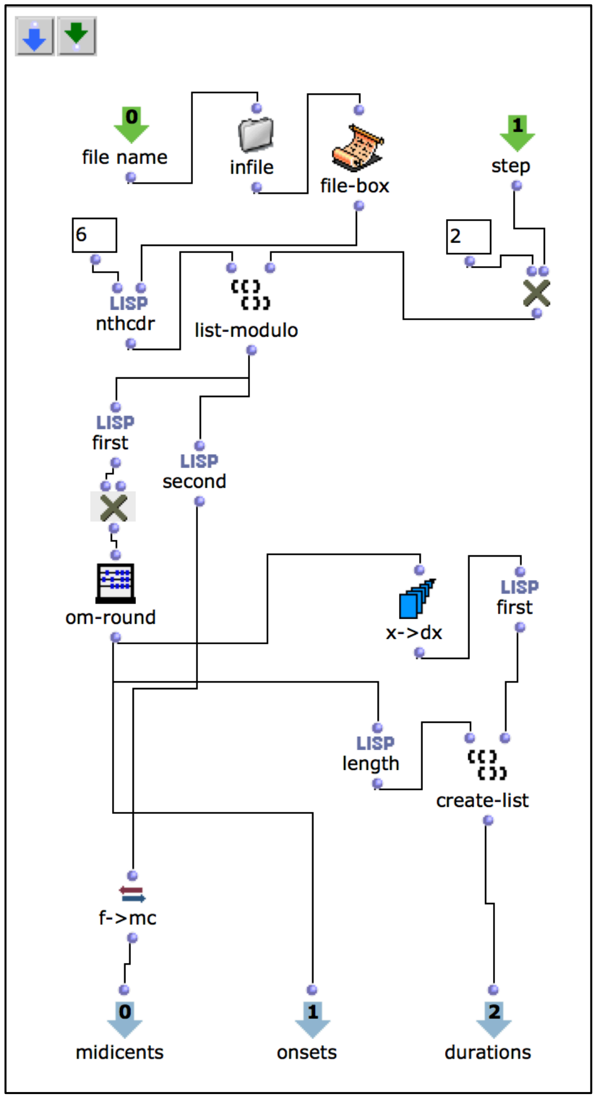



**Figure 22:** ***praat-pitchtier*, my OpenMusic patch for importing Praat fundamental frequency (F0) data in the PitchTier format**

# Appendix D

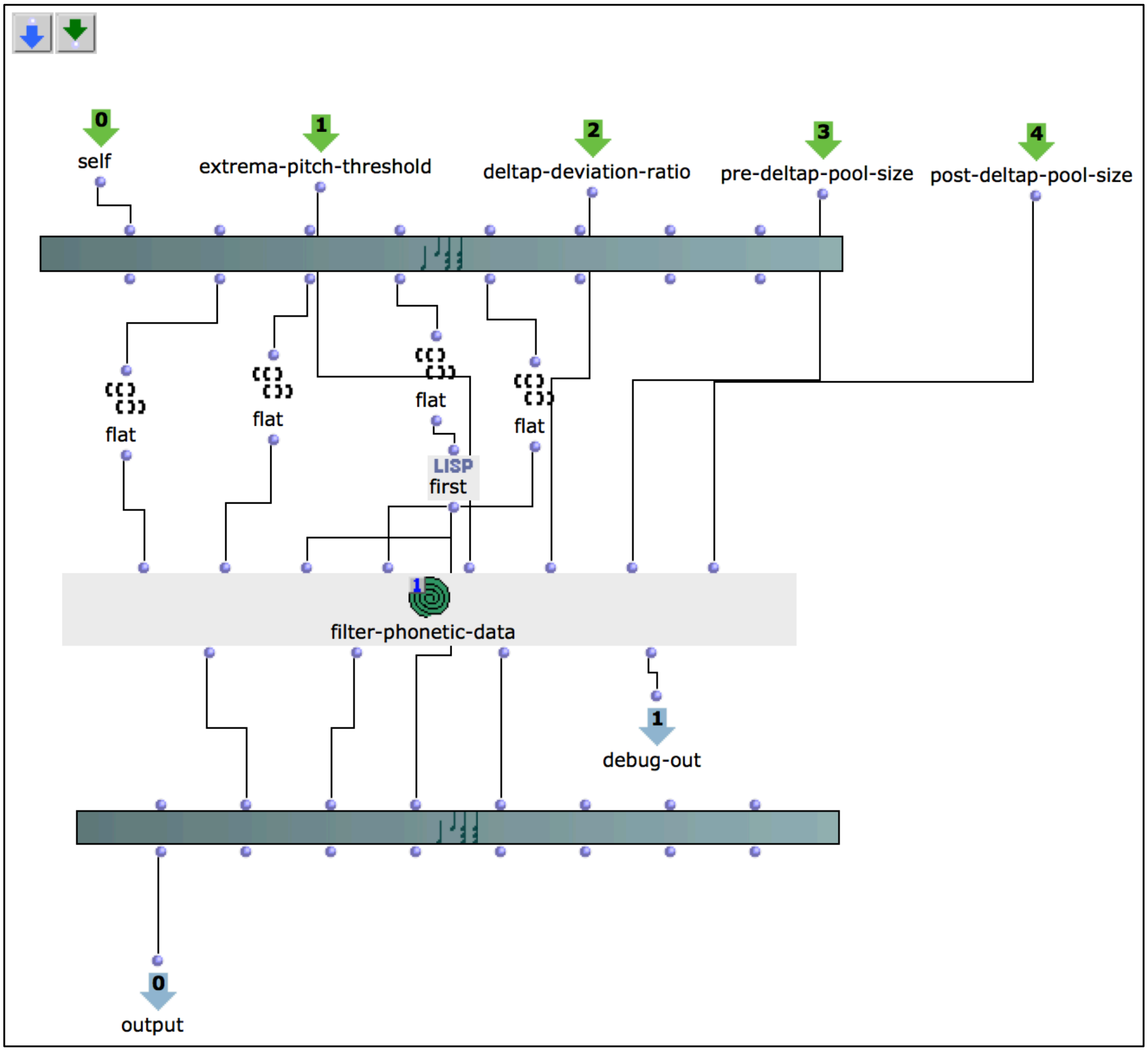



**Figure 23:** ***praat-data-filter*****, my OpenMusic patch for filtering and optimizing the formant and fundamental frequency data imported from Praat**

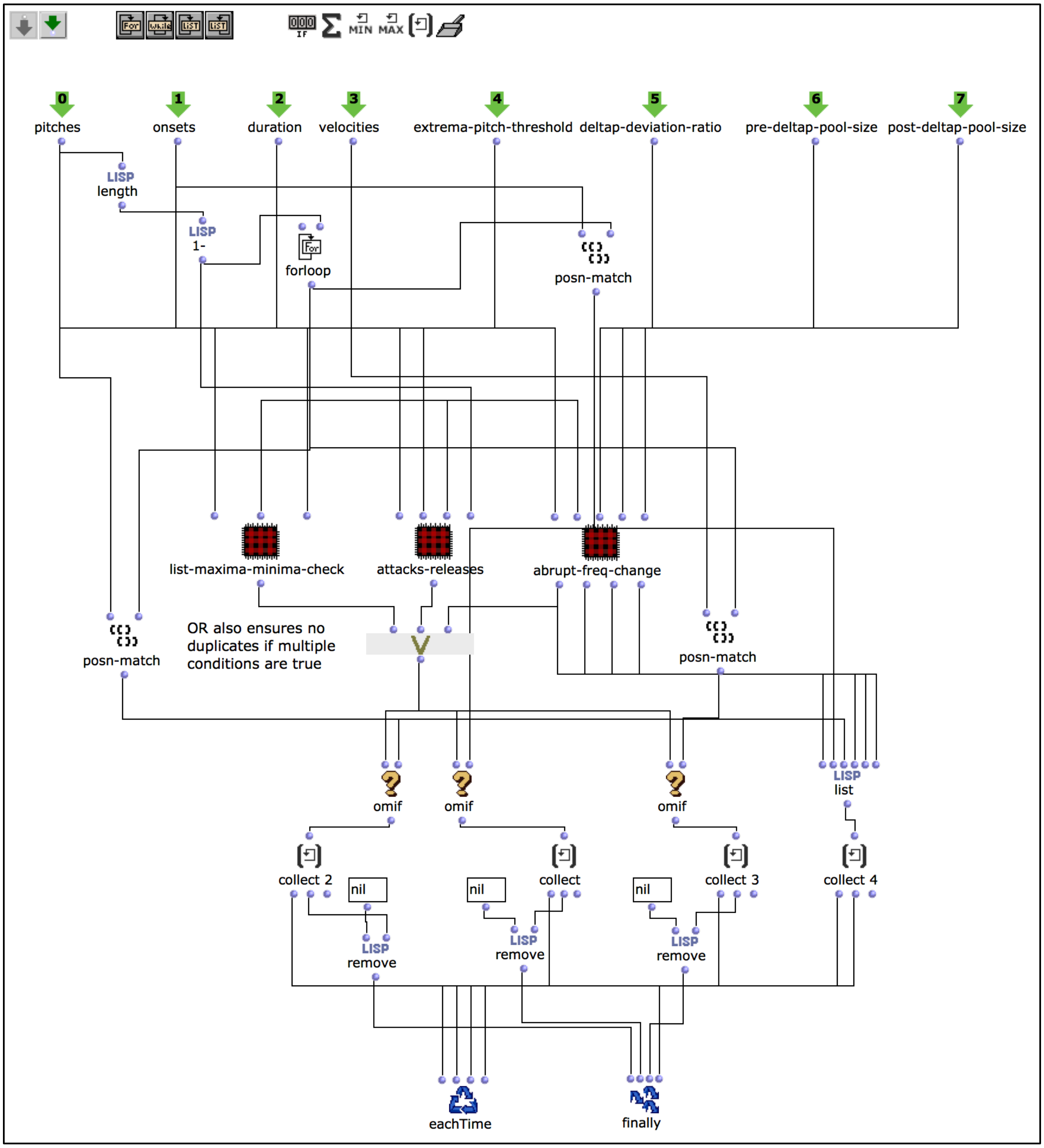



**Figure 24:** ***praat-data-filter*** **(continued), contents of the omloop** ***filter-phonetic-data***

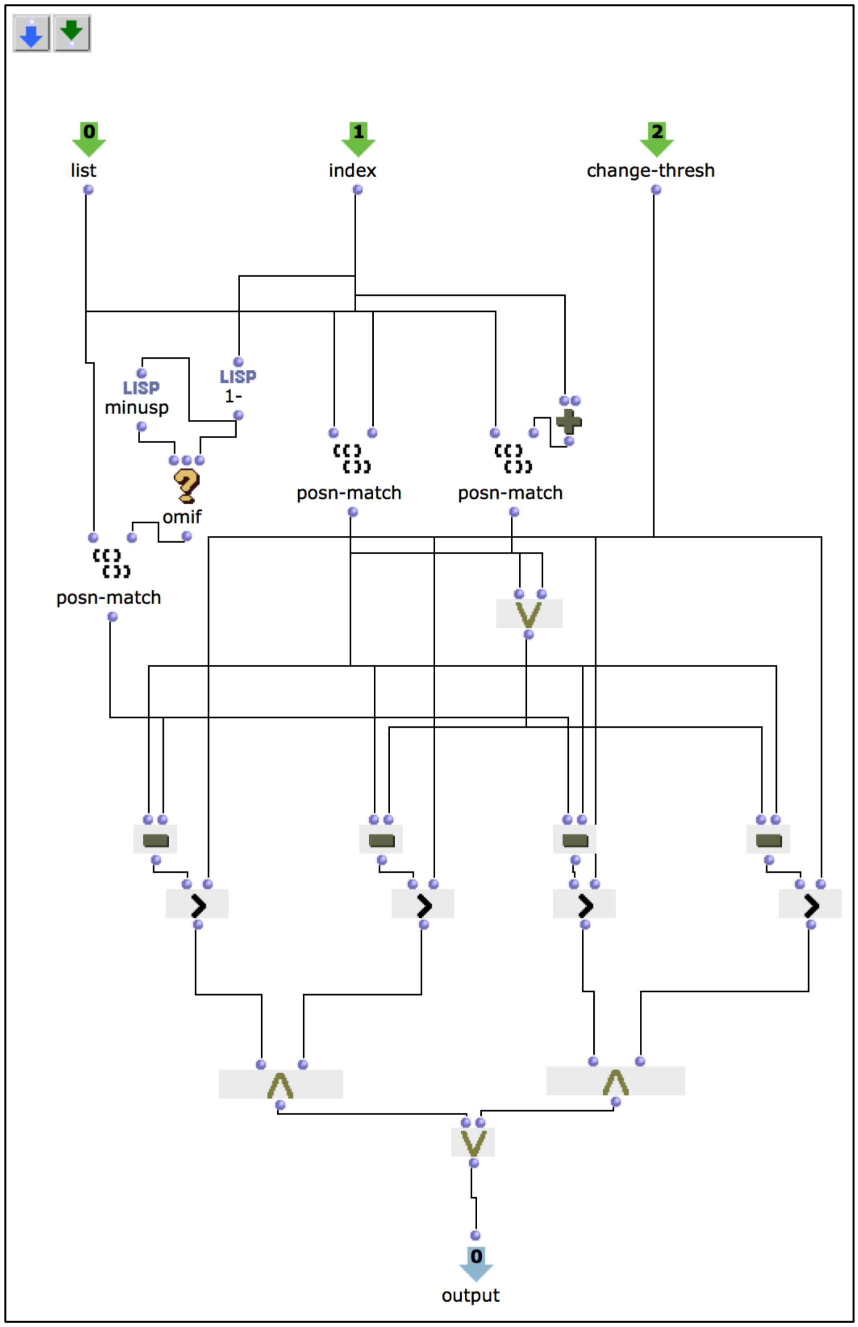



**Figure 25:** ***praat-data-filter*** **(continued), contents of the** ***list-maxima-minima-check*** **subpatch**

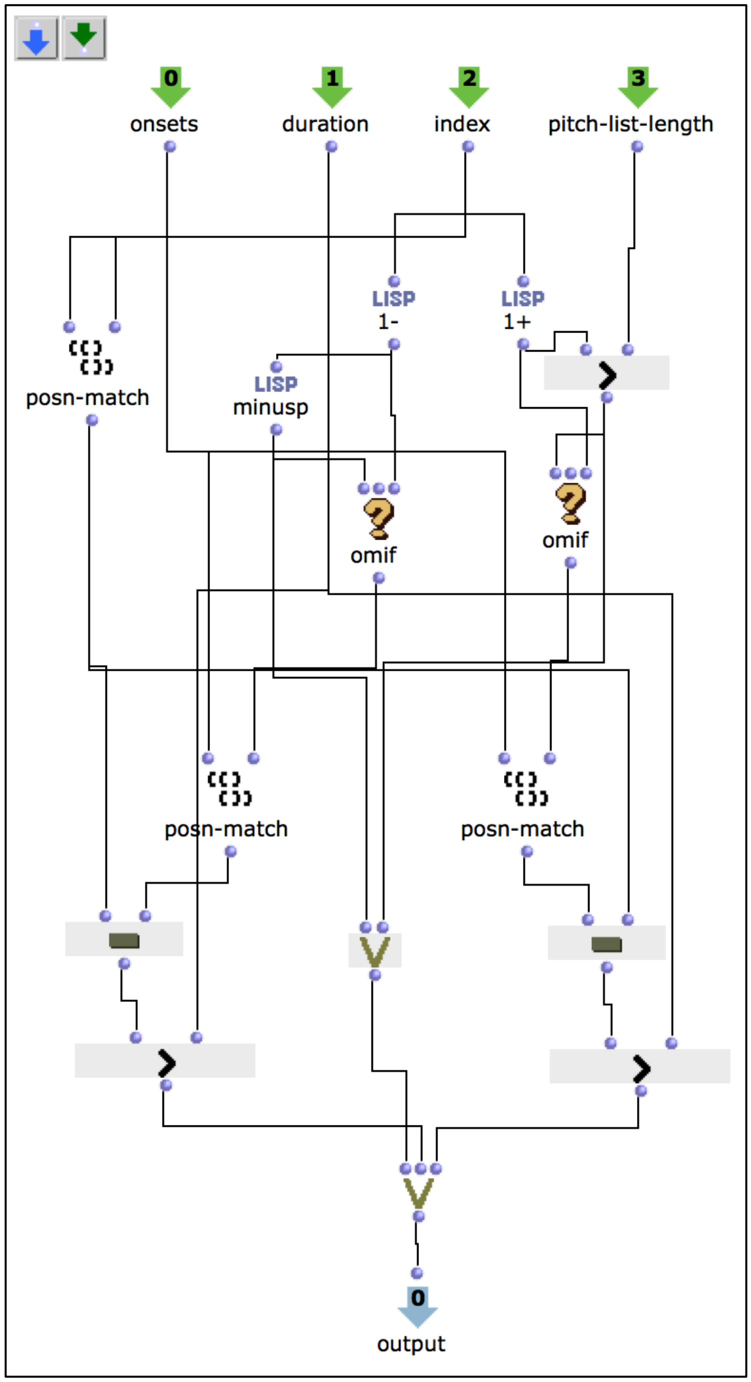



**Figure 26:** ***praat-data-filter*** **(continued), contents of the** ***attacks-releases*** **subpatch**

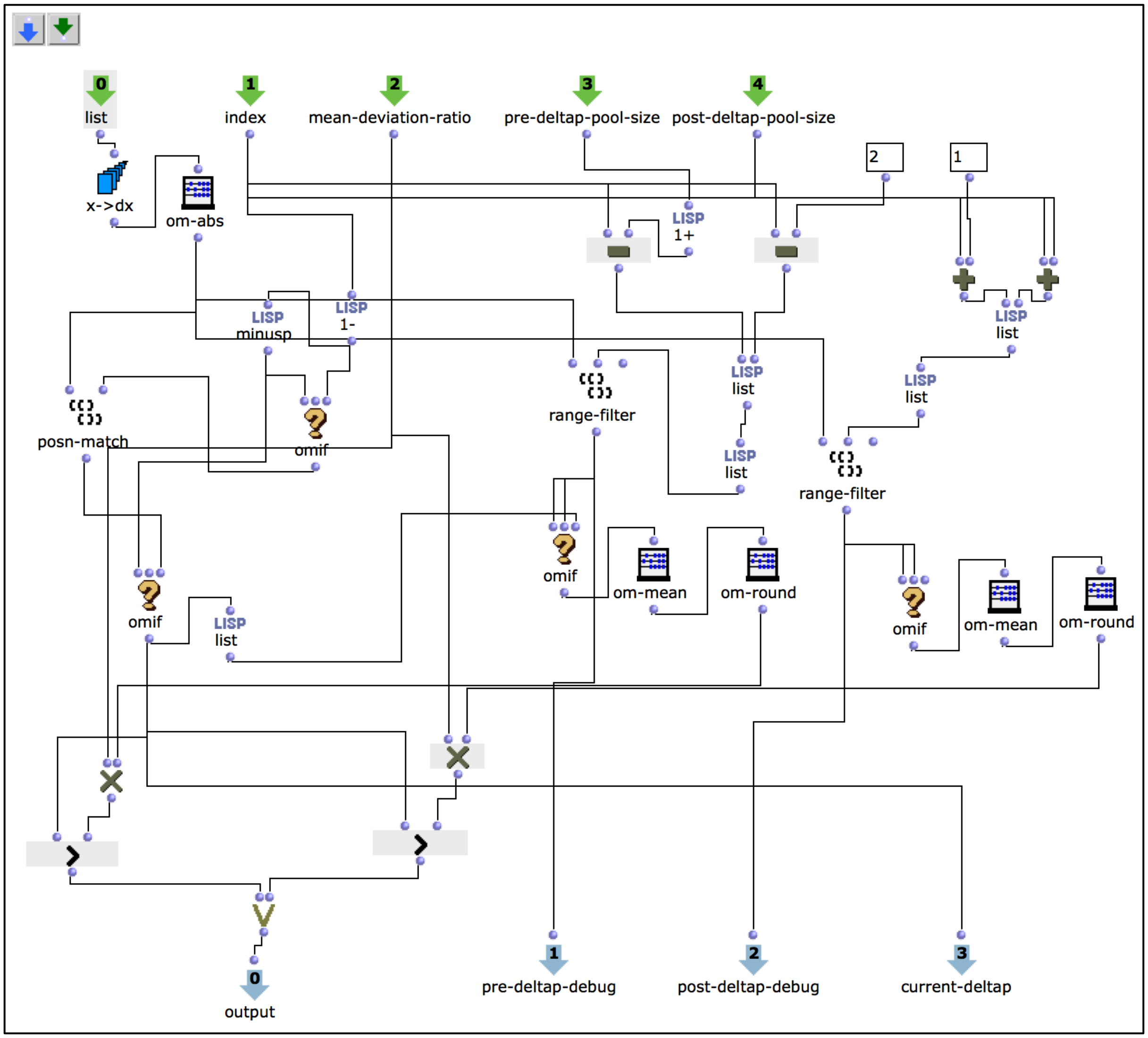



**Figure 27:** ***praat-data-filter*** **(continued), contents of the** ***abrupt-freq-change*** **subpatch**

# Appendix E

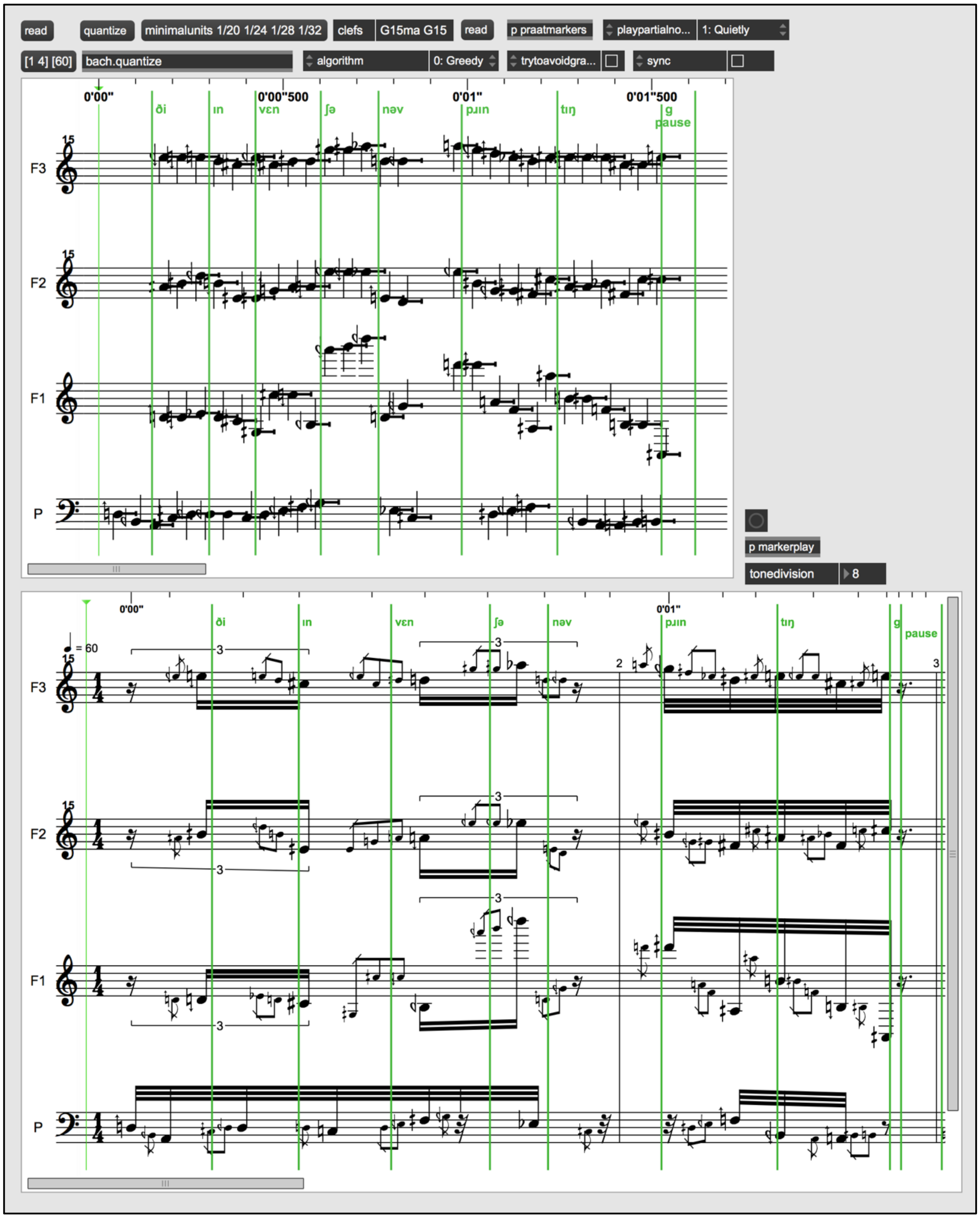



**Figure 28: My main bach/Max environment containing a *bach.roll* object with data imported from OpenMusic, *bach.score* displaying the quantized version, and my *praatmarker* subpatch used for extracting phoneme segments and labels from Praat annotation files and overlaying them on *bach.roll* or *bach.score* objects as markers**

# Music Score A: *ŠÀ {karāz}* (2017, rev. 2021)

ONUR YILDIRIM

# ŠÀ {karāz}

for large ensemble

## Instrumentation

Alto Flute
Clarinet in B♭
Bassoon

Trumpet in B♭ (+ wa-wa mute)
Trombone (+ plunger)

Percussion: snare drum, 2 bongos, 3 temple blocks (plastic "granite block" type), kutu wapa, 2 egg shakers (different sizes), reibestock

Piano

Violin 1
Violin 2
Viola
Cello
Double bass

The score is notated in C except double bass, which is written in its usual octave transposition (along with the necessary octave clef).

## Program Notes

The acts of excavation, decipherment and reconstruction form the core of *ŠÀ {karāz}* ("innards, inner soul, entrails" in Hittite). Most of the musical material in the piece is based on phonetic analyses of hypothetical pronunciations of phrases from cuneiform tablets written in several ancient languages.

## Performance Instructions

### General Instructions

➤ Grace notes, trills and tremolos should be played as fast as possible.

➤ All instruments should play without vibrato, unless indicated otherwise.

| | |
|---|---|
| 𝄲 | Quarter-tone higher |
| 𝄳 | Quarter-tone lower |
| ♯𝄲 | Three quarter-tones higher |
| 𝄳♭ | Three quarter-tones lower |

➤ Quarter-tones are employed as **tempered** intervals and should be played as precisely as possible.

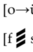
| | |
|---|---|
| ♯↑ ♮↑ ♭↑ | Slightly higher than ♯ ♮ ♭ (less than a ¼ tone) |
| ♯↓ ♮↓ ♭↓ | Slightly lower than ♯ ♮ ♭ (less than a ¼ tone) |

➤ **Woodwinds:** Where possible, appropriate quarter-tone fingerings should be used, rather than manipulating the pitch solely with the embouchure (except lip glissandi). "Slightly sharper/flatter" pitches can, in most cases, be produced with the embouchure, but fingerings are supplied wherever more pitch precision is required.

➤ **Trumpet:** The quarter tones on the trumpet are played via extending the 3rd valve slide and playing the pitches via the given valve combinations that include the 3rd valve. Whenever the slide is extended for quarter-tone pitches, it stays in that position until an indication to return it to its normal position.

| | |
|---|---|
|  | Crescendo dal niente |
|  | Decrescendo al niente |
| ⟶ | Gradual transition from one way of playing to another |
| n.v. | Non vibrato |
| vib. | Ordinary vibrato |
| m.v. | Molto vibrato |
|  | Used when a rapid transition between different kinds of vibrato is employed; try to match the line in your vibrato speed and amplitude. |

### All winds:

| | |
|---|---|
| ← | Inhale |
| → | Exhale |
| [k], [u], [hi] | Use the indicated vowel/consonant phonemes or syllables in the brackets to articulate the attack. If the syllable structure is [consonant + vowel] (such as [ta], [shi], etc.), attack the note with the consonant and sustain it with the vowel. In case of a fricative consonant (without a vowel at the end) such as [f], [s], [sh] (=ʃ), [th] (=θ), or [x] (=χ, kh), as well as the resonant [r], don't just articulate the beginning of the note, but keep sustaining the consonant phoneme throughout the indicated duration (e.g. [fffffff], [sssssss]). |
| [f(i)], [s(ü)] | Simultaneously sustain the fricative consonant and the vowel in parentheses |
| [o→ü] | Gradually move from one phoneme to another over the given duration. |
| [f  s] | Tremolo between the indicated phonemes (unmeasured) |

➤ The vowels—whether they are by themselves or combined with a consonant—should never be voiced (i.e. vocal cords shouldn't vibrate); they are always (forcefully) whispered.

➤ All wind instruments, except **Trombone**, should use the necessary fingerings for the indicated pitches while pronouncing the phonemes. **Trombone** should use the low-mid register ad libitum during the spoken actions.

Here is a pronunciation guide for the unusual phonemes used in the piece, along with examples:

[ə] = IPA: /ə/ — English *father*

[ü] = IPA: /y/ — German *Tür*, French *lune*, Turkish ***üç***

[l'a] = IPA /lʲa/ — "Palatal l", like Russian *лягушка*, Turkish ***lâf***

[k'] = IPA: /k'/ — Like [k], but with a sort abrupt pause (i.e. glottal stop) before the following vowel

[r] = IPA: /r/ — "Rolled r", like the Italian or Russian pronunciations; never like the English, German or French varieties

[th] = IPA: /θ/ — Modern Greek *θάλασσα*, English *thick*

[ts] = IPA: /t͡sʰ/ — German *Zeit*, English *cats*

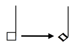
[ts'] = IPA: /t͡s'/ — Like [ts], but with a short, abrupt pause (i.e. glottal stop before the following vowel

[w] = IPA: /w̥/—like English *water*, but voiceless

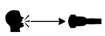
[x] = IPA: /χ/ — German *doch*, Dutch *goed*

[xw] = IPA: /χʷ/ — Simultaneously pronounce [x] and [w]; similar to the pronunciation of ***wh***- (/hw/) in some American English dialects

### Woodwinds

| | |
|---|---|
|  | ”Toneless”—predominance of breath sound on the indicated pitches—**very faint or no pitch**<br>* This notehead is also used whenever a phoneme/syllable is spoken into the instrument |
|  | Half breath sound, half normal tone |
|  | Transition between different mixtures of air/normal tone |
| W.T. | Whistle tones; follow the given contour as closely as possible |

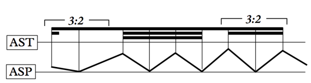
### Brass

| | |
|---|---|
|  | ”Toneless”—breath sound with the **mouthpiece on** (usually accompanied by a coloring vowel and/or consonant)<br>➤ **Trumpet** should always use the necessary fingerings for the indicated fundamentals. **Trombone** should use any low-mid fundamental fingering/position ad libitum and realize any given contour/register by using the indicating vowel(s). |
|  | Start blowing 3 cm away from the mouthpiece and gradually come closer. |
| ○ /⊕/+ | **Trumpet and Trombone:** Plunger/wa-wa mute – open / half-open / closed |
|  | Tremolo between open and closed (unmeasured) |

### Strings

| | |
|---|---|
|  | On the bridge—most of the bow hair should be directly on the bridge, but some of it should still touch the respective string to create a faint pitch marginally affected by the fingered note. |
| ASP | Alto sul ponticello—literally next to the bridge |
| SP | Sul ponticello |
| ORD | Ordinario |
| ST | Sul tasto |
| AST | Alto sul tasto—well over the fingerboard; as close as possible to the left hand fingers |

**Bow position staff:**




A small upper staff is employed whenever more rhythmic and/or gestural precision is needed in regards to the bow position changes. The uppermost and lowermost positions on the staff occasionally change and should, therefore, be observed carefully.

➤ Whenever **saltando** bowing is used in conjunction with the bow pos. staff, dotted lines are used, instead of solid lines (e.g. Vln 1 mm. 5-6).

➤ **Important:** The orientation of the bow positions on the staff is reversed for **Cello** and **Double Bass** in accordance with the instruments' reverse direction.

| | |
|---|---|
| c.l.b. | Col legno battuto |
| c.l.t. | Col legno tratto |
| ½ legno/c.l.t. | The hair and wood of the bow should simultaneously touch the string |
|  ←crini ←legno | The hair of the of the bow should be positioned behind the bridge while the wood bows the respective string at the ASP position; this should produce a metallic sound with a faint pitch content determined by the fingered note. |
| crini | Return to playing with the hair of the bow; cancels c.l.b. and c.l.t. |

flaut. Flautando

p. flaut Poco flautando

m. flaut Molto flautando—extremely light bow

n. flaut. Non flautando—normal bow pressure

➤ **Important**: While playing at any kind of flautando pressure, louder dynamics should be realized with faster bow speed, rather than applying more bow pressure.

 Overpressure of bow on string, scratch tone

½ o.p. Half over-pressure, premuto; it should never be vulgar like overpressure; a slight increase in the intensity of the note

 Return to normal bow pressure; cancels the over-pressure and half over-pressure signs

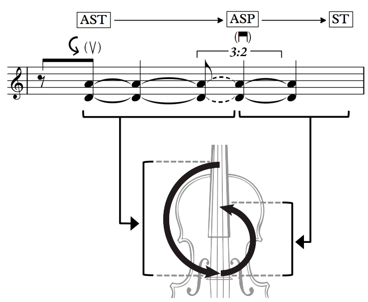
 Gradually increase/decrease bow pressure; the beginning and ending bow pressure levels are indicated (flaut → ½ o.p., etc.)

 Saltando

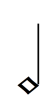
 Circular bowing

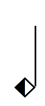
➤ The direction, duration and the radius of the circular movement are indicated by the ( ) / ( ) signs and the bow position labels (or the bow position staff for more complex rhythms).

**Example:**

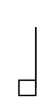



 "Toneless" bowing; dampen the indicated string/pitch so that no clear pitch is audible; mostly bow noise

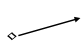
 Harmonic pressure; also used for non-harmonic points on the string

 Half-harmonic pressure

## Piano

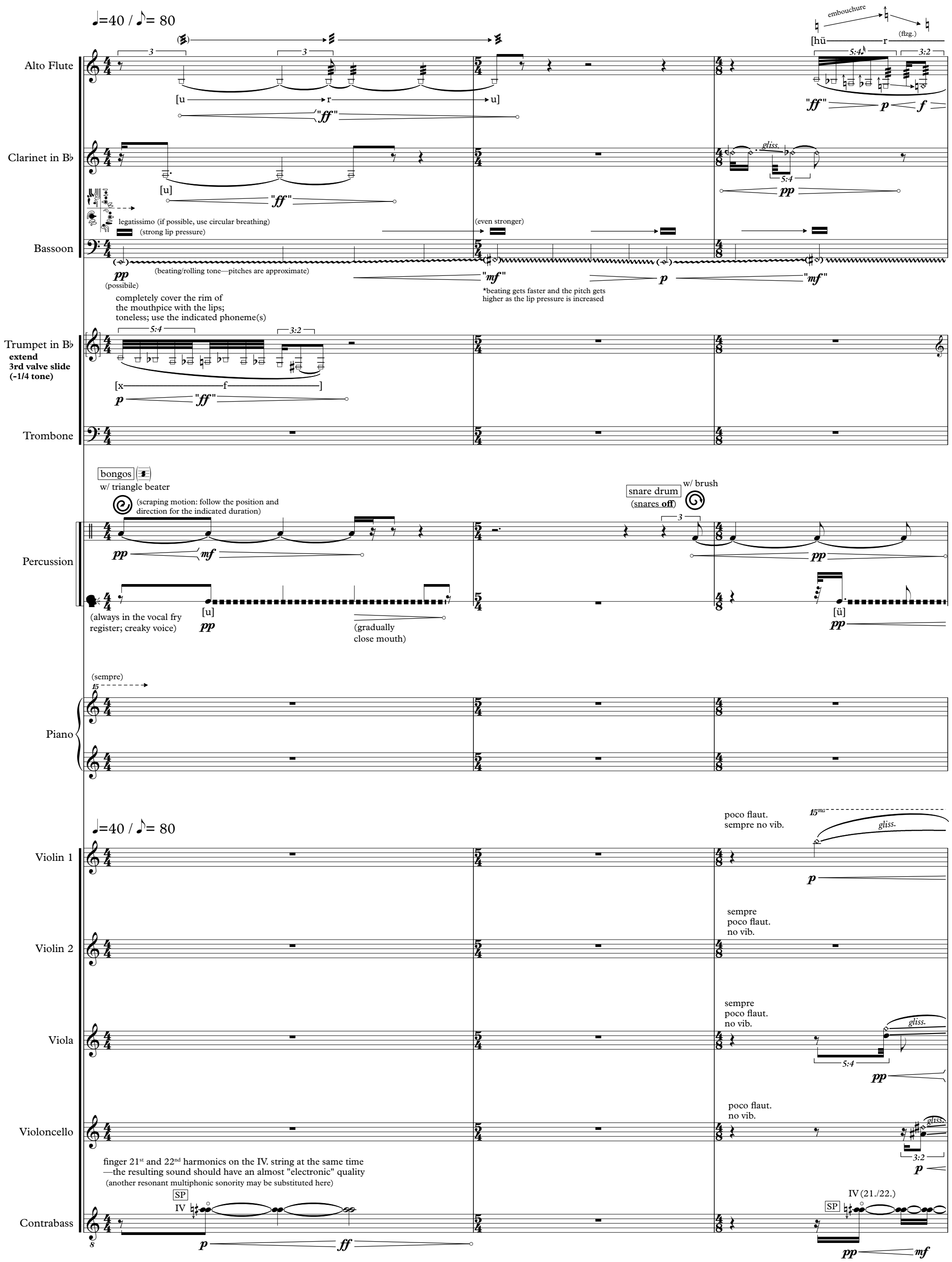
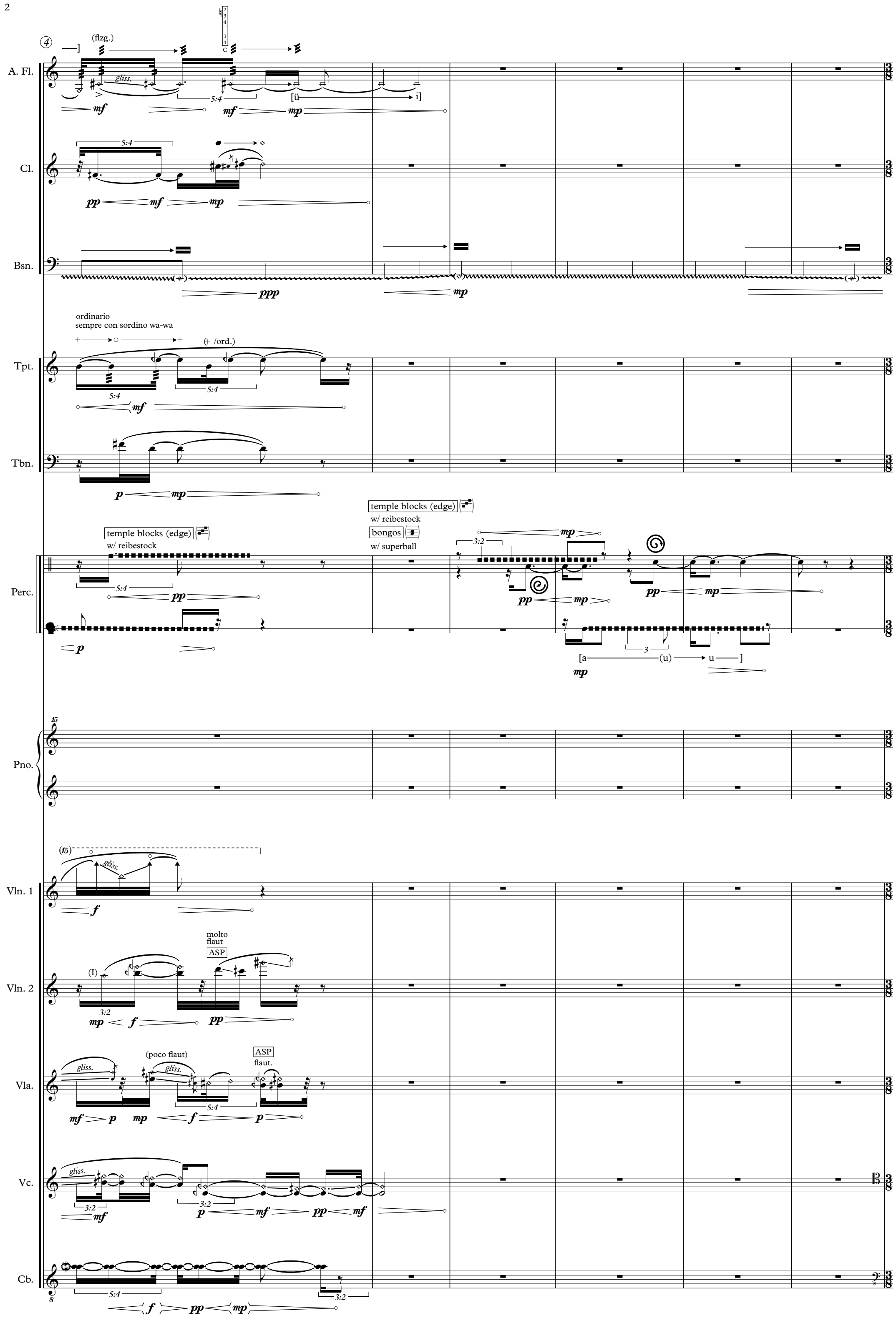
 Silently depress keys

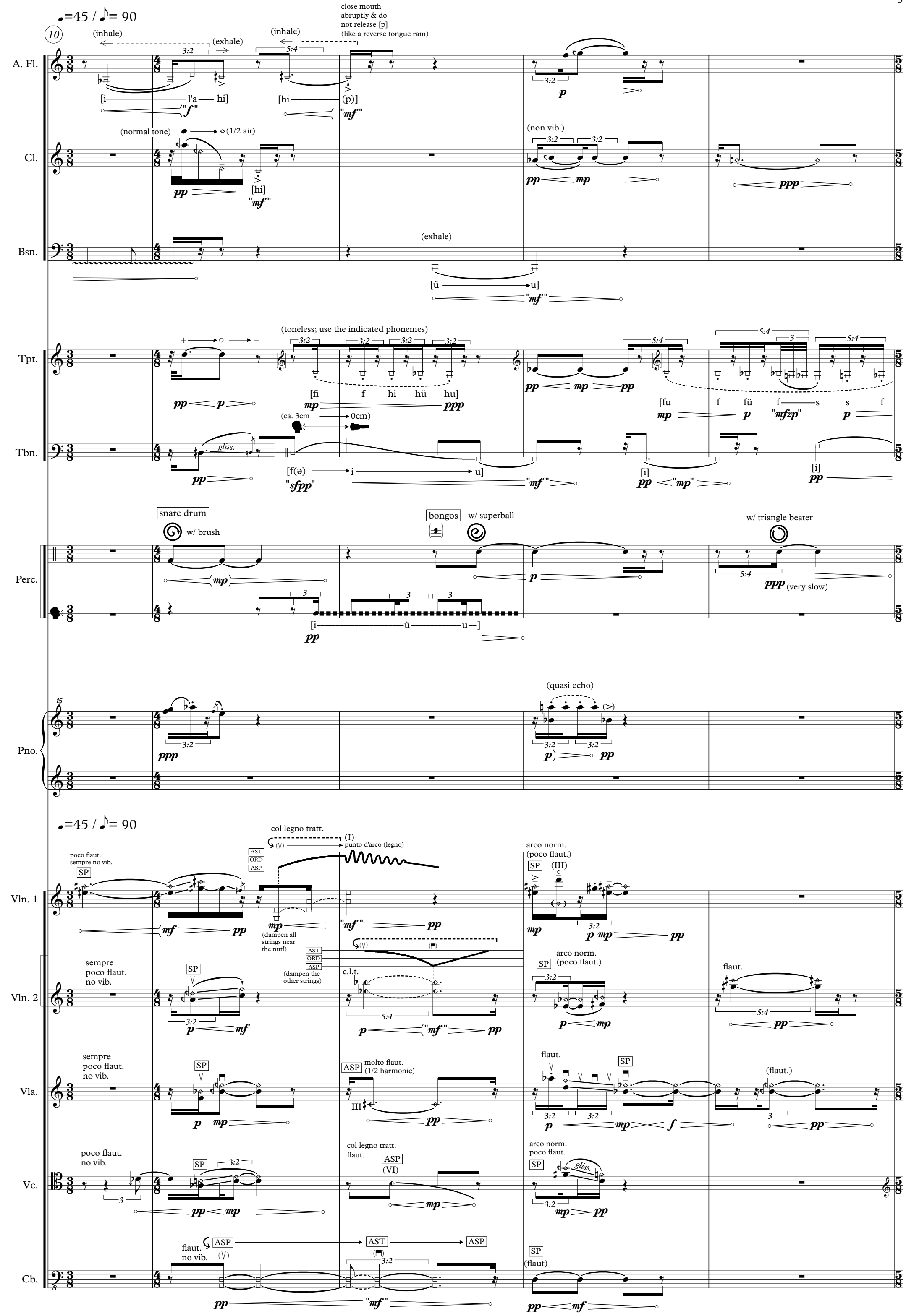
 Dampen the note inside the piano

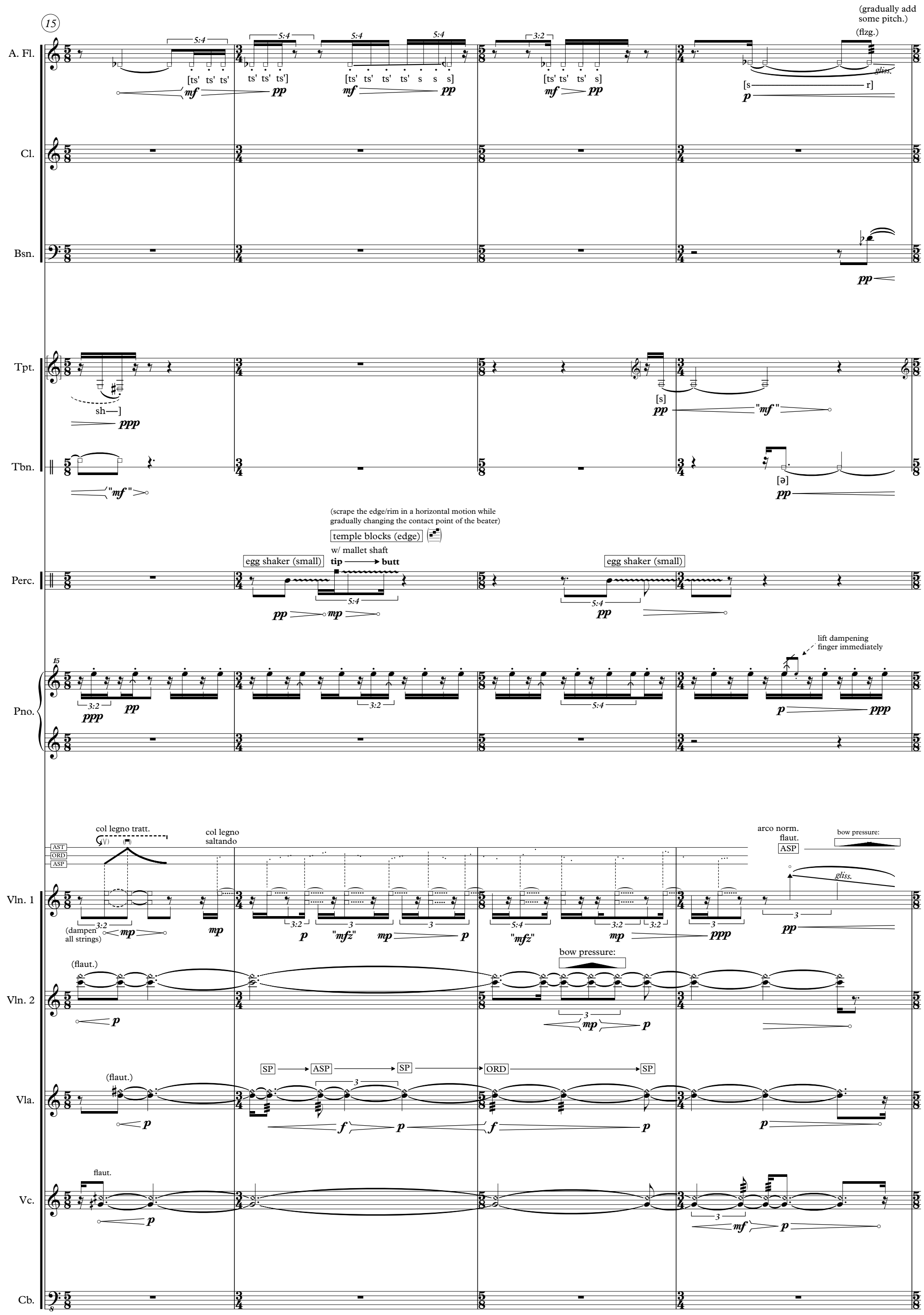
 Gradually slide dampening finger towards the hammers

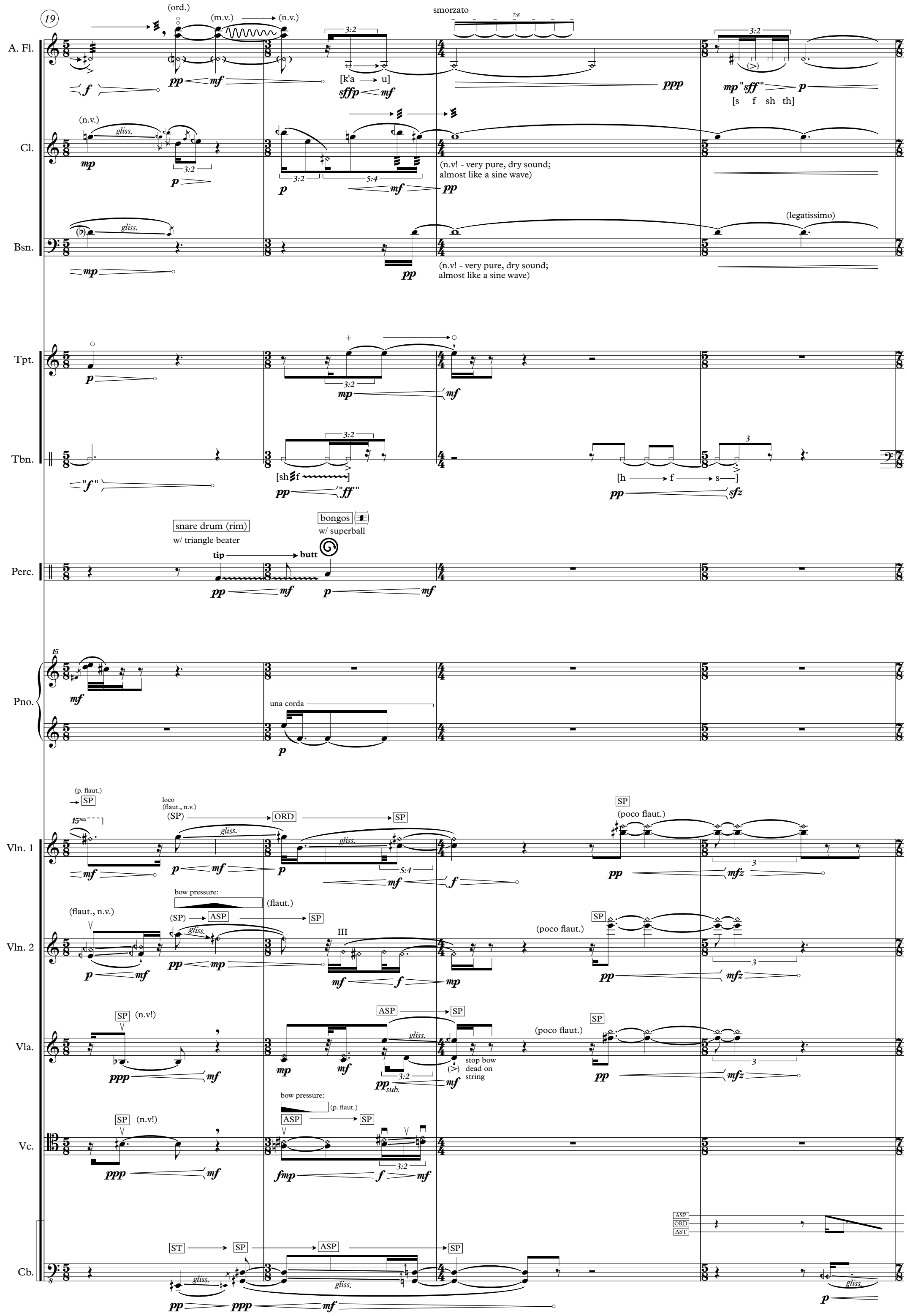
# ŠÀ {karāz}

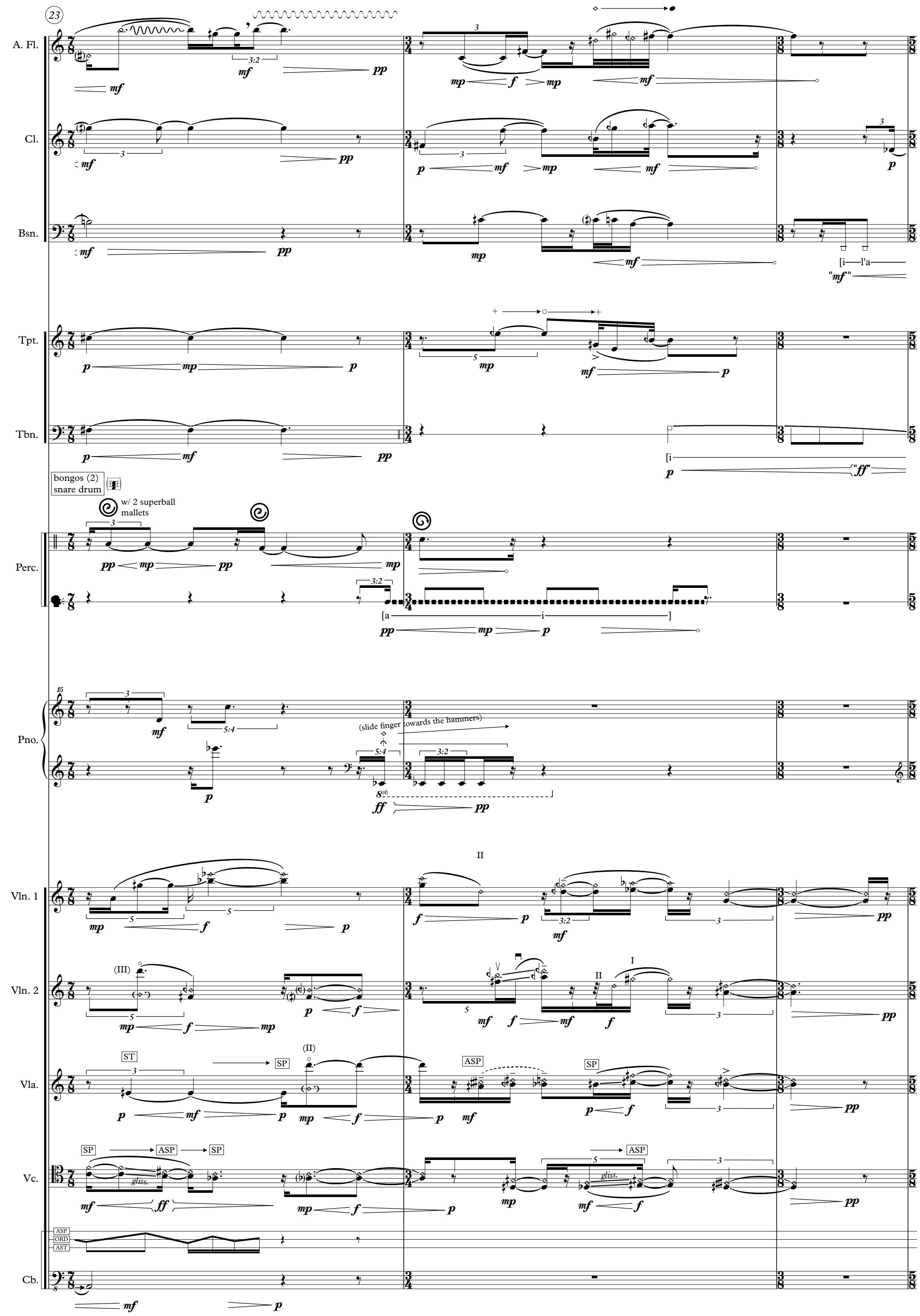
for large ensemble

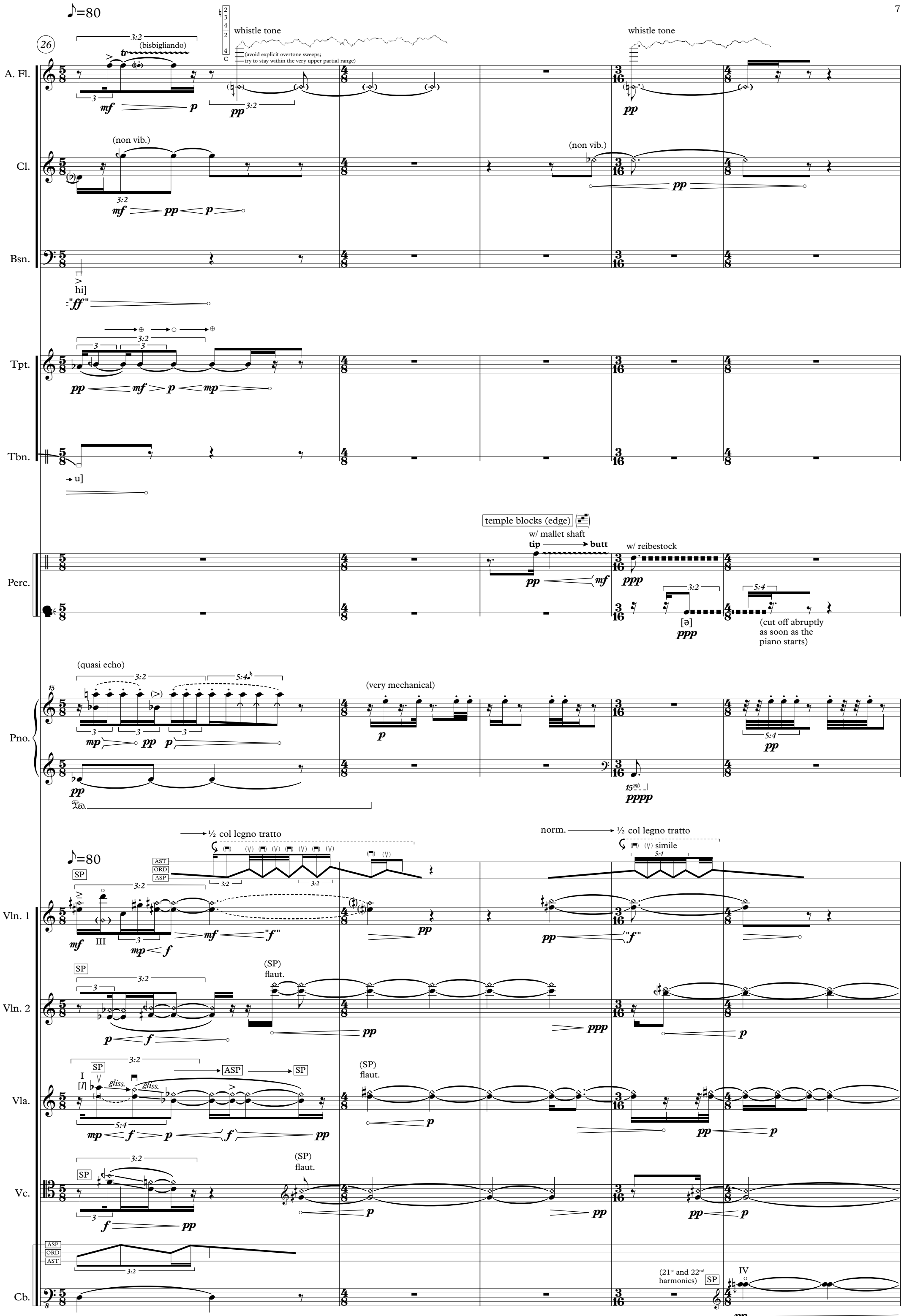
onur yıldırım (2017, rev. 2021)



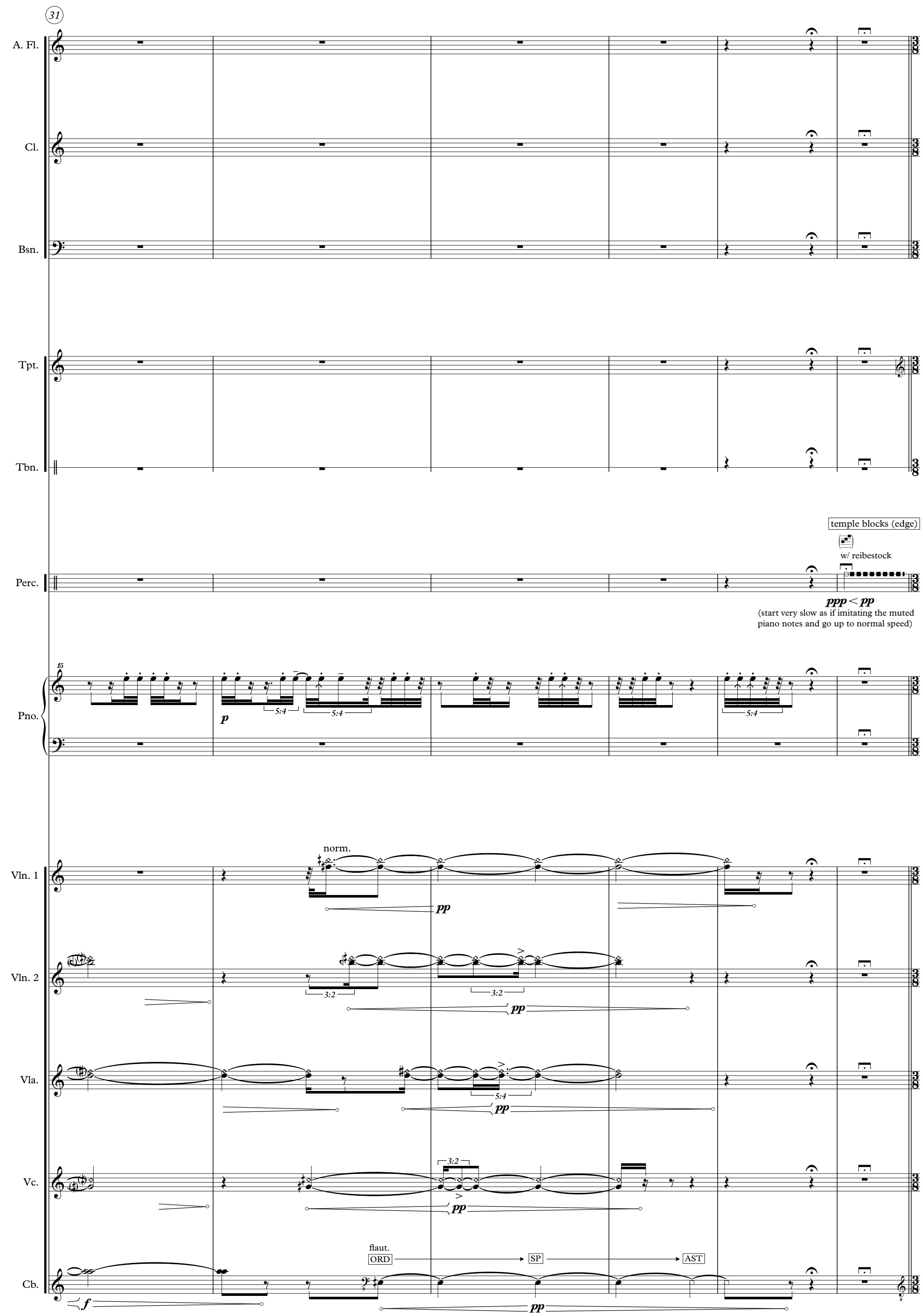
A. Fl.
(flzg.)
gliss.
[ü
i]
Cl.
Bsn.
ordinario
sempre con sordino wa-wa
(+ /ord.)
Tpt.
Tbn.
temple blocks (edge)
w/ reibestock
bongos
w/ superball
Perc.
[a
(u)
u
]
Pno.
Vln. 1
molto
flaut
ASP
Vln. 2
(poco flaut)
flaut.
Vla.
Vc.
Cb.


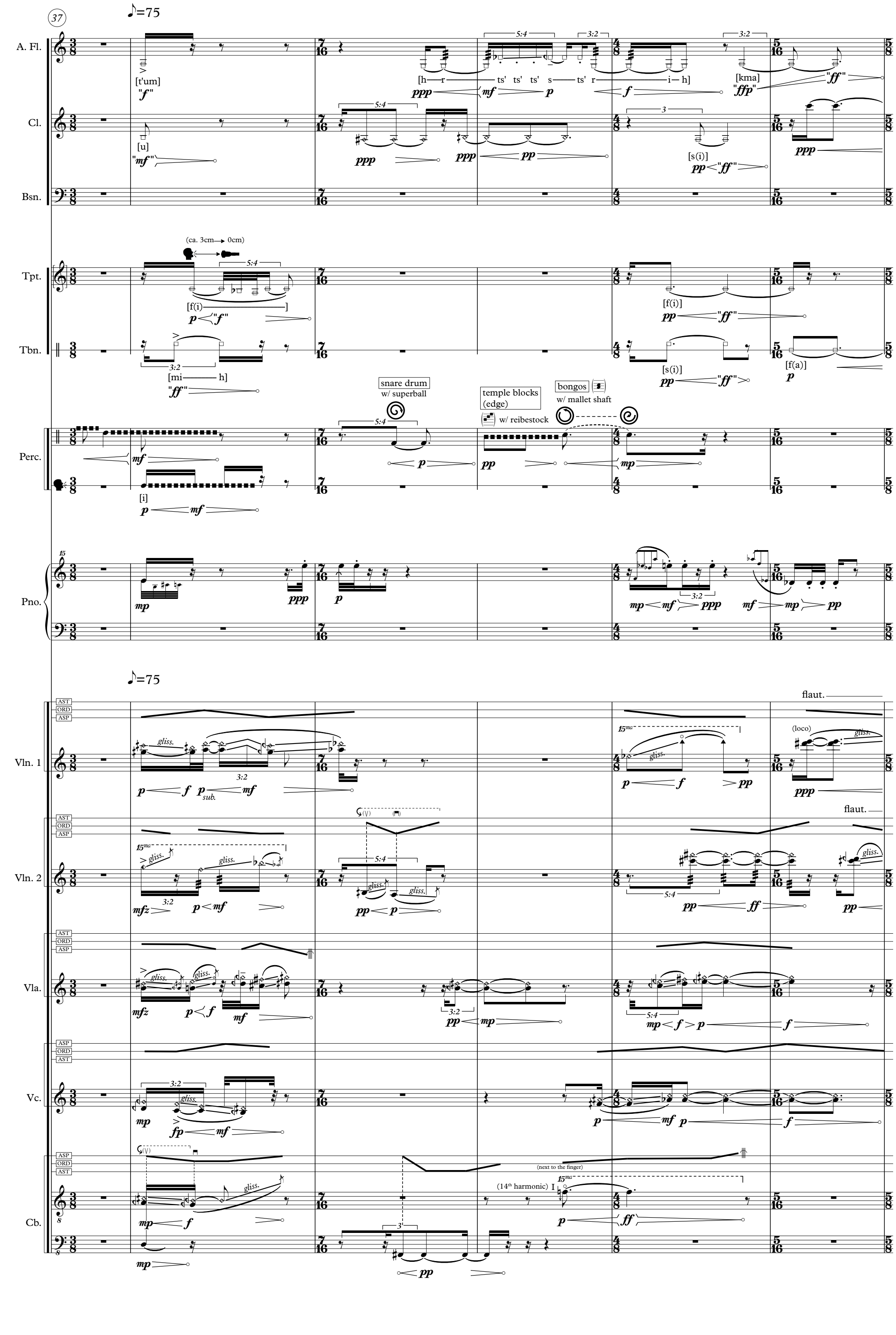
♩=45 / ♪= 90
10
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
(inhale)
(exhale)
close mouth
abruptly & do
not release [p]
(like a reverse tongue ram)
[i l'a hi]
[hi (p)]
(normal tone)
(1/2 air)
(non vib.)
[ü u]
(toneless; use the indicated phonemes)
[fi f hi hü hu]
[fu f fü f s s f]
(ca. 3cm 0cm)
[f(ə) i u]
[i]
gliss.
snare drum
w/ brush
bongos
w/ superball
w/ triangle beater
(very slow)
[i ü u]
(quasi echo)
poco flaut.
sempre no vib.
col legno tratt.
punto d'arco (legno)
AST
ORD
ASP
SP
(dampen all strings near the nut!)
(dampen the other strings)
c.l.t.
arco norm.
(poco flaut.)
sempre poco flaut. no vib.
flaut.
molto flaut. (1/2 harmonic)
(flaut.)
poco flaut. no vib.
col legno tratt. flaut.
arco norm. poco flaut.
flaut. no vib.
(flaut)


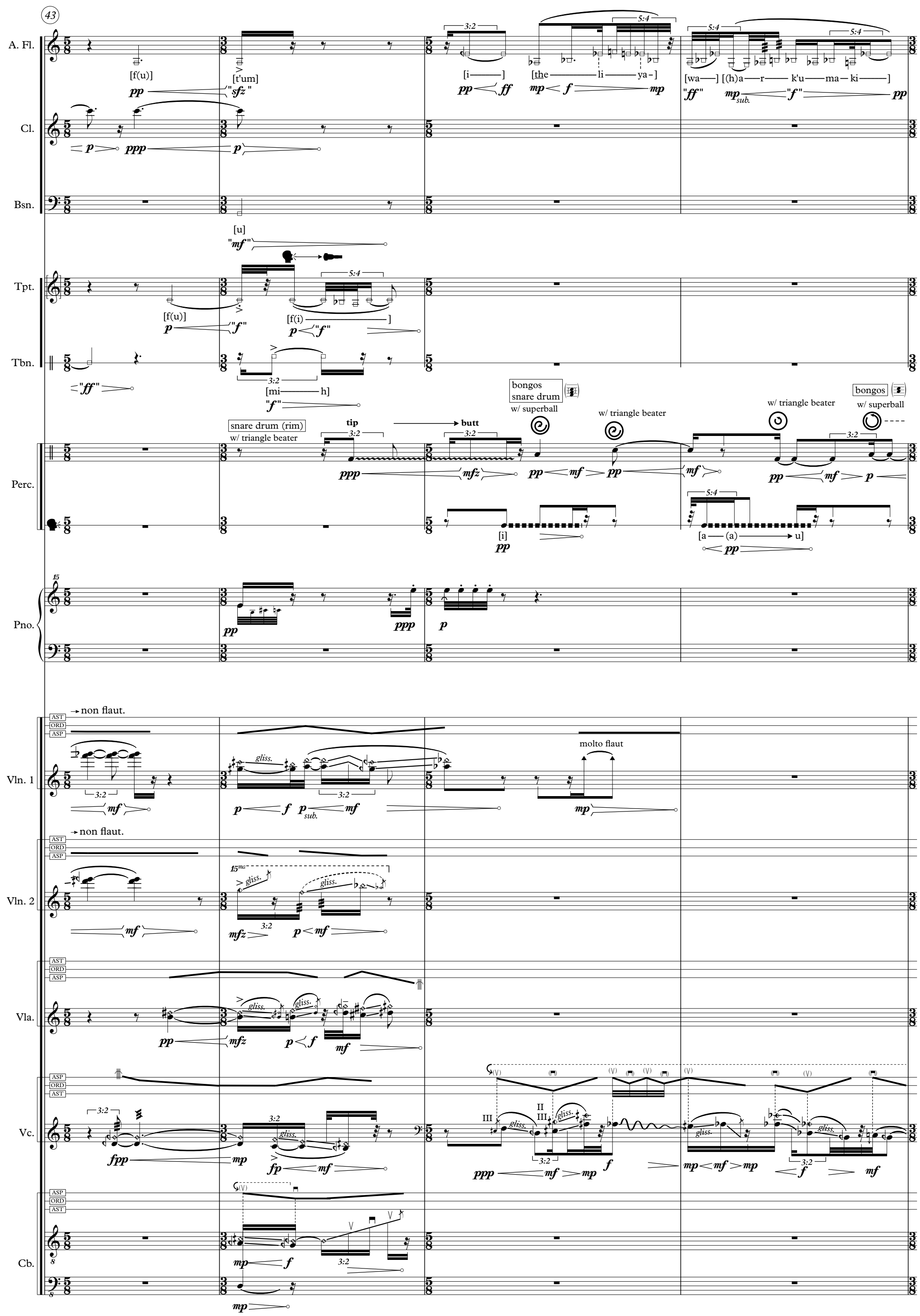
15
(gradually add some pitch.)
(flzg.)
A. Fl.
[ts' ts' ts'
ts' ts' ts']
[ts' ts' ts' ts' s s s]
[ts' ts' ts' s]
[s r]
gliss.
Cl.
Bsn.
Tpt.
sh—]
[s]
Tbn.
[ə]
(scrape the edge/rim in a horizontal motion while gradually changing the contact point of the beater)
temple blocks (edge)
egg shaker (small)
w/ mallet shaft
tip
butt
Perc.
lift dampening finger immediately
Pno.
col legno tratt.
col legno saltando
arco norm.
flaut.
bow pressure:
AST
ORD
ASP
Vln. 1
(dampen all strings)
(flaut.)
Vln. 2
SP
ASP
ORD
Vla.
flaut.
Vc.
Cb.


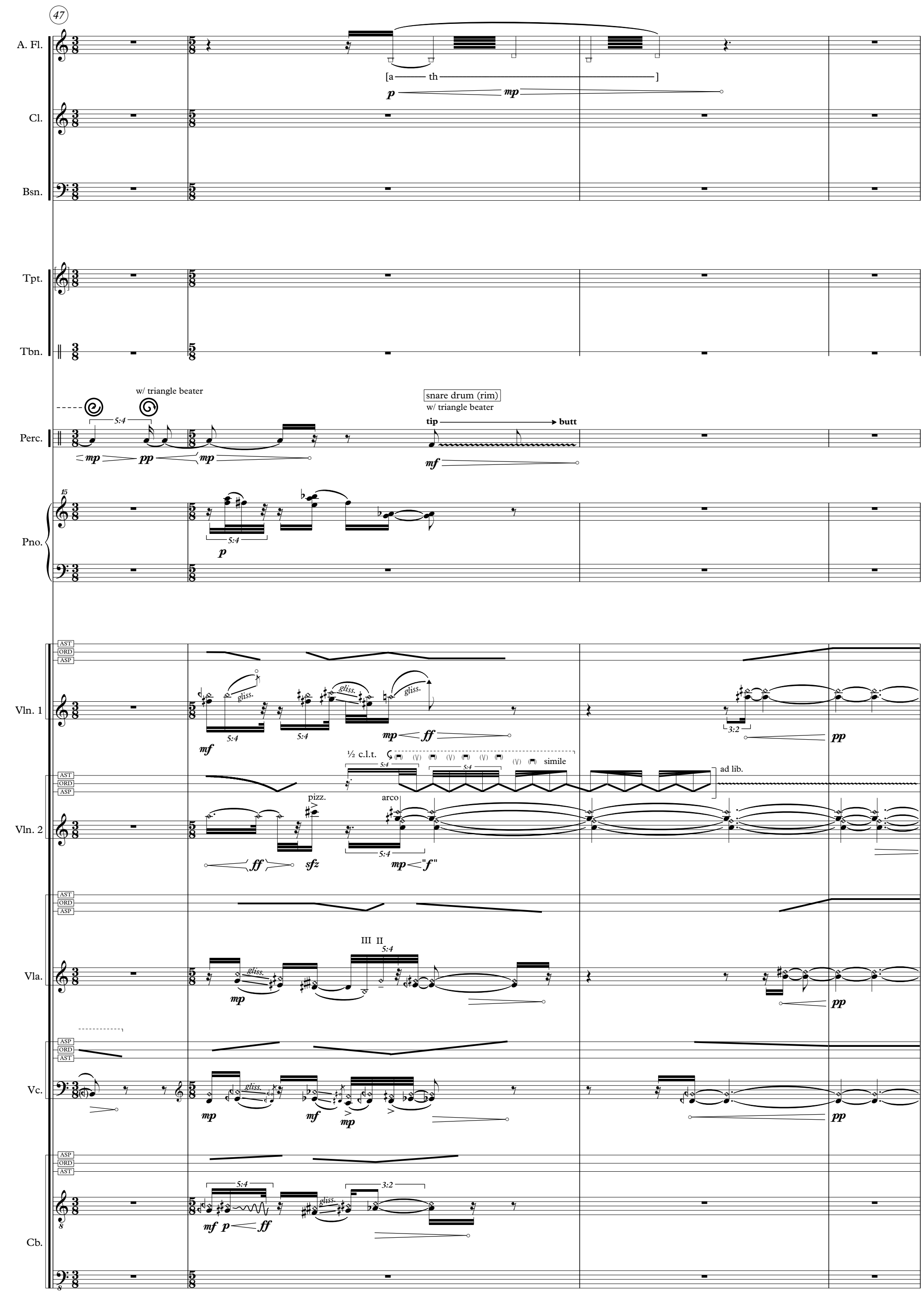
19
A. Fl.
(ord.)
(m.v.)
(n.v.)
3:2
[k'a u]
smorzato
7:8
[s f sh th]
Cl.
(n.v.)
gliss.
(n.v! - very pure, dry sound; almost like a sine wave)
Bsn.
(legatissimo)
Tpt.
Tbn.
[sh f]
[h f s]
snare drum (rim)
w/ triangle beater
tip
butt
bongos
w/ superball
Perc.
Pno.
una corda
Vln. 1
(p. flaut.)
SP
loco (flaut., n.v.)
(SP)
ORD
(poco flaut.)
Vln. 2
(flaut., n.v.)
bow pressure:
ASP
(flaut.)
Vla.
(n.v!)
stop bow dead on string
Vc.
(p. flaut.)
Cb.
ST
ORD
AST


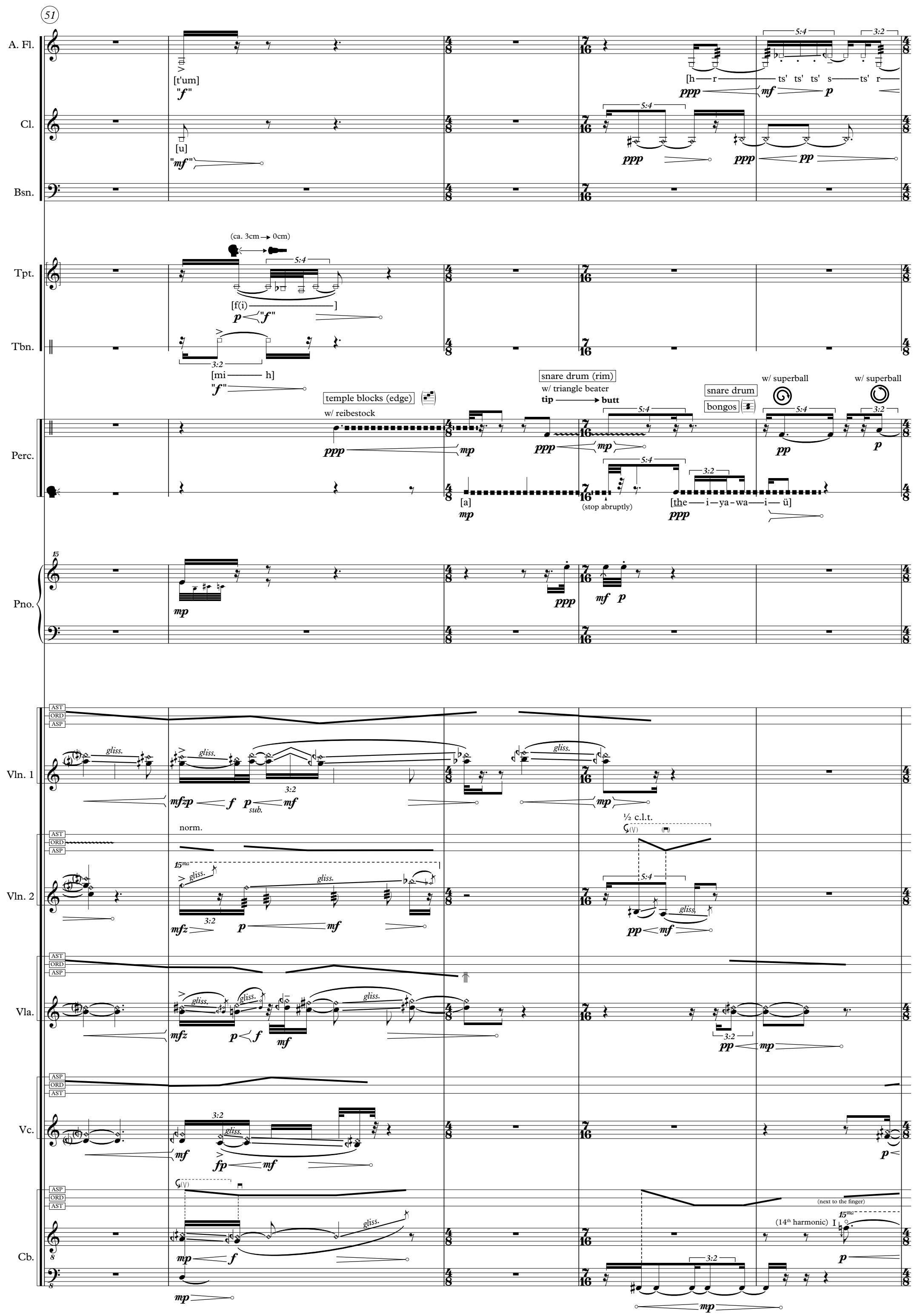
23
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
bongos (2)
snare drum
w/ 2 superball mallets
(slide finger towards the hammers)
ST
SP
ASP
ORD
AST
gliss.


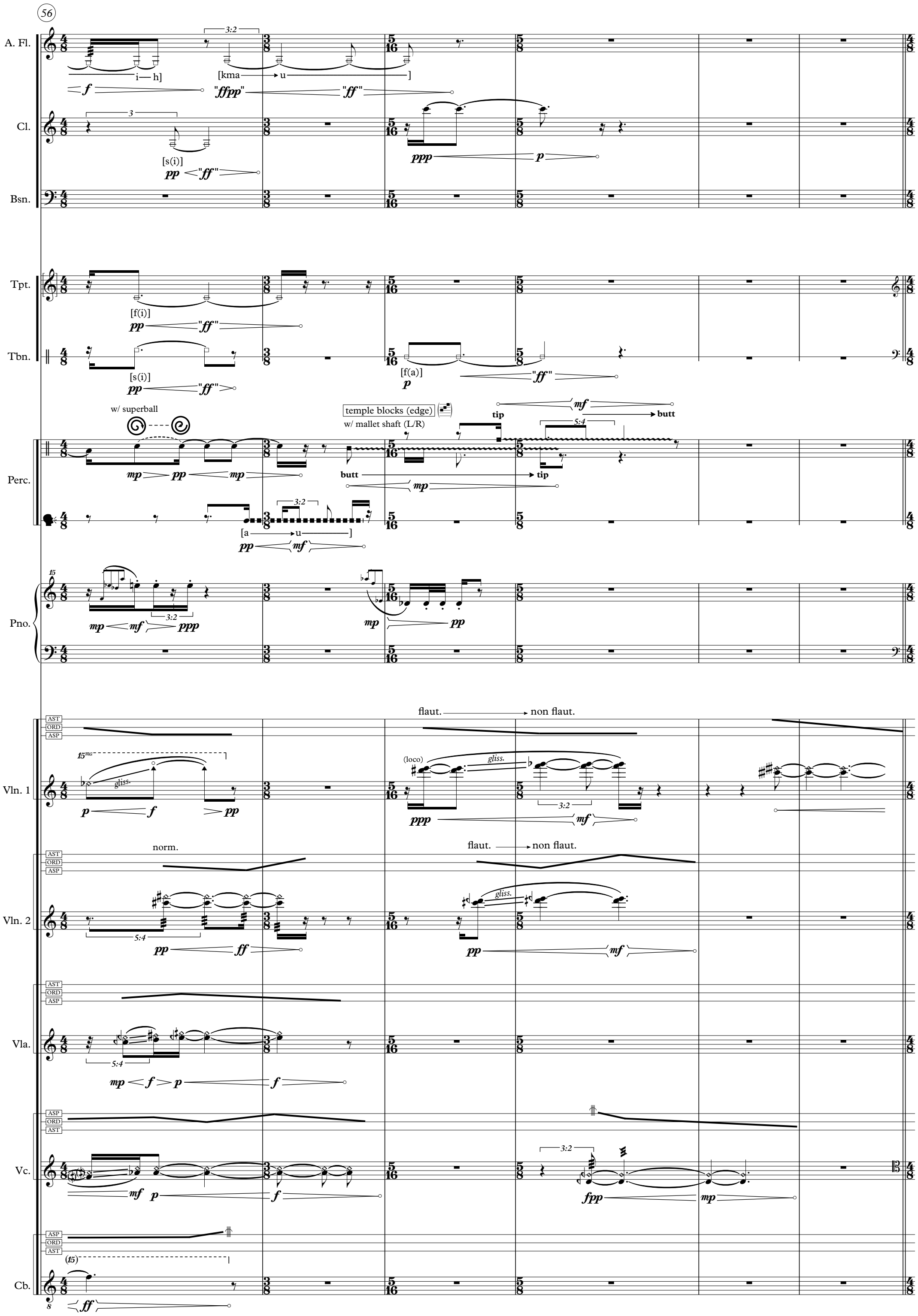
♪=80
26
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
(bisbigliando)
whistle tone
(avoid explicit overtone sweeps; try to stay within the very upper partial range)
(non vib.)
temple blocks (edge)
w/ mallet shaft
tip
butt
w/ reibestock
(cut off abruptly as soon as the piano starts)
(quasi echo)
(very mechanical)
½ col legno tratto
norm.
simile
flaut.
gliss.
(21st and 22nd harmonics)


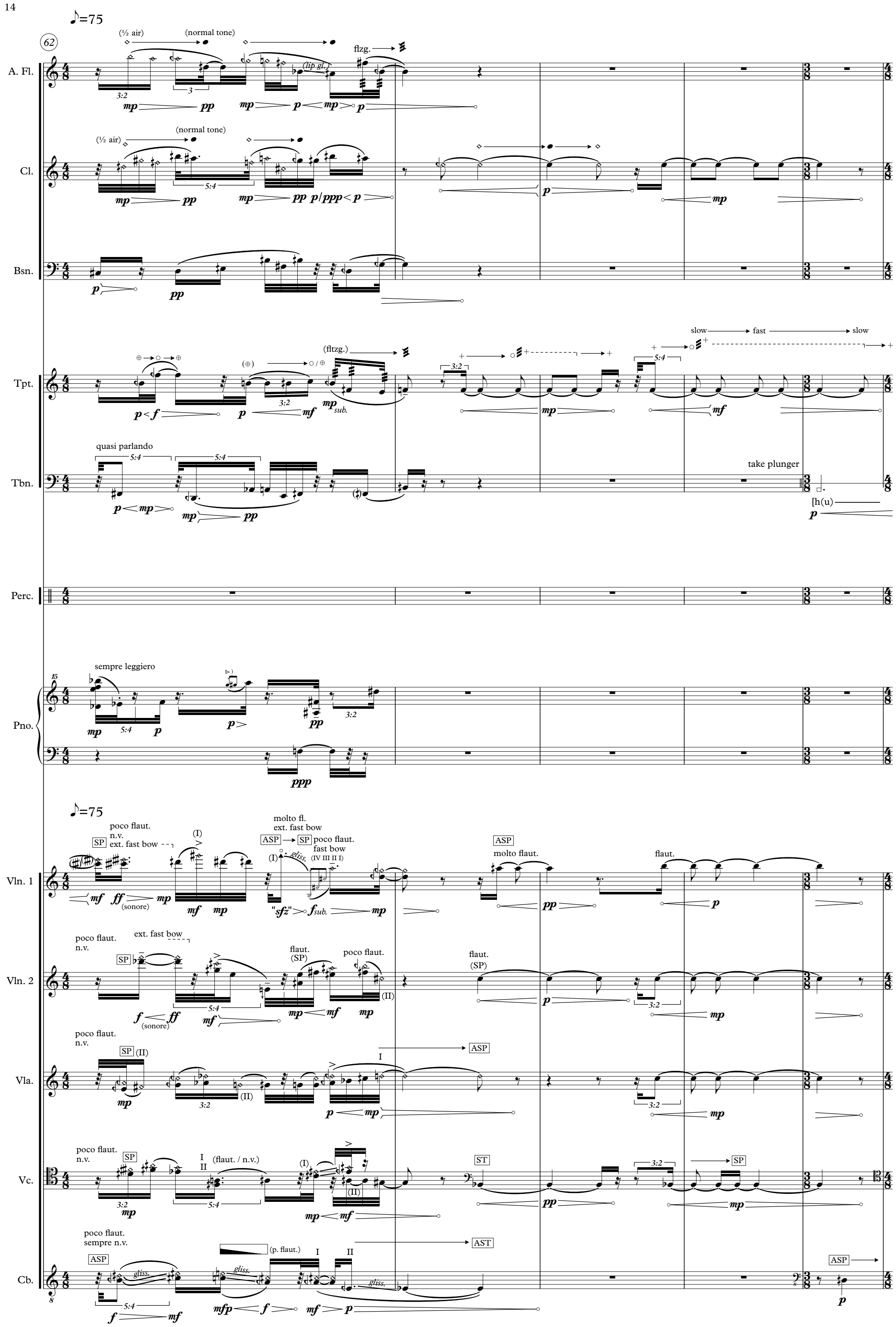
31
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
temple blocks (edge)
w/ reibestock
Perc.
ppp < pp
(start very slow as if imitating the muted
piano notes and go up to normal speed)
15
Pno.
p
5:4
norm.
Vln. 1
pp
Vln. 2
3:2
pp
Vla.
5:4
pp
Vc.
3:2
pp
flaut.
ORD
SP
AST
Cb.
f
pp


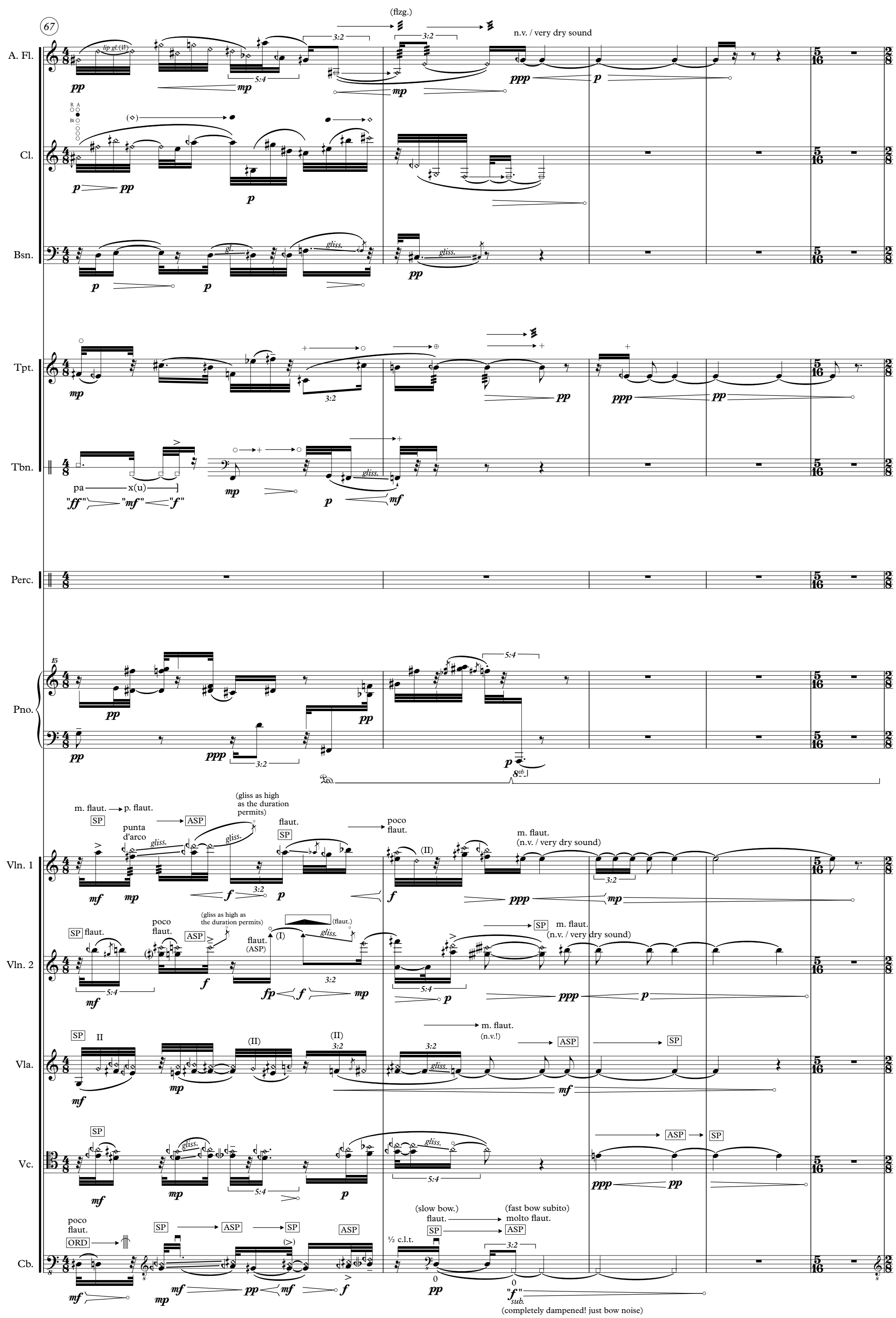
37
♪=75
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
[t'um]
[u]
[h r ts' ts' ts' s ts' r i h]
[kma]
[s(i)]
(ca. 3cm → 0cm)
[f(i)]
[mi h]
[f(a)]
snare drum
w/ superball
temple blocks
(edge)
w/ reibestock
bongos
w/ mallet shaft
[i]
flaut.
(loco)
gliss.
AST
ORD
ASP
(next to the finger)
(14th harmonic)


43
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
[f(u)]
[t'um]
[i]
[the li ya]
[wa] [(h)a r k'u ma ki]
[u]
[f(u)]
[f(i)]
[mi h]
bongos
snare drum
w/ superball
w/ triangle beater
snare drum (rim)
tip
butt
bongos
[a (a) u]
non flaut.
molto flaut
gliss.
AST
ORD
ASP


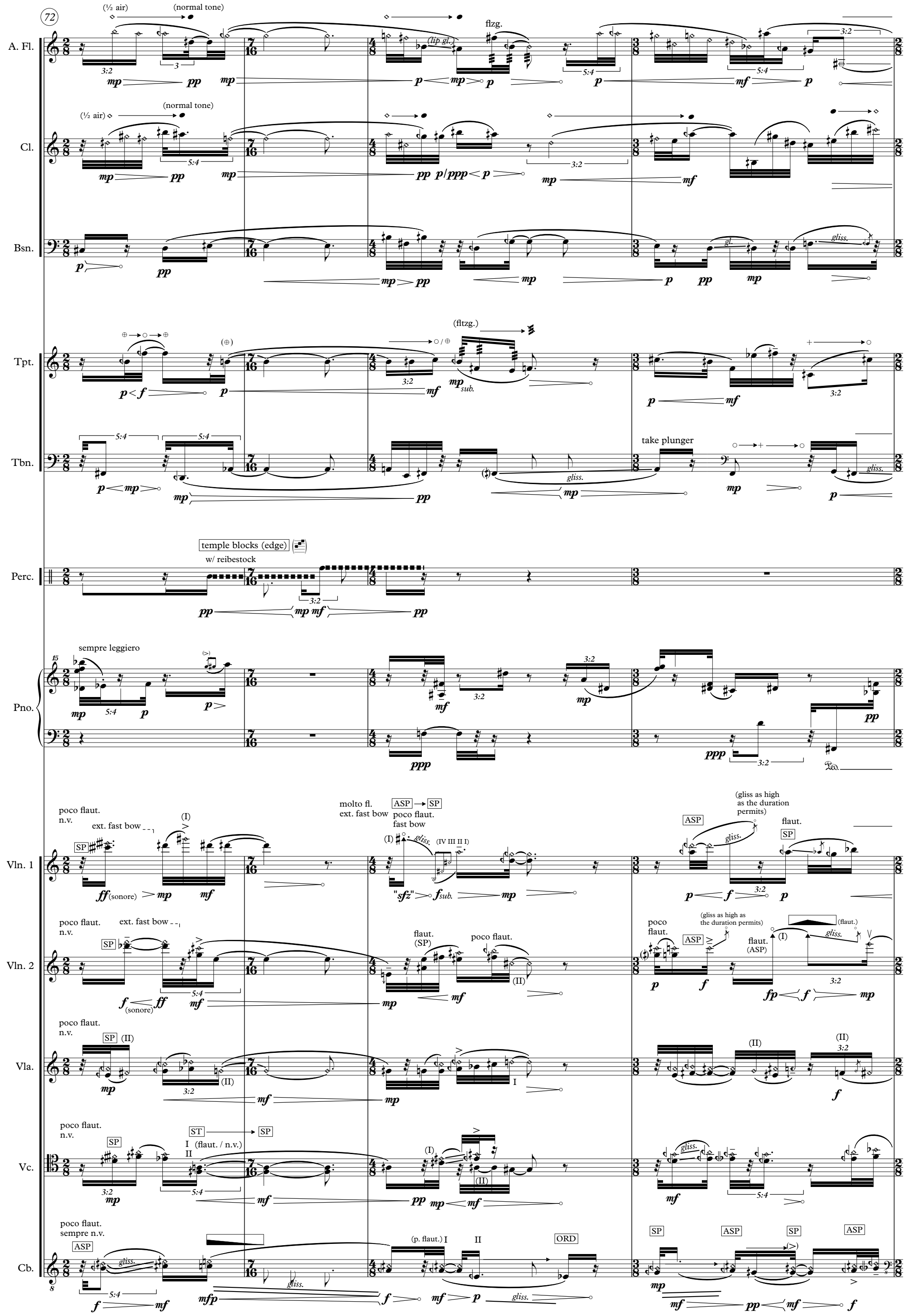
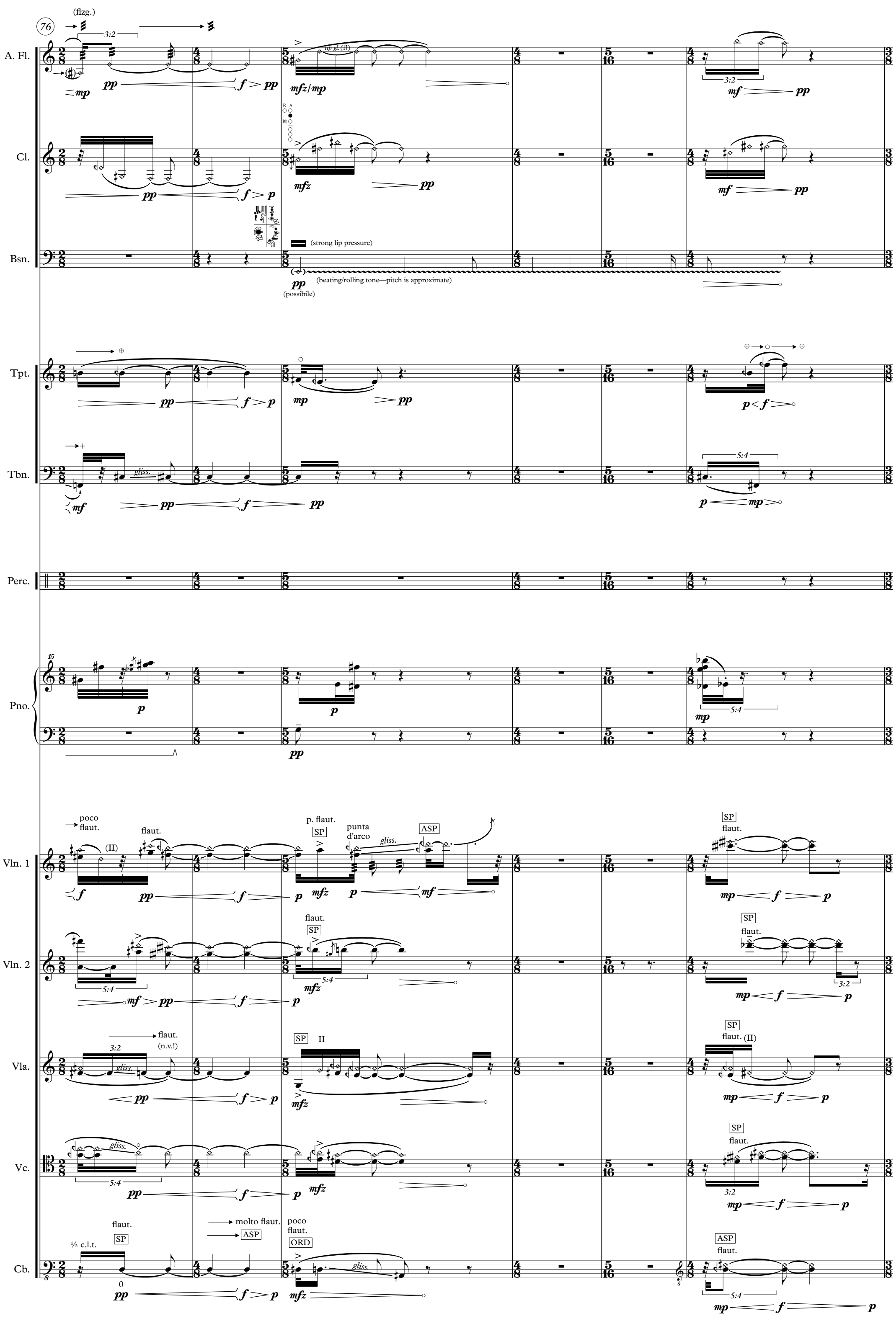
47
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
[a th ]
w/ triangle beater
snare drum (rim)
w/ triangle beater
tip
butt
gliss.
½ c.l.t.
simile
ad lib.
pizz.
arco
III II
AST
ORD
ASP


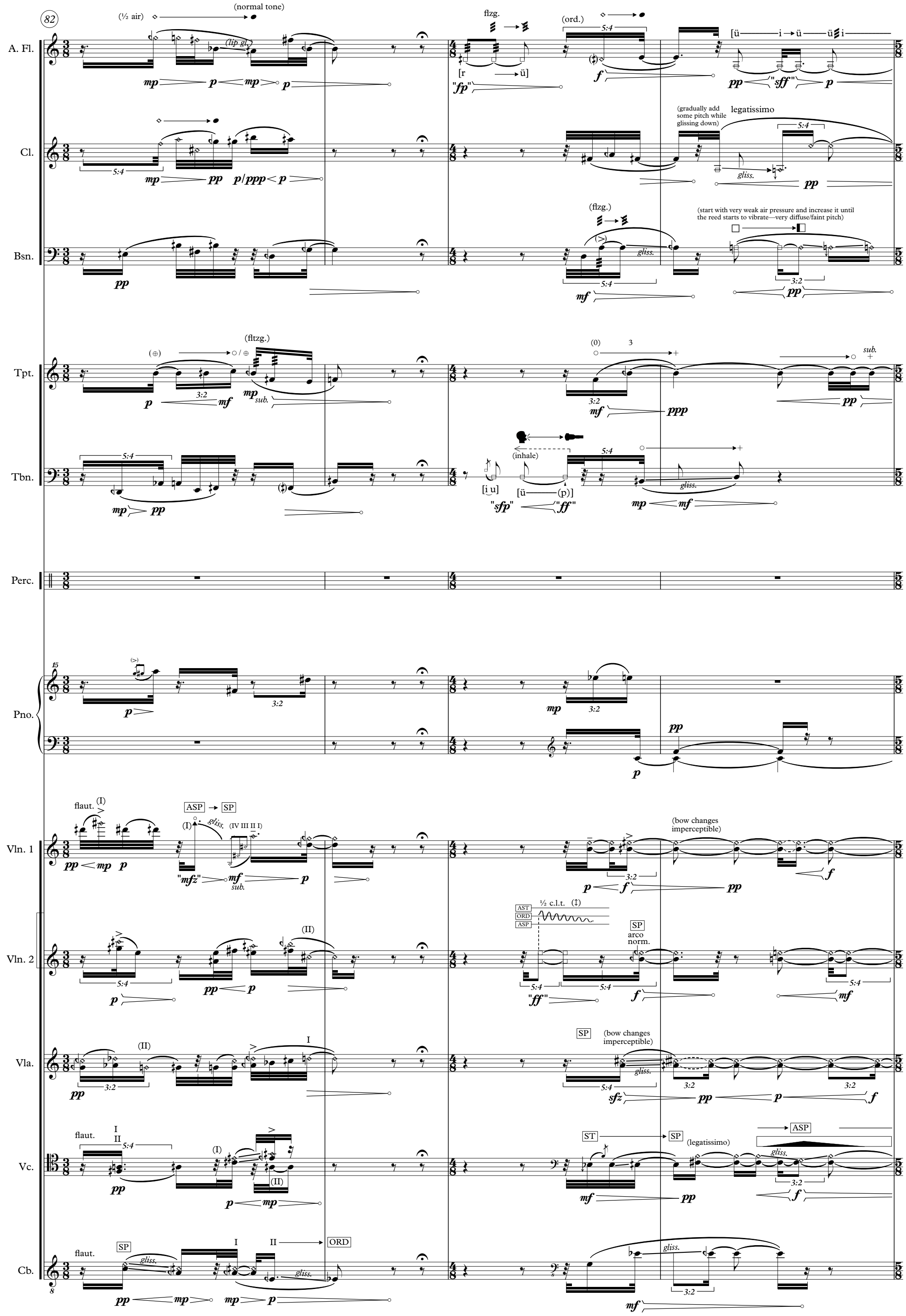
51
A. Fl.
[t'um]
"f"
[h — r — ts' ts' ts' s — ts' r
ppp mf p
5:4
3:2
Cl.
[u]
"mf"
ppp pp
Bsn.
(ca. 3cm → 0cm)
Tpt.
[f(i) — ]
p "f"
Tbn.
[mi — h]
"f"
3:2
temple blocks (edge)
w/ reibestock
snare drum (rim)
w/ triangle beater
tip → butt
snare drum
bongos
w/ superball
w/ superball
Perc.
ppp mp ppp mp pp p
[a]
mp
(stop abruptly)
[the — i — ya — wa — i — ü]
ppp
Pno.
mp ppp mf p
AST
ORD
ASP
Vln. 1
gliss.
mfzp f p sub. mf mp
norm.
Vln. 2
½ c.l.t.
mfz p mf pp mf
Vla.
mfz p f mf pp mp
ASP
ORD
AST
Vc.
mf fp mf p
Cb.
(next to the finger)
(14th harmonic)
mp f p
mp mp


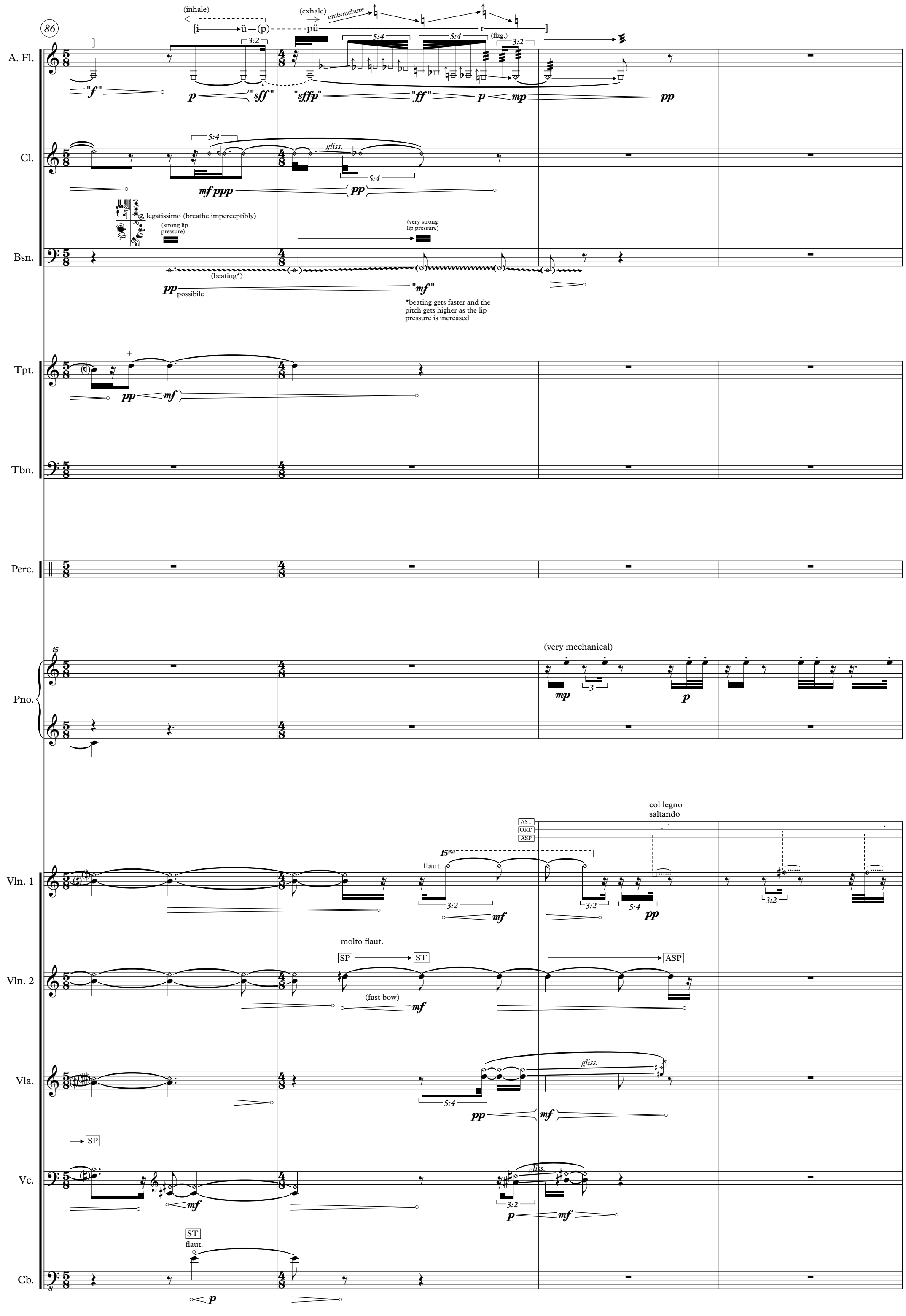
56
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
[kma
u
]
[s(i)]
[f(i)]
[f(a)]
w/ superball
temple blocks (edge)
w/ mallet shaft (L/R)
tip
butt
[a
u
]
flaut.
non flaut.
(loco)
gliss.
norm.
AST
ORD
ASP


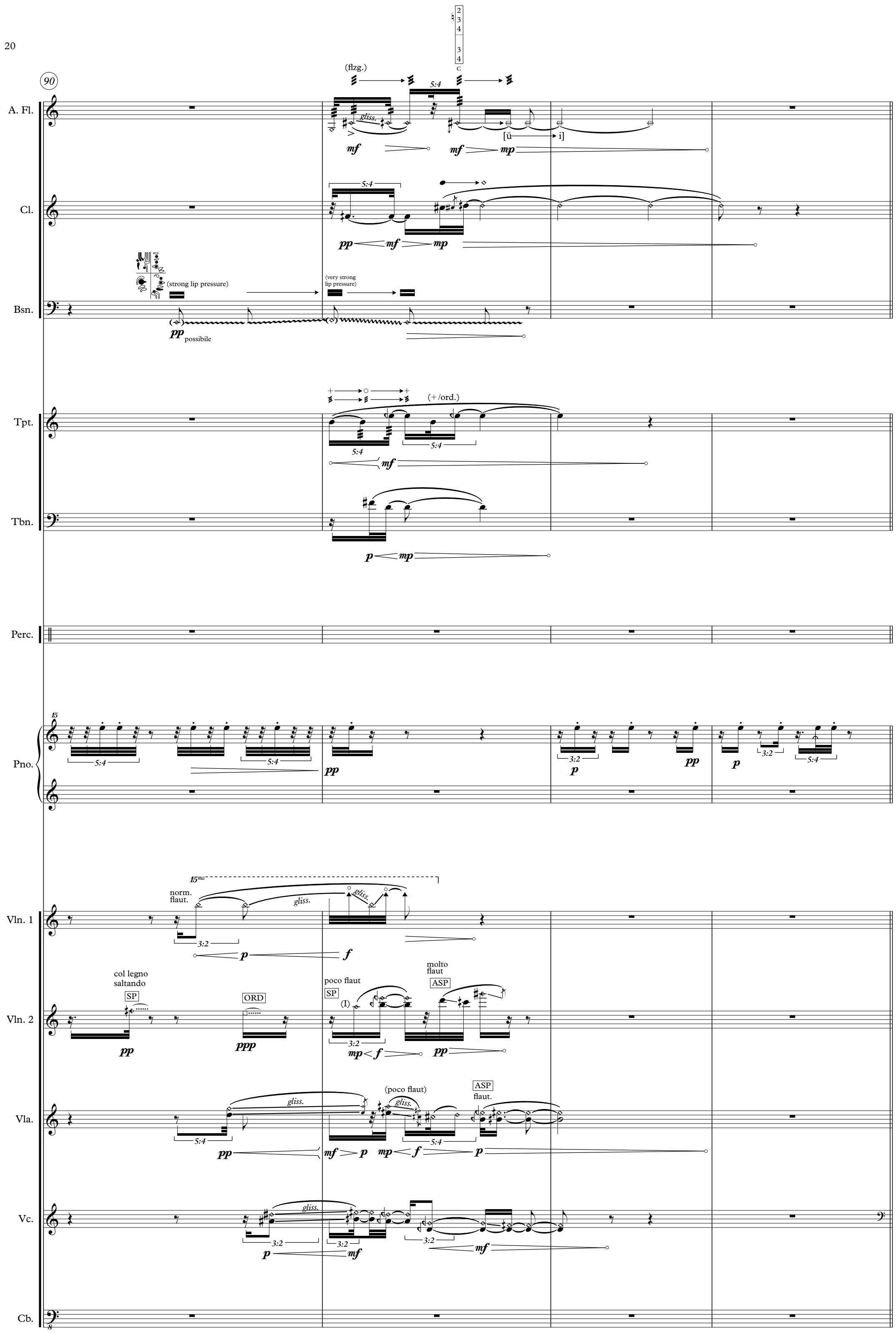
♪=75
62
A. Fl.
(½ air)
(normal tone)
(lip gl.)
flzg.
mp pp mp p mp p
Cl.
(½ air)
(normal tone)
5:4
mp pp mp pp p / ppp p
p mp
Bsn.
p pp
Tpt.
(fltzg.)
slow fast slow
p f p mf mp sub.
mp mf
Tbn.
quasi parlando
take plunger
[h(u)
p mp mp pp p
Perc.
Pno.
sempre leggiero
mp p p pp ppp
♪=75
Vln. 1
poco flaut.
n.v.
ext. fast bow
SP
ASP
molto fl.
poco flaut.
fast bow
(IV III II I)
gliss.
molto flaut.
flaut.
mf ff (sonore) mp mf mp "sfz" f sub. mp pp p
Vln. 2
poco flaut.
n.v.
ext. fast bow
flaut.
(SP)
(II)
f ff (sonore) mf mp mf mp p mp
Vla.
poco flaut.
n.v.
SP (II)
ASP
mp p mp mp
Vc.
poco flaut.
n.v.
(flaut. / n.v.)
ST
SP
mp mp mf pp mp
Cb.
poco flaut.
sempre n.v.
ASP
(p. flaut.)
gliss.
AST
f mf mfp f mf p p


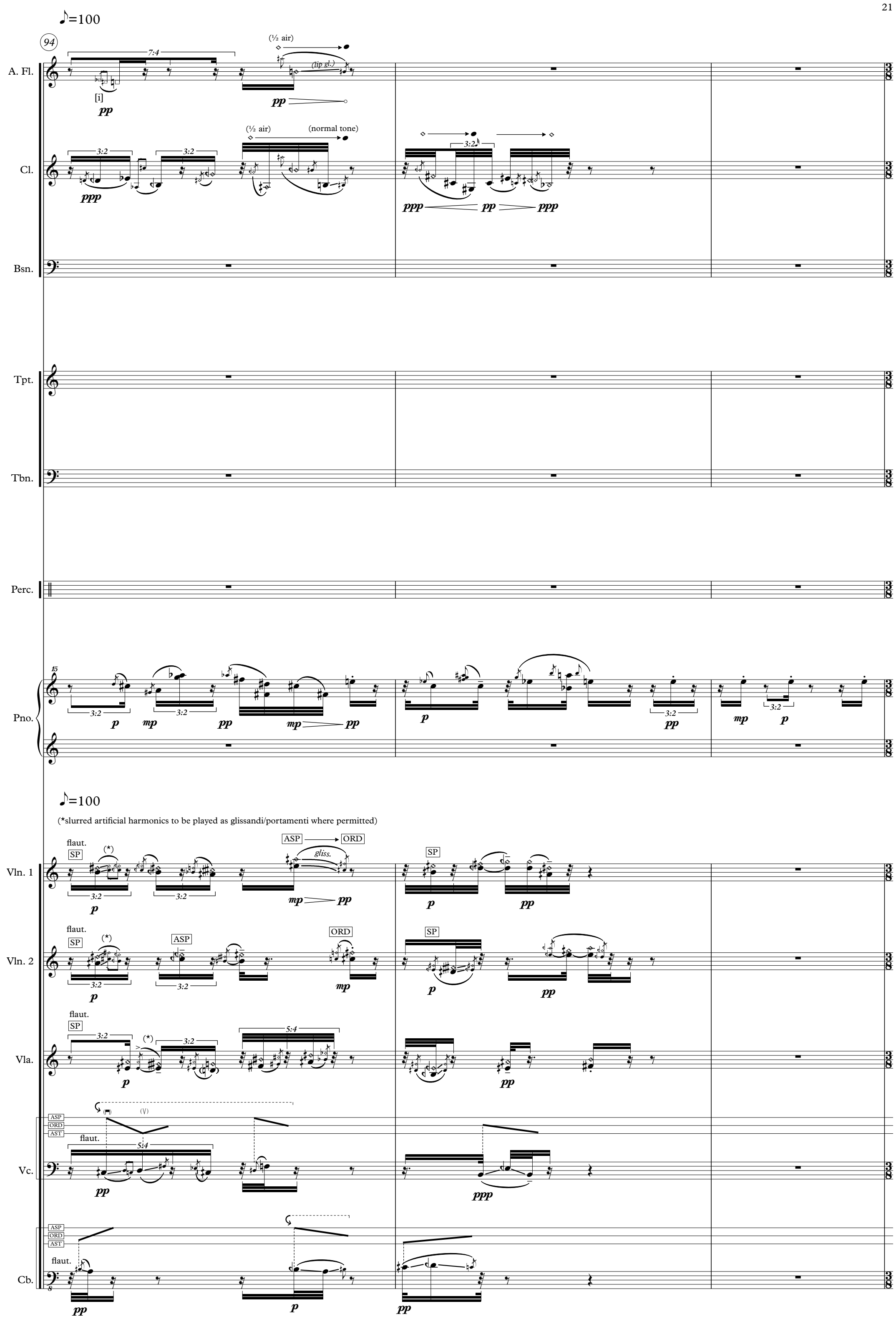
67
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
(flzg.)
n.v. / very dry sound
gliss.
pa
x(u)
m. flaut.
p. flaut.
punta d'arco
(gliss as high as the duration permits)
flaut.
poco flaut.
m. flaut. (n.v. / very dry sound)
flaut. (ASP)
(flaut.)
m. flaut. (n.v.!)
ORD
½ c.l.t.
(slow bow.)
(fast bow subito) molto flaut.
(completely dampened! just bow noise)

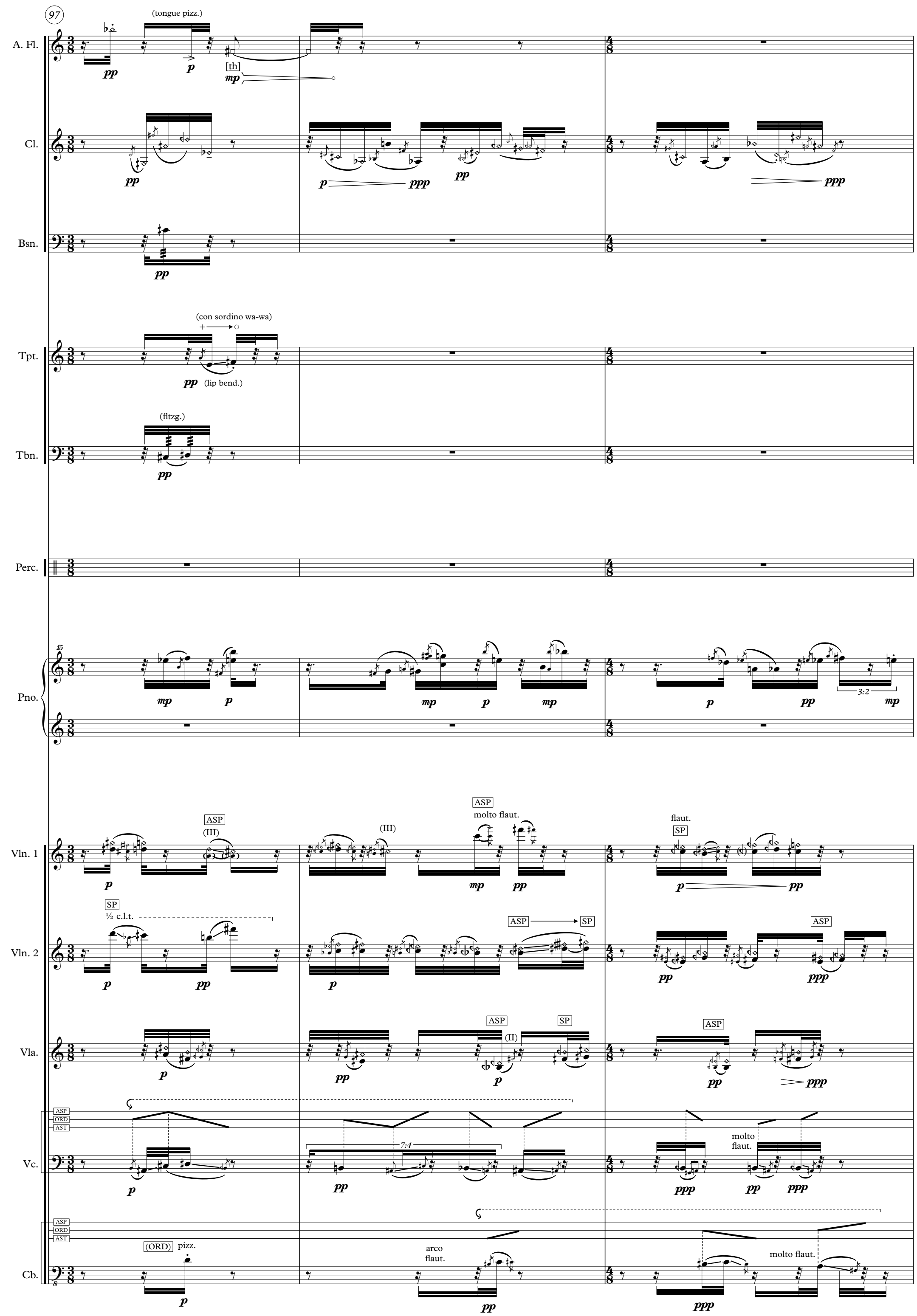

72
A. Fl.
(½ air)
(normal tone)
flzg.
Cl.
Bsn.
Tpt.
(fltzg.)
Tbn.
take plunger
Perc.
temple blocks (edge)
w/ reibestock
Pno.
sempre leggiero
Vln. 1
poco flaut.
n.v.
ext. fast bow
molto fl.
ext. fast bow
poco flaut.
fast bow
(gliss as high as the duration permits)
flaut.
Vln. 2
poco flaut.
(sonore)
(gliss as high as the duration permits)
Vla.
Vc.
I (flaut. / n.v.)
Cb.
poco flaut.
sempre n.v.
(p. flaut.)
ORD


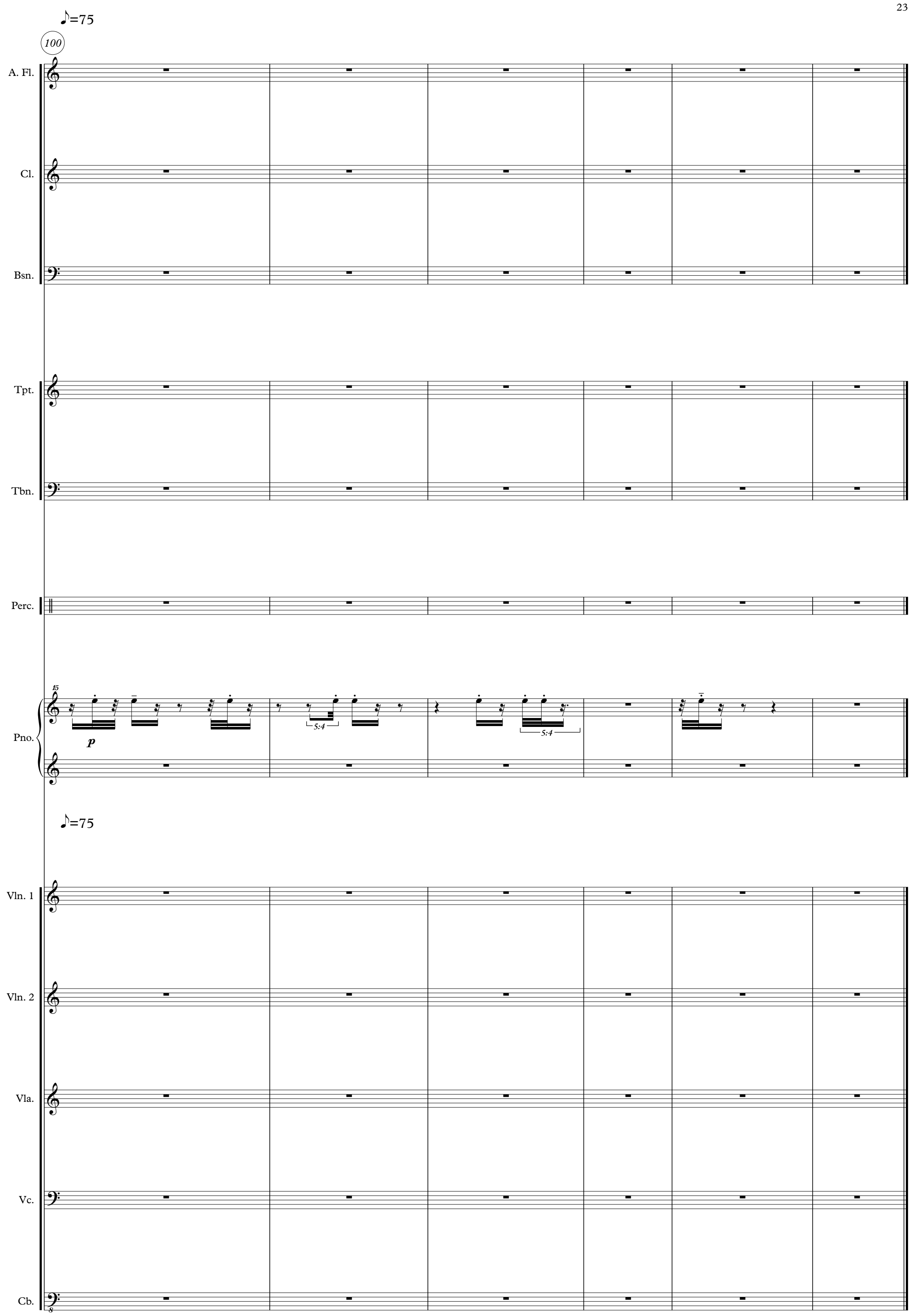
76
(flzg.)
A. Fl.
Cl.
Bsn.
(strong lip pressure)
(beating/rolling tone—pitch is approximate)
(possibile)
Tpt.
Tbn.
gliss.
Perc.
Pno.
Vln. 1
poco flaut.
flaut.
p. flaut.
SP
punta d'arco
gliss.
ASP
Vln. 2
flaut.
SP
Vla.
flaut. (n.v.!)
gliss.
SP
II
flaut.
(II)
Vc.
gliss.
SP
flaut.
Cb.
½ c.l.t.
flaut.
SP
molto flaut.
ASP
poco flaut.
ORD
gliss.
ASP
flaut.


82
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
(½ air)
(normal tone)
(lip gl.)
flzg.
(ord.)
legatissimo
(gradually add some pitch while glissing down)
(start with very weak air pressure and increase it until the reed starts to vibrate—very diffuse/faint pitch)
(flzg.)
(fltzg.)
(inhale)
flaut.
ASP
SP
ORD
ST
gliss.
(bow changes imperceptible)
½ c.l.t.
arco norm.
(legatissimo)


86
(inhale)
(exhale)
embouchure
[i ü (p) pü r]
(flzg.)
A. Fl.
Cl.
gliss.
legatissimo (breathe imperceptibly)
(strong lip pressure)
(very strong lip pressure)
Bsn.
(beating*)
possibile
*beating gets faster and the pitch gets higher as the lip pressure is increased
Tpt.
Tbn.
Perc.
(very mechanical)
Pno.
col legno
saltando
AST
ORD
ASP
flaut.
Vln. 1
molto flaut.
SP
ST
ASP
(fast bow)
Vln. 2
Vla.
SP
Vc.
ST
flaut.
Cb.


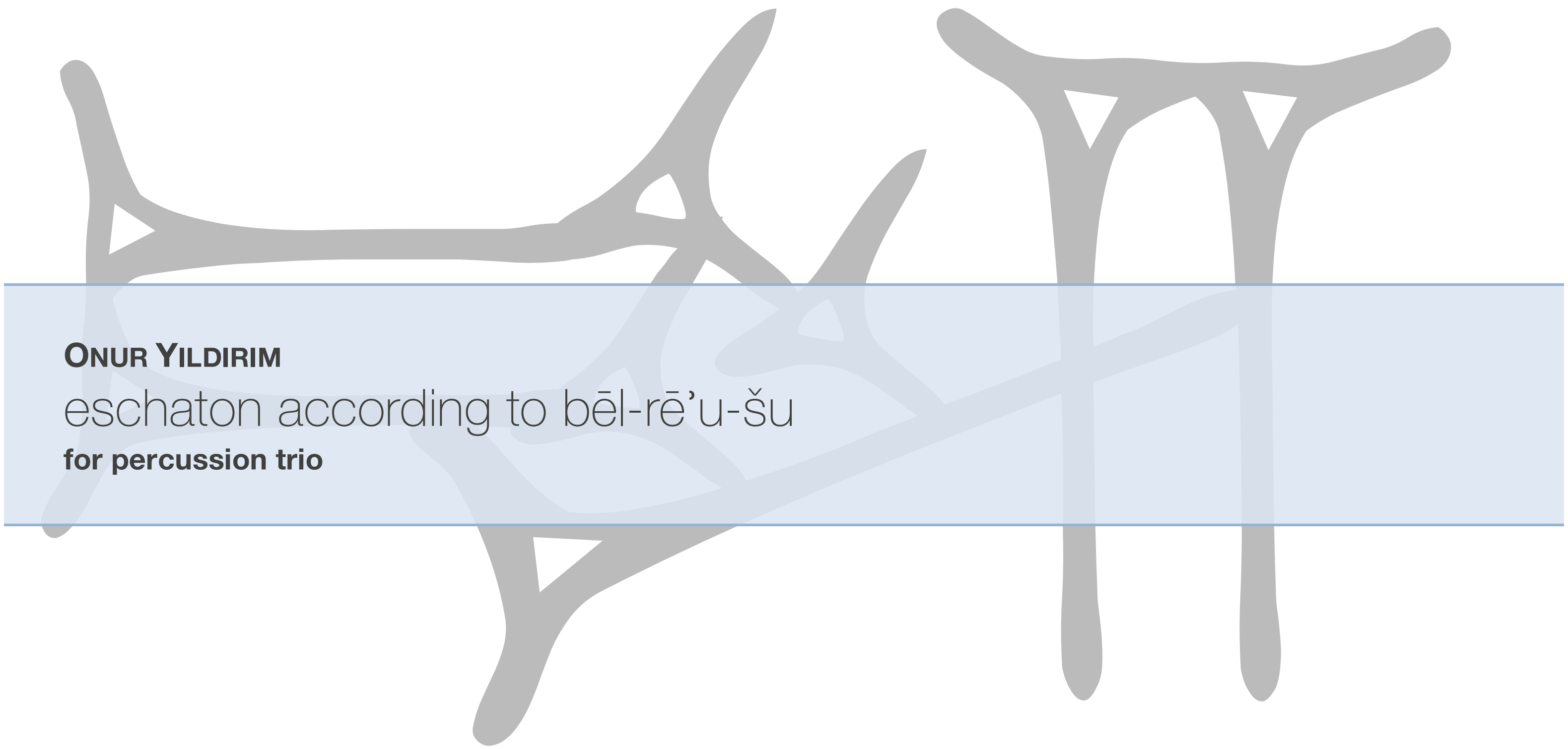
90
A. Fl.
(flzg.)
5:4
gliss.
[ü
i]
mf
mf
mp
Cl.
5:4
pp
mf
mp
Bsn.
(strong lip pressure)
(very strong lip pressure)
pp possibile
Tpt.
(+/ord.)
5:4
5:4
mf
Tbn.
p
mp
Perc.
Pno.
15
5:4
5:4
pp
3:2
p
pp
p
3:2
5:4
Vln. 1
15ma
norm. flaut.
gliss.
gliss.
3:2
p
f
Vln. 2
col legno saltando
SP
ORD
pp
ppp
poco flaut
SP
(I)
3:2
mp
f
molto flaut
ASP
pp
Vla.
gliss.
(poco flaut)
gliss.
ASP
flaut.
5:4
5:4
pp
mf
p
mp
f
p
Vc.
gliss.
3:2
3:2
3:2
p
mf
mf
Cb.


♪=100
94
A. Fl.
7:4
(½ air)
(tip gl.)
[i]
pp
Cl.
3:2
(normal tone)
ppp
Bsn.
Tpt.
Tbn.
Perc.
Pno.
p
mp
♪=100
(*slurred artificial harmonics to be played as glissandi/portamenti where permitted)
Vln. 1
flaut.
SP
(*)
ASP
ORD
gliss.
Vln. 2
Vla.
5:4
Vc.
AST
Cb.


97
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
Vln. 1
Vln. 2
Vla.
Vc.
Cb.
(tongue pizz.)
[th]
(con sordino wa-wa)
(lip bend.)
(fltzg.)
ASP
(III)
SP
½ c.l.t.
molto flaut.
flaut.
(II)
ORD
AST
arco
(ORD)
pizz.
7:4
3:2


♪=75
100
A. Fl.
Cl.
Bsn.
Tpt.
Tbn.
Perc.
Pno.
p
5:4
5:4
♪=75
Vln. 1
Vln. 2
Vla.
Vc.
Cb.

## Music Score B: *eschaton according to bēl-rē’u-šu* (2016, rev. 2021)



# Instrumentation

## Percussion I

snare drum
2 bongos
1 boobam (A3)
3 temple blocks
gong (+bucket of water)
1 rin (G#4)
simantra
flexatone
maracas

**Beaters:** reibestock, 2 triangle beaters, drum stick, brush, superball mallet, gong beater, bow

## Percussion II

1 chinese gong - descending (+bucket of water)
1 rin (A#4)
simantra
kutuwapa (D4)
4 round stones
maracas
claves
polystyrene

**Beaters:** reibestock, rubber mallet, drum stick, gong beater, bow

## Percussion III

tenor drum
symphonic bass drum
4 bongos
3 temple blocks
cabasa

**Beaters:** reibestock, triangle beater, drum stick, brush, 2 superball mallets

**Duration:** ca. 4'30"

# Program Notes

Inspired by Attilâ İlhan's poem *Yanılsama* ("Illusion" in Turkish), *eschaton according to bēl-rēʾu-šŭ* ΒΗΡΩΣΣΟΣ
is a meditation

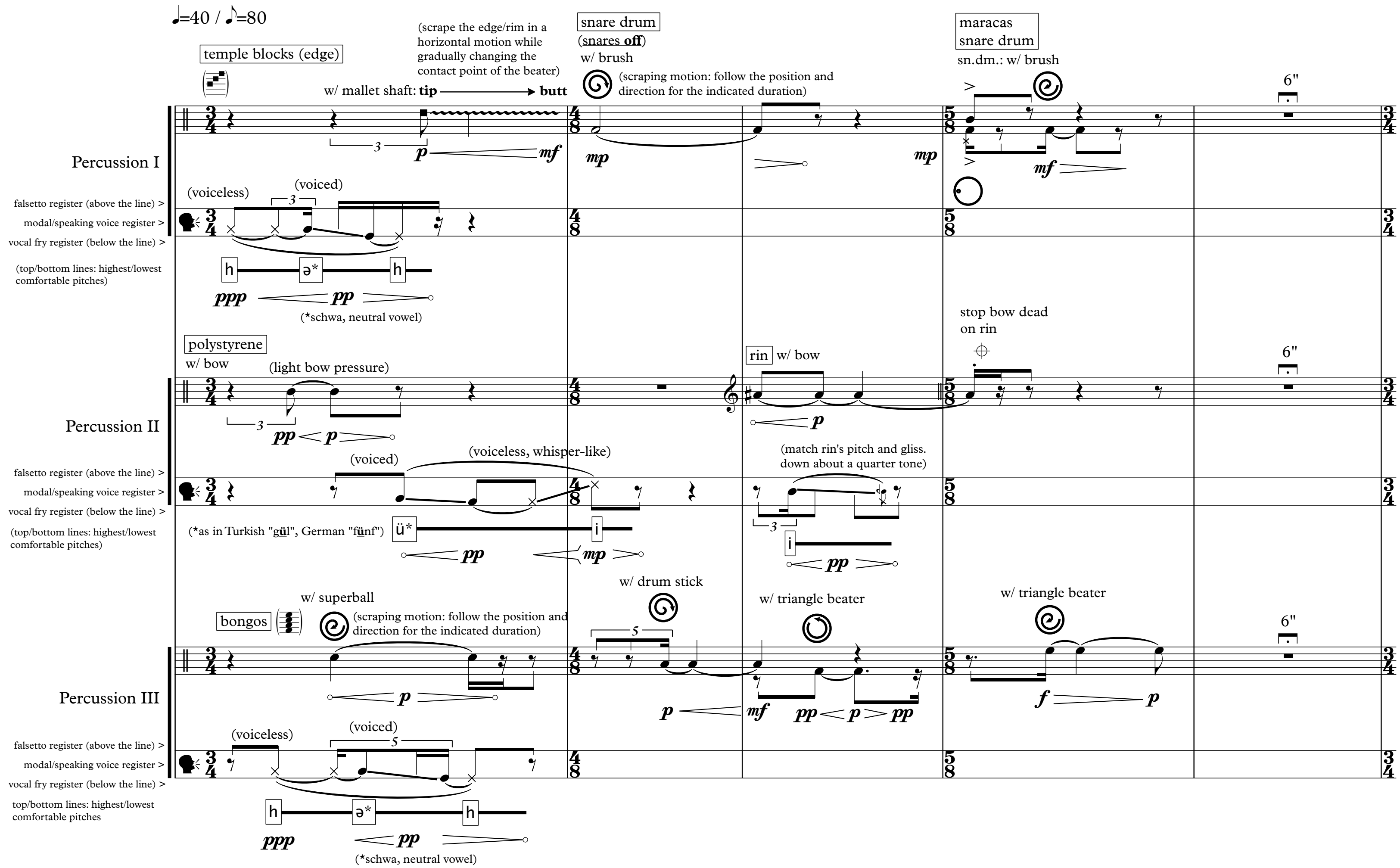
on the ultimate fate

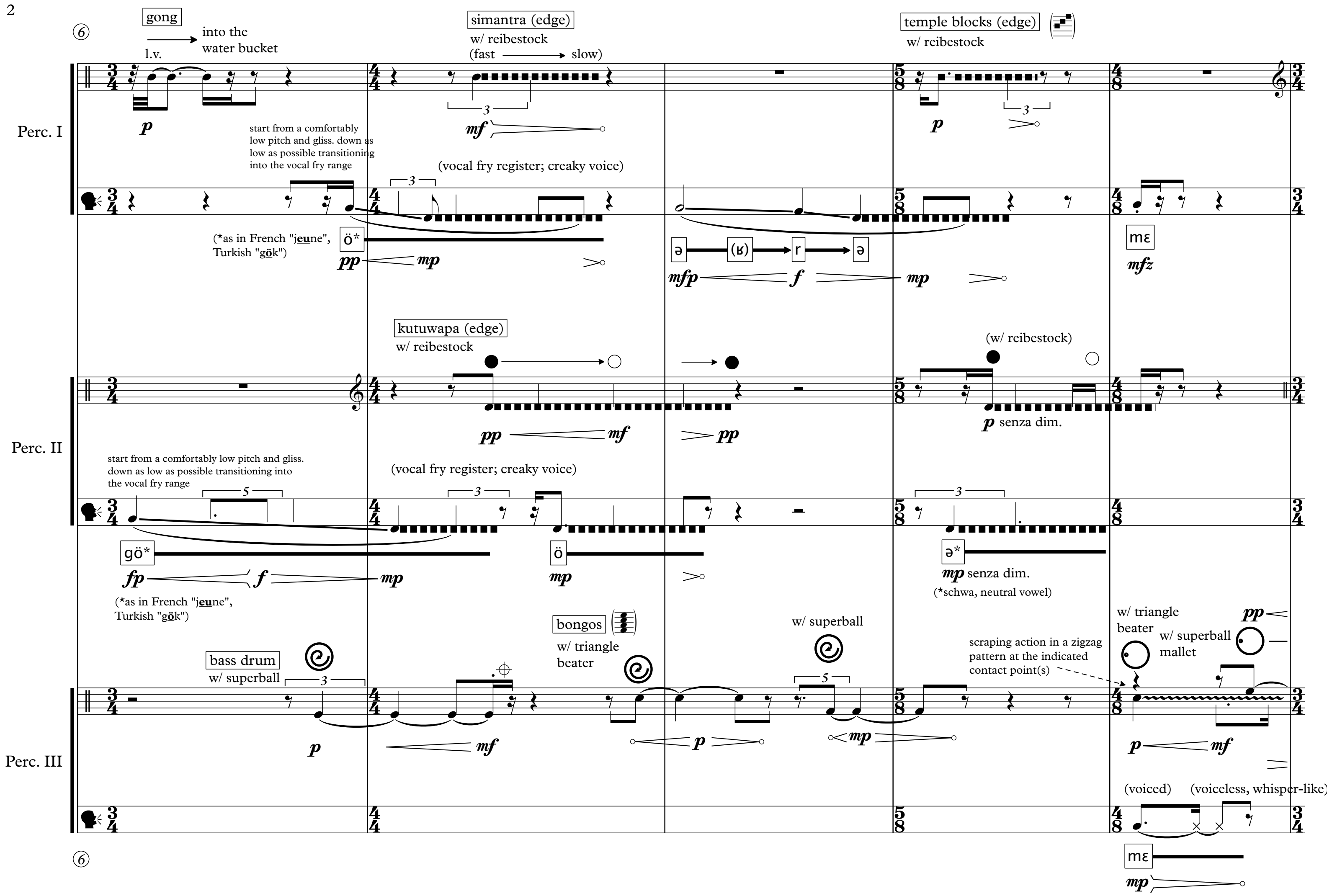
of

space, time and matter.

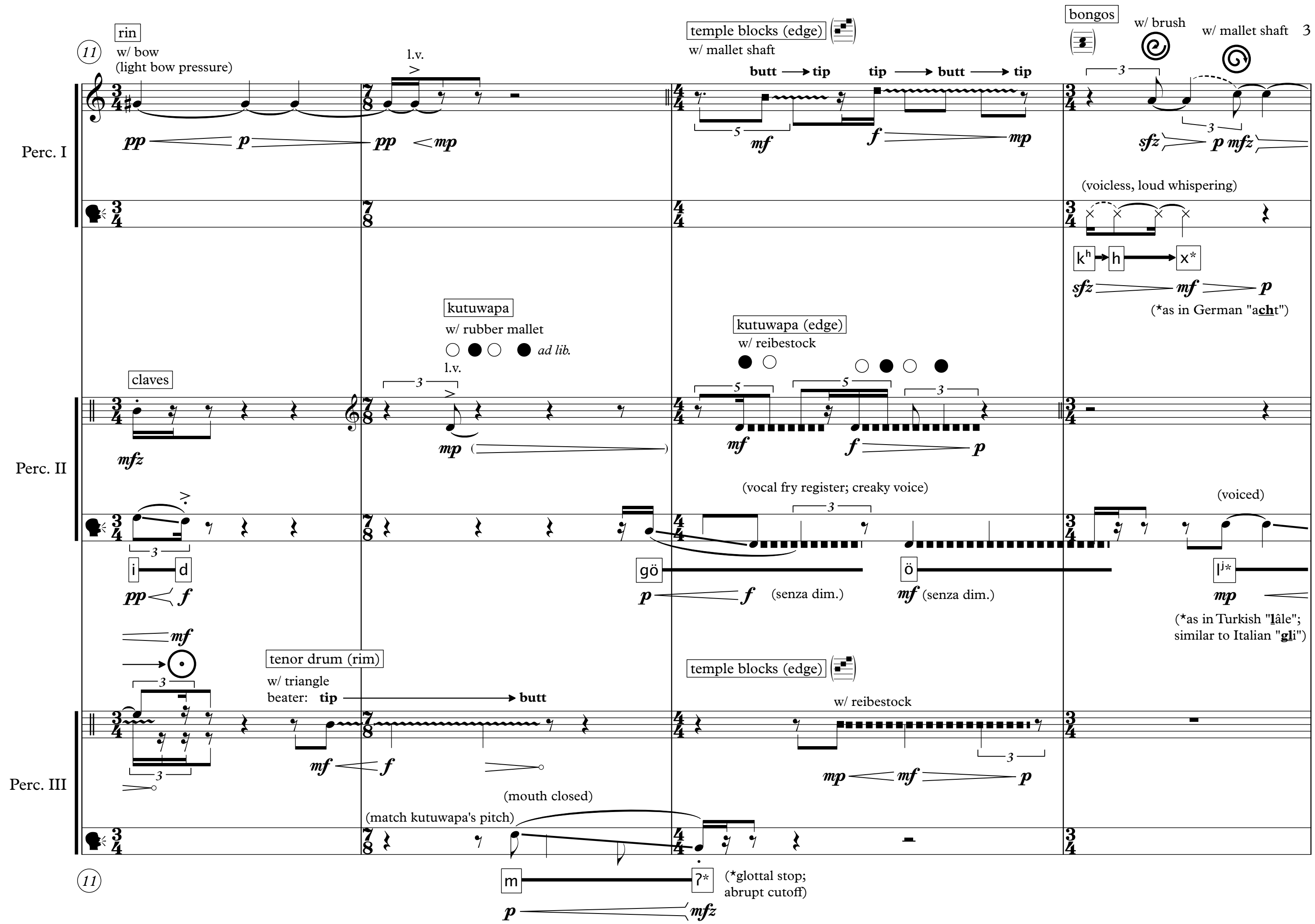
# eschaton according to bēl-rēʾu-šu

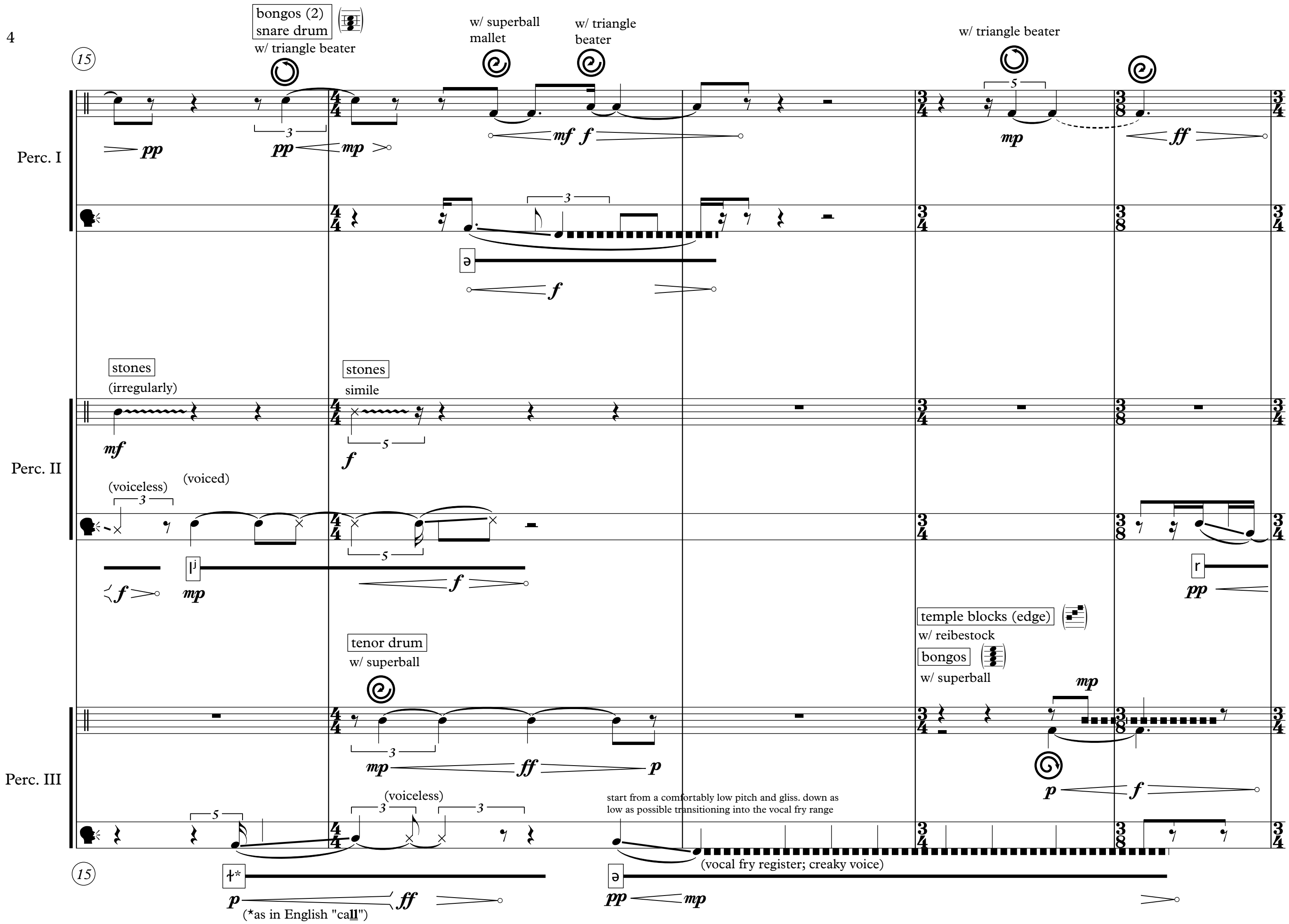
onur yıldırım (2016, rev. 2021)

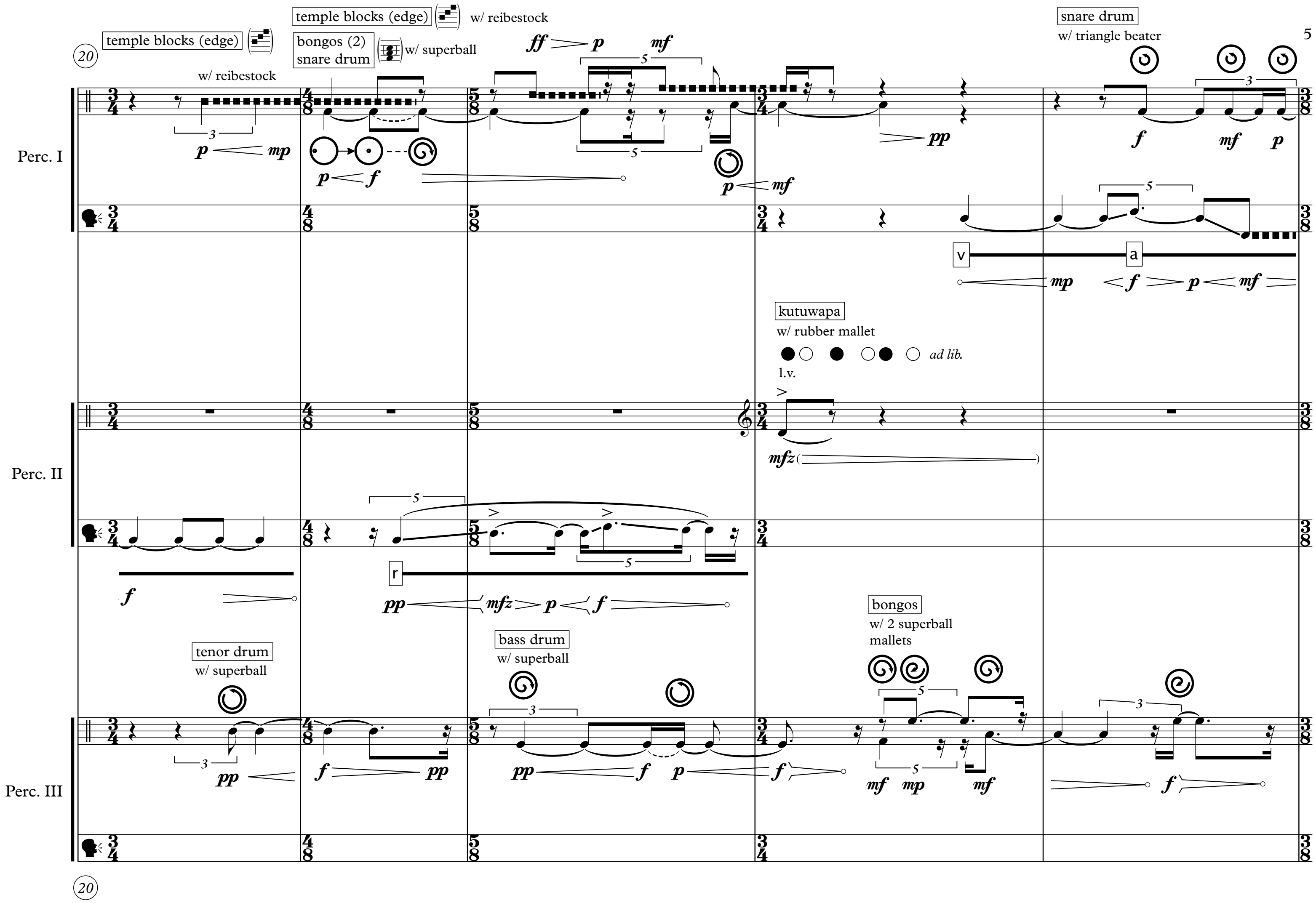


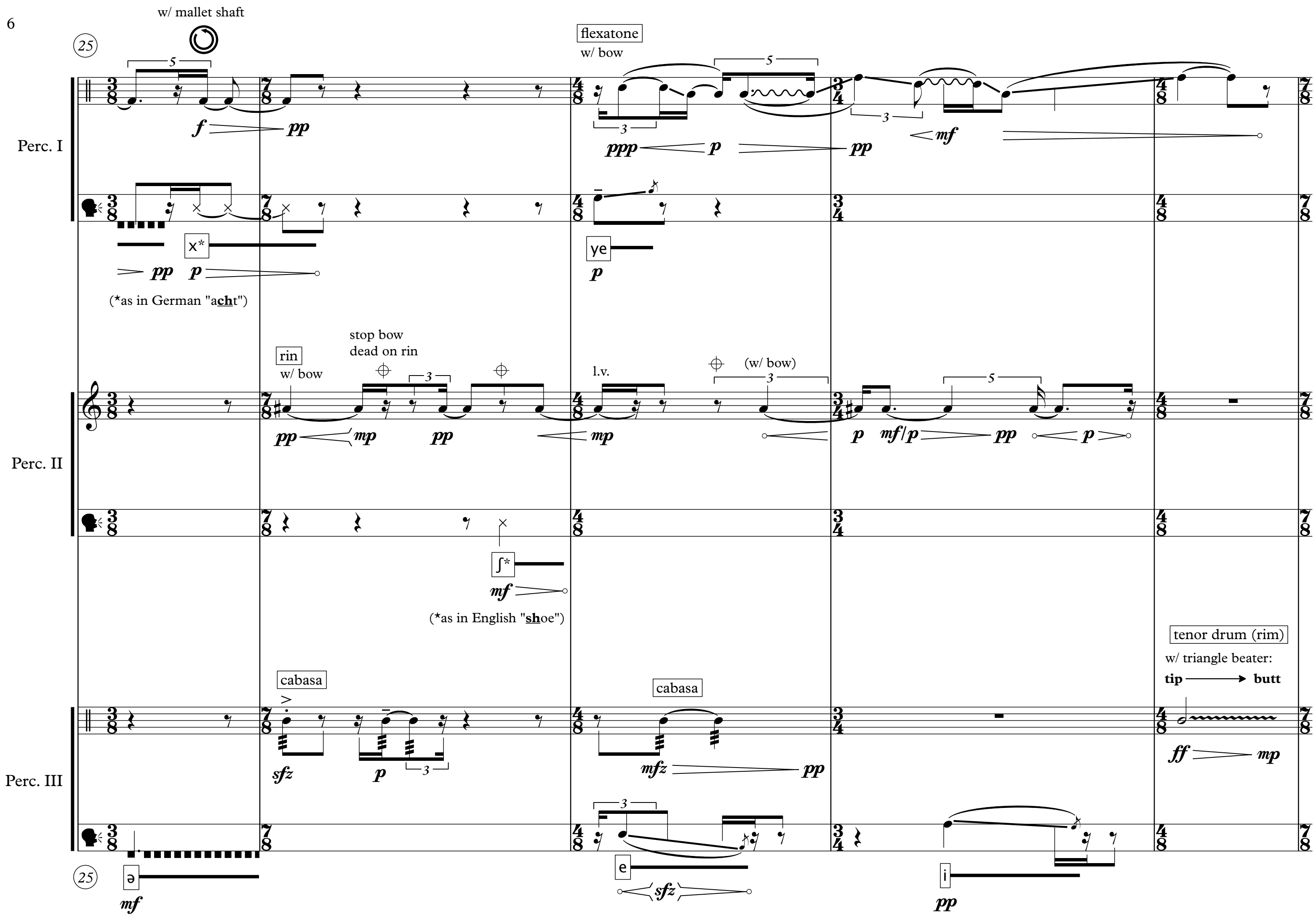

6
gong
l.v.
into the water bucket
simantra (edge)
w/ reibestock
(fast slow)
temple blocks (edge)
w/ reibestock
Perc. I
p
mf
p
start from a comfortably low pitch and gliss. down as low as possible transitioning into the vocal fry range
(vocal fry register; creaky voice)
(*as in French "jeune", Turkish "gök")
ö*
pp mp
ə (ʁ) r ə
mfp f mp
mɛ
mfz
kutuwapa (edge)
w/ reibestock
(w/ reibestock)
Perc. II
pp mf pp
p senza dim.
start from a comfortably low pitch and gliss. down as low as possible transitioning into the vocal fry range
(vocal fry register; creaky voice)
gö*
fp f mp
ö
mp
ə*
mp senza dim.
(*schwa, neutral vowel)
(*as in French "jeune", Turkish "gök")
bass drum
w/ superball
bongos
w/ triangle beater
w/ superball
w/ triangle beater
w/ superball mallet
scraping action in a zigzag pattern at the indicated contact point(s)
Perc. III
p
mf
p
mp
p mf
pp
(voiced)
(voiceless, whisper-like)
mɛ
mp
6

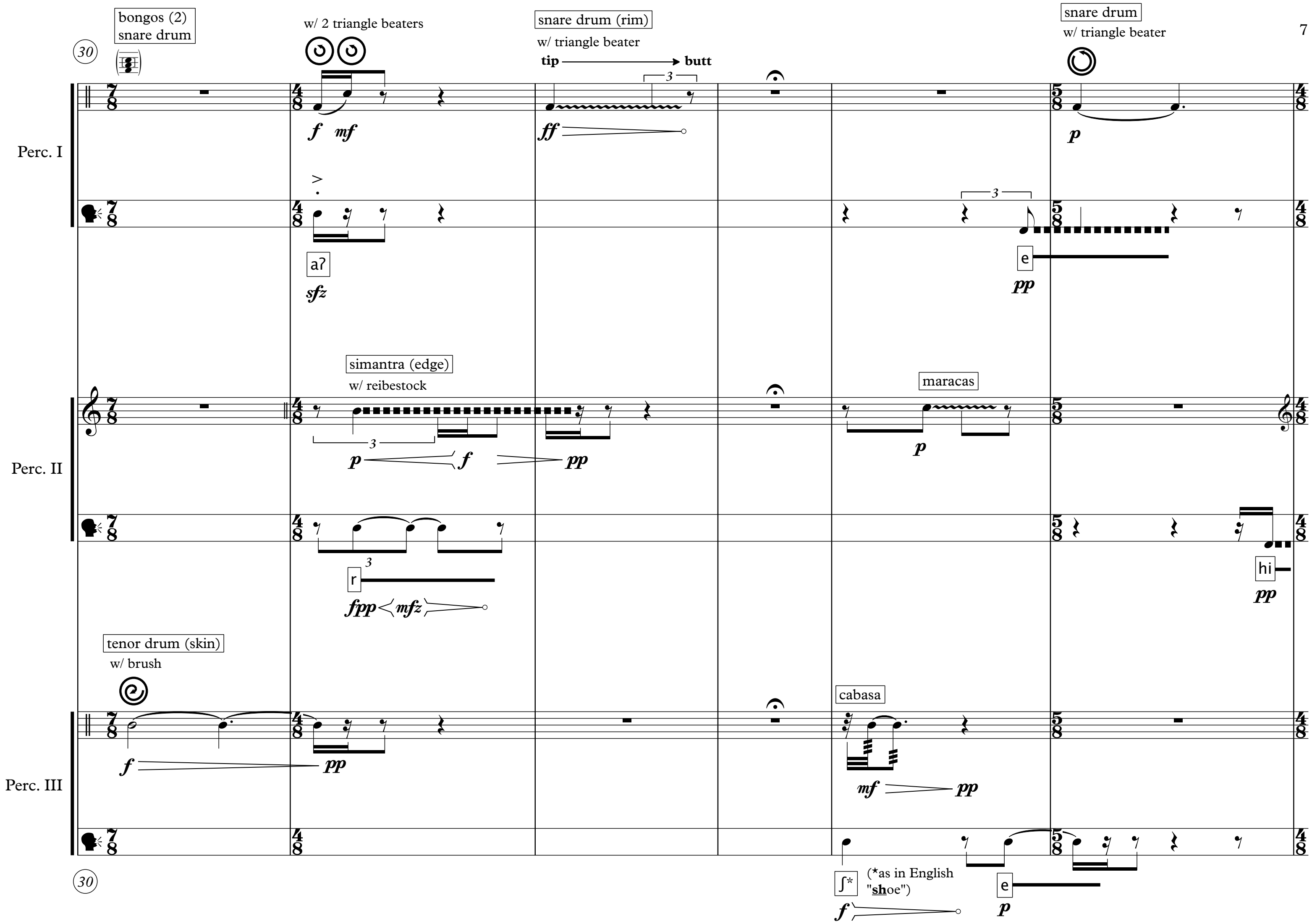
88

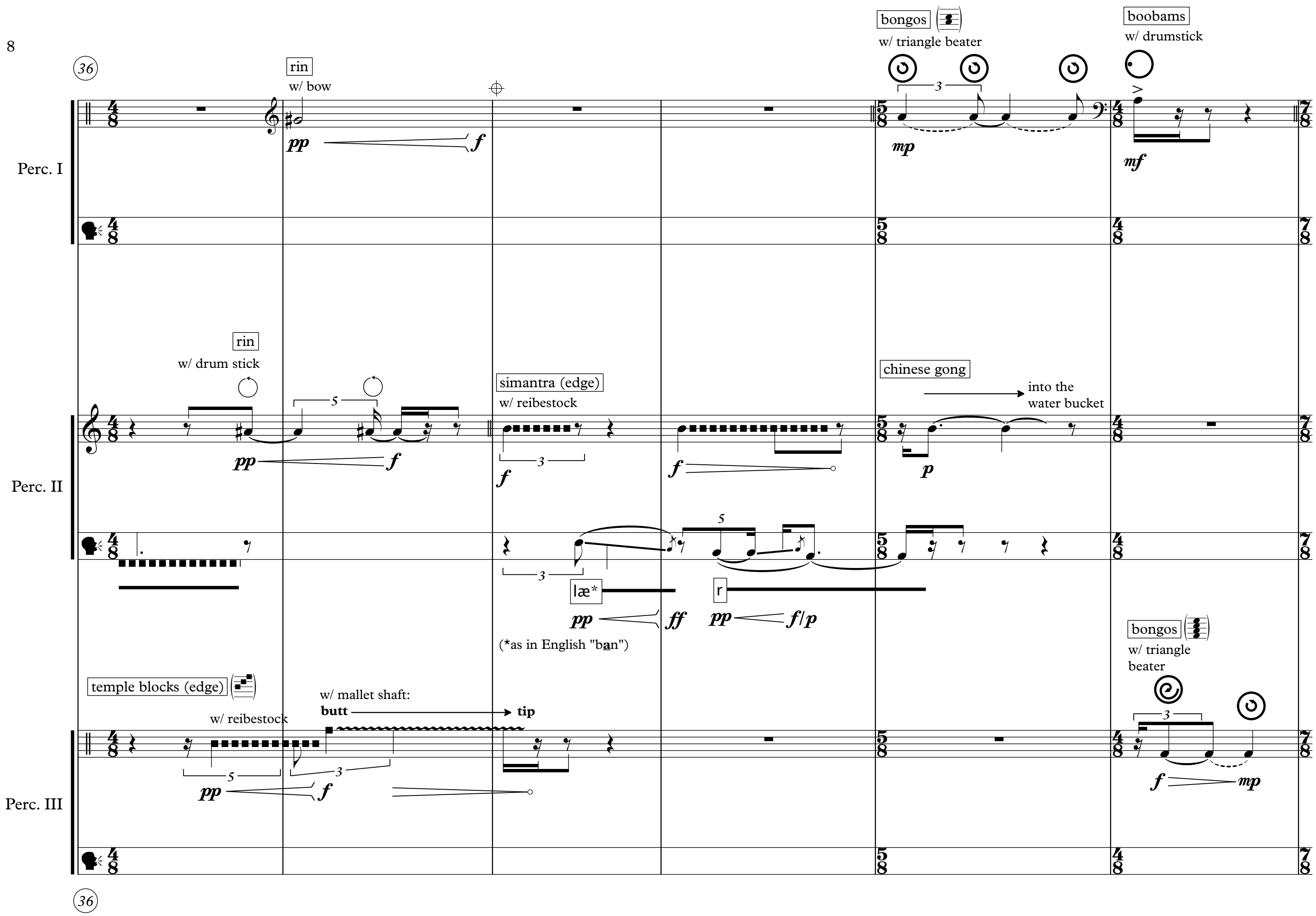

11
rin
w/ bow
(light bow pressure)
l.v.
temple blocks (edge)
w/ mallet shaft
butt → tip
tip → butt → tip
bongos
w/ brush
w/ mallet shaft
3
Perc. I
pp p pp mp
mf f mp
sfz p mfz
(voiceless, loud whispering)
kʰ h x*
sfz mf p
(*as in German "acht")
kutuwapa
w/ rubber mallet
ad lib.
l.v.
kutuwapa (edge)
w/ reibestock
claves
Perc. II
mfz
mp
mf f p
(vocal fry register; creaky voice)
(voiced)
i d
pp f
gö
p f (senza dim.)
ö
mf (senza dim.)
lʲ*
mp
(*as in Turkish "lâle"; similar to Italian "gli")
mf
tenor drum (rim)
w/ triangle
beater: tip → butt
temple blocks (edge)
w/ reibestock
Perc. III
mf f
mp mf p
(mouth closed)
(match kutuwapa's pitch)
m
ʔ*
(*glottal stop; abrupt cutoff)
p mfz
11

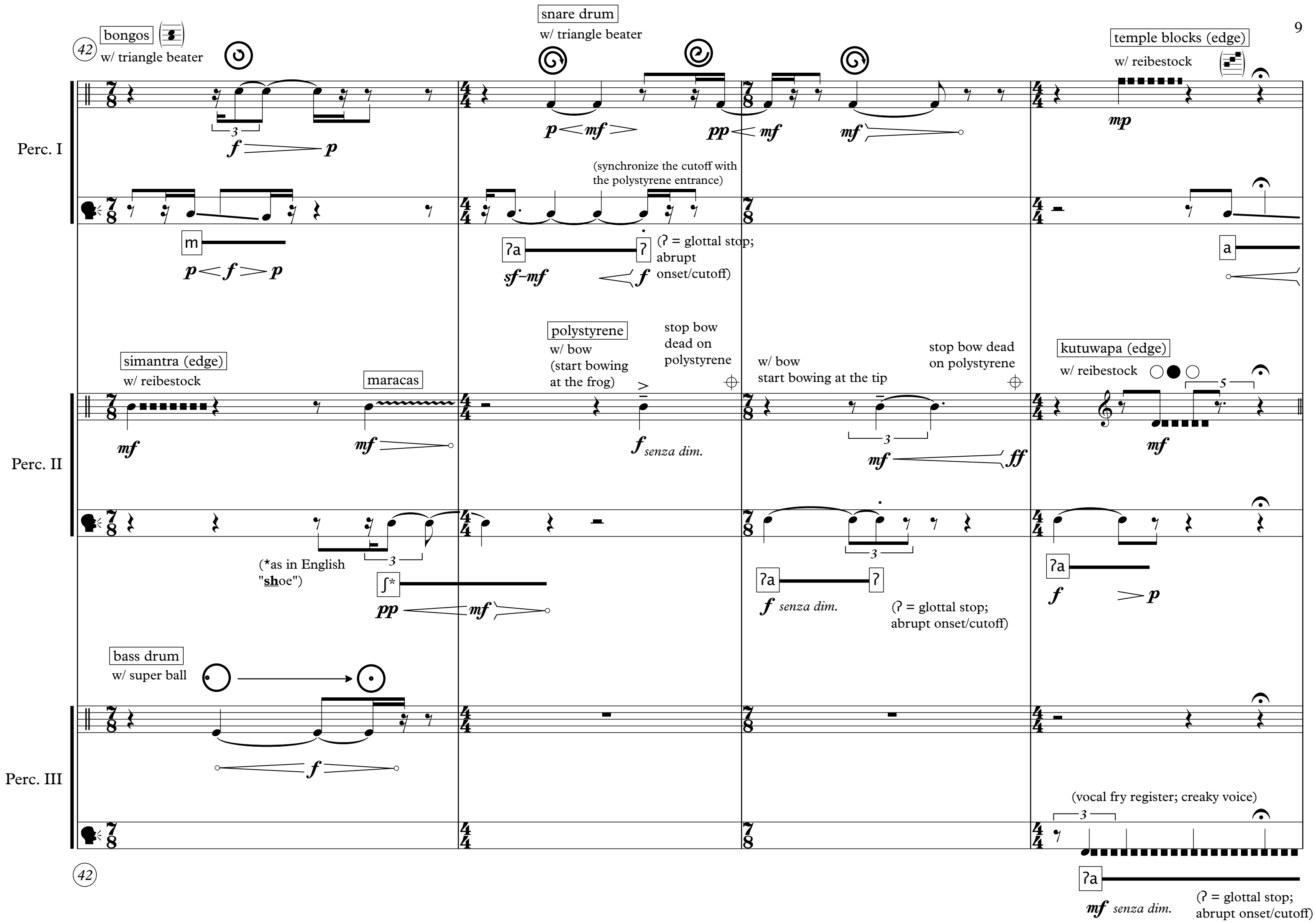
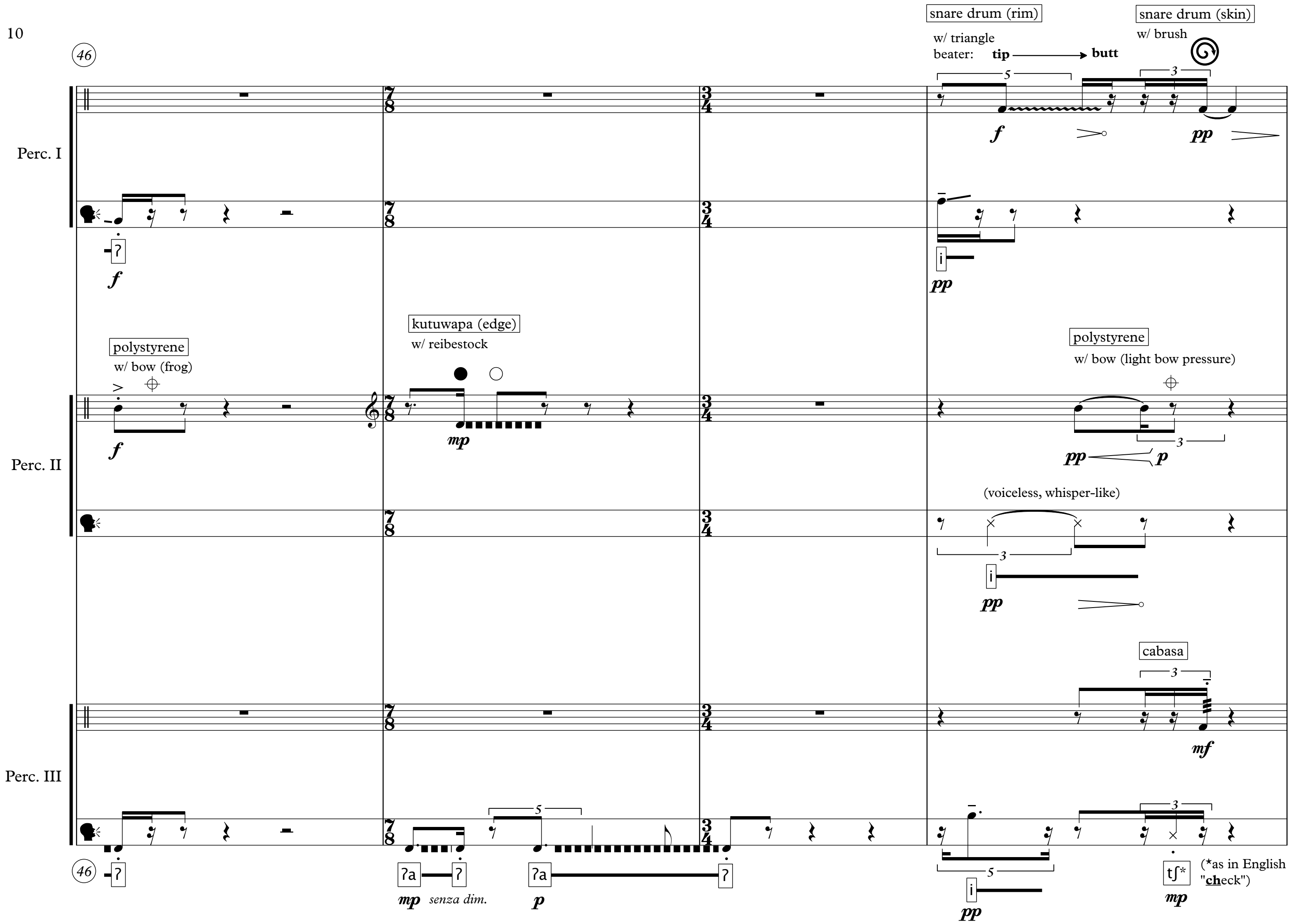

15
bongos (2)
snare drum
w/ triangle beater
w/ superball mallet
w/ triangle beater
w/ triangle beater
Perc. I
stones
(irregularly)
stones
simile
(voiceless)
(voiced)
Perc. II
temple blocks (edge)
w/ reibestock
bongos
w/ superball
tenor drum
w/ superball
(voiceless)
start from a comfortably low pitch and gliss. down as low as possible transitioning into the vocal fry range
(vocal fry register; creaky voice)
Perc. III
(*as in English "call")
15


20
temple blocks (edge)
w/ reibestock
temple blocks (edge)
w/ reibestock
bongos (2)
snare drum
w/ superball
snare drum
w/ triangle beater
5
Perc. I
v
a
kutuwapa
w/ rubber mallet
ad lib.
l.v.
Perc. II
r
tenor drum
w/ superball
bass drum
w/ superball
bongos
w/ 2 superball
mallets
Perc. III
20

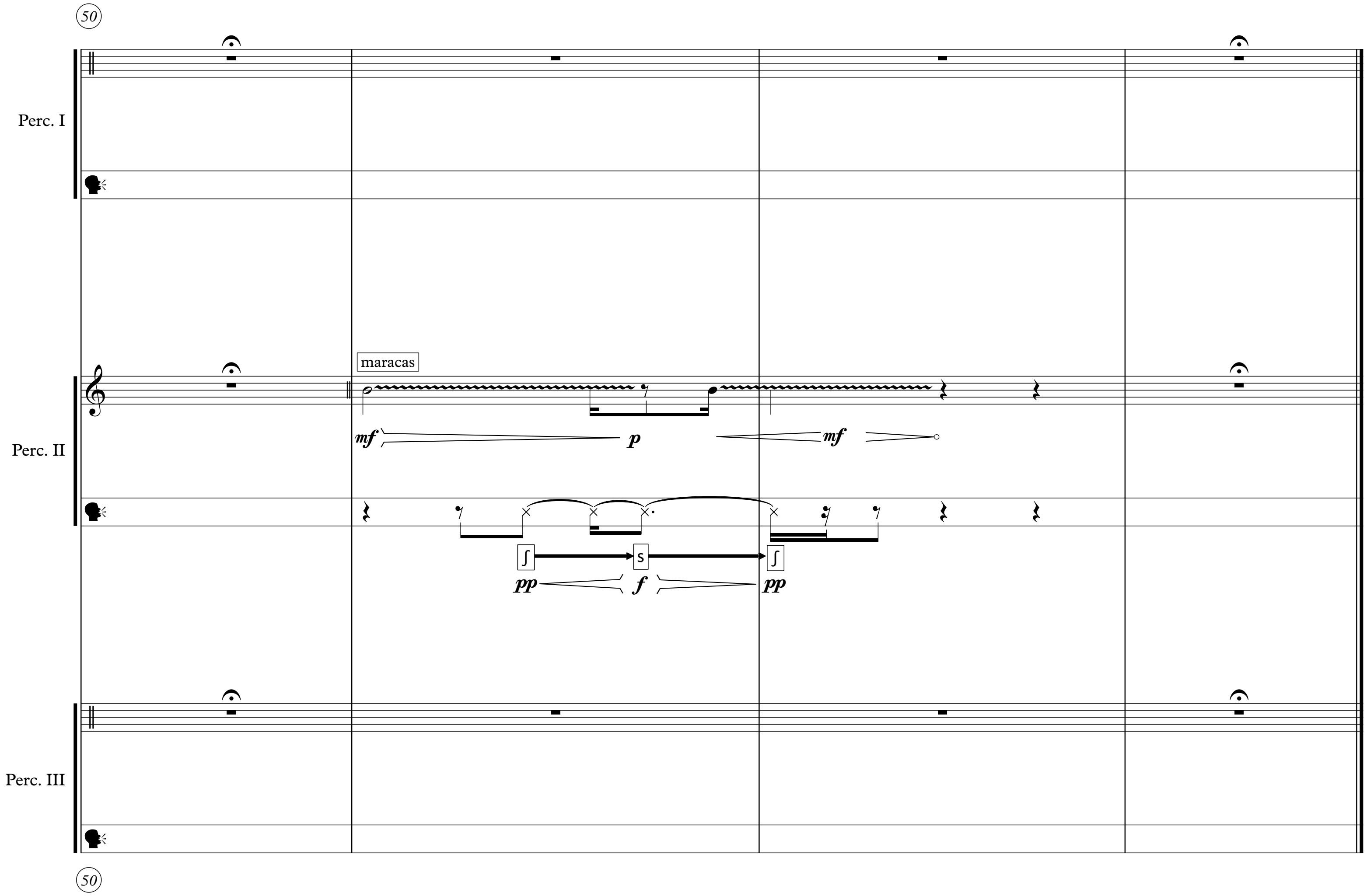

25
w/ mallet shaft
flexatone
w/ bow
5
3
f
pp
ppp
p
mf
Perc. I
x*
(*as in German "acht")
ye
rin
w/ bow
stop bow
dead on rin
l.v.
(w/ bow)
mp
mf/p
Perc. II
ʃ*
(*as in English "shoe")
tenor drum (rim)
w/ triangle beater:
tip
butt
cabasa
sfz
mfz
ff
Perc. III
ə
e
i


30
bongos (2)
snare drum
w/ 2 triangle beaters
snare drum (rim)
w/ triangle beater
tip
butt
snare drum
w/ triangle beater
Perc. I
f
mf
ff
p
a?
sfz
e
pp
simantra (edge)
w/ reibestock
maracas
Perc. II
p
f
pp
r
fpp
mfz
hi
tenor drum (skin)
w/ brush
cabasa
Perc. III
f
pp
mf
ʃ*
(*as in English
"shoe")
e
p
30


36
Perc. I
Perc. II
Perc. III
rin
w/ bow
bongos
w/ triangle beater
boobams
w/ drumstick
rin
w/ drum stick
simantra (edge)
w/ reibestock
chinese gong
into the
water bucket
læ*
r
(*as in English "ban")
bongos
w/ triangle
beater
temple blocks (edge)
w/ reibestock
w/ mallet shaft:
butt
tip
36


42
bongos
w/ triangle beater
snare drum
w/ triangle beater
temple blocks (edge)
w/ reibestock
Perc. I
f p
p mf
pp mf
mf
mp
(synchronize the cutoff with
the polystyrene entrance)
m
p f p
ʔa
ʔ
sf–mf
f
(ʔ = glottal stop;
abrupt
onset/cutoff)
a
simantra (edge)
w/ reibestock
maracas
polystyrene
w/ bow
(start bowing
at the frog)
stop bow
dead on
polystyrene
w/ bow
start bowing at the tip
stop bow dead
on polystyrene
kutuwapa (edge)
w/ reibestock
Perc. II
mf
mf
f senza dim.
mf ff
(*as in English
"shoe")
ʃ*
pp mf
ʔa
ʔ
f senza dim.
(ʔ = glottal stop;
abrupt onset/cutoff)
f p
bass drum
w/ super ball
Perc. III
f
(vocal fry register; creaky voice)
ʔa
mf senza dim.
(ʔ = glottal stop;
abrupt onset/cutoff)
42


46
Perc. I
Perc. II
Perc. III
snare drum (rim)
w/ triangle
beater: tip → butt
snare drum (skin)
w/ brush
polystyrene
w/ bow (frog)
kutuwapa (edge)
w/ reibestock
polystyrene
w/ bow (light bow pressure)
(voiceless, whisper-like)
cabasa
senza dim.
(*as in English "check")


50
Perc. I
Perc. II
Perc. III
maracas
mf
p
mf
pp
s
f
pp
50