SONGS OF LOVE AND DEATH: THE ACT OF SOUND AS NOTATION

Gil Dori
ENTI-UB
gil.dori@enti.cat

ABSTRACT

Songs of Love and Death is a composition for voice and electronics, which revolves around vowel sounds - real and implied. It features a real-time graphic score, which uses mouth contour data to recreate the shapes of vowels on the screen. Although they are represented graphically, these recreated vowels imply the sound that is tied to each one of them. The performer sings the text, imitating the given vowel shapes, and thus reshaping the sounds of the sung text. Through that, the work explores unusual acoustic results that emerge from the contrast between the implied vowel sound and the text needed to be sung. This approach to action-based notation incorporates aspects of audio scores into a visual one. The paper describes the composition and performance of *Songs of Love and Death*, discussing this notion of implied sounds, or acts of sound, as the driving principle of the notation.

.

1. INTRODUCTION

1.1. Action-Based Visual and Audio Scores

One of the types of real-time scores concerns notating gestures and action. The underlying notion of such action-based notation is to enable quick, accurate interpretation on the spot, through clear, immediate instructions [1]. Often action based scores are notated graphically, using visual symbols to represent gestures, but some utilize audio messages to convey performance instructions and information of musical gestures [2, p. 2].

Visual action-based scores rely on sound morphology. That is the gestural structure of sound manifested in time, an inherited physical interrelation between visible gesture and sound [3]. This allows the performers to instantaneously and intuitively interpret the score by imitating the given action. When coupled with a gesture data approach to notation, it produces results that are completely idiomatic to the instrument [4]. Imitation is also an important interpretation method of audio scores. However, as opposed to the visual score, audio score imitation embraces performance possibilities that fall outside the idiomatic practice, producing results that would not be habitually performed [2, pp. 4-5].

The composition discussed in this paper, *Songs of Love and Death* (2025), uses an action-based visual graphic notation. It features actions on the screen, recreated in real-time directly from gesture data, which the performer needs to imitate. It also incorporates the audio score notion of eschewing idiomatic performance practice. It aims to be intuitive as well as to produce unconventional performance and acoustic results. While not featuring actual audio instructions, it utilizes implied sounds, or the act of sound, to achieve this outcome. This is done in the context of vowels, which lend themselves naturally to the idea of implied sounds.

1.2. Audio-Visual Perception of Vowels

Vowels are speech sounds made with an open vocal tract, without significant restriction of the air flow, and with specific tongue and lips position that define the sound through filtering. This can be measured through spectral analysis, showing the formats that make the vowel sound the way it is [5]. Vowels, then, are an act of sound that presents a distinct morphology. Only seeing the act of producing vowels implies the sound of the vowel. As opposed to the understanding of the morphology of an instrument, which requires specific training, the morphology of vowels is understood naturally and intuitively.

McGurk and MacDonald, in their seminal paper *Hearing lips and seeing voices*, illustrated our innate ability to interpret vowels by both audio and visual information [6]. Other researchers have further studied this phenomenon, dubbed the McGurk Effect. While results vary, depending on the vowels and languages, the audio-visual perception of vowels is clear [7]. Conrad even determined that the difference in lip reading between deaf and hearing 15-year-olds is negligible [8], and Havy et al. demonstrated that not only both toddlers and adults can recognize words visually, but adults can also even learn new words this way [9].

Songs of Love and Death, in a way, is a live musical performance of the McGurk Effect. Its notation shows the mouth shapes and the text, each with its implied vowel

sound. But the acoustic result is a fusion of both. Or, in audio terms, considering the vowels as filtered sound, it is akin to convolution, in which the characteristics of an implied sound is used to alter the produced sound.

2. COMPOSITION PROCESS

2.1. Key Elements

Songs of Love and Death is a composition for voice and electronics, with a real-time graphic score. The score was made in Processing. Some of the audio material was made in Pure Data, and gesture data acquisition was done with FaceOSC.

The three elements–the signer, the synthesized accompaniment, and the notation–are all united through vowels. The composition is defined by the relationship between the act of producing vowels and the acoustic results of the actual produced vowels. It explores the contrast between the implied vowel sound and the text needed to be sung. The vowels used are the five most basic ones: open front unrounded ⟨a⟩, close-mid front unrounded ⟨e⟩, close front unrounded ⟨i⟩, close-mid back rounded ⟨o⟩, and close back rounded ⟨u⟩ [10].

The singer, could be any type of range, is instructed to sing a given text from various poems about love and death. The text is fragmented, and presented as isolated short phrases, words, and individual syllables. The score shows the text one line at a time. The text is fixed, and is introduced in the same order, so it can be learned in advance. The graphic notation indicates relative pitch, dynamic, and most importantly, mouth shape to imitate. When singing the text, the singer must produce its vowels according to the given mouth shape (figure 1).

The electronics accompany the singer with synthesized vowels and fricatives, utilizing formant synthesis and source-filter modeling respectively. Initially I tried to have all the audio synthesis happen in real-time and have it done in Processing, but in the case of the vowels it yielded poor results, both aesthetically and functionally. Instead, I synthesized the vowels in Pure Data, with added vibrato and reverb, recording four sound files for each of the five vowels in variation of pitch and duration. The audio files, then, were called up by the score algorithm when needed, and, in some cases, played at different rates to alter the pitch. In the composition, the synthesized vowels function as sustained choir-like pads, with long attack and release. In contrast, the synthesized fricatives are more abrupt, and in some cases, they are played all together randomly, as a rapidly changing texture. They are synthesized in Processing in real-time. Four types of fricatives are synthesized: voiceless labiodental ⟨f⟩, voiceless dental ⟨θ⟩, voiceless alveolar ⟨s⟩, and voiceless postalveolar ⟨ʃ⟩ [Ibid.]. Overall, the electronics accompaniment is sparse, and it is used mostly to emphasize certain moments in the piece.

The most prominent component of the score is the vowel shape indication. This is the act of sound which informs the voice production of the performer. Essentially, an animated mouth that changes according to the vowel selected by the score algorithm. I used FaceOSC to record mouth shape data of each of the five vowels into text files. Then, in Processing, I recreated a mouth with the same points as in FaceOSC, and that is capable of the same movements. For this process I made use of the FaceOSC Receiver Templates repository by Wilcox et al. [11]. Finally, when a vowel is selected by the program, it reads the data from the corresponding text file, and the mouth changes its shape accordingly (figure 2).

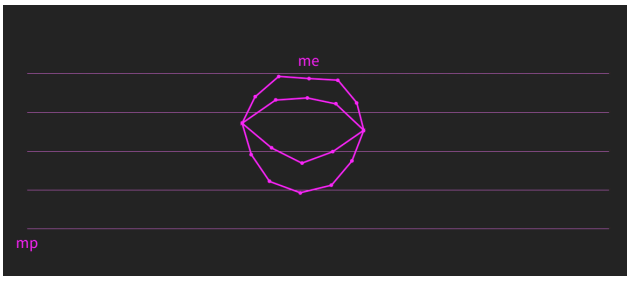

Figure 1. *Songs of Love and Death* score, showing the mouth shape indication, text to be sung, pitch and dynamic indication.

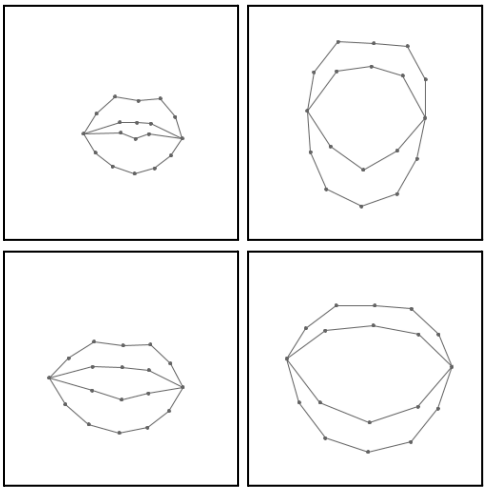

Figure 2. Preliminary mouth shapes recreated in Processing from FaceOSC data.

2.2. Score Design

In the score itself, the animated mouth also moves vertically, to indicate the pitch. A higher position corresponds to a higher pitch, and vice versa. While pitch is indeterminate, the mouth is placed over a five-line staff (without a clef), and it moves within the common singable range from slightly below the bottom line to slightly above the top line (figure 3). Dynamics are indicated by the color of the background; the brighter it is, the louder the level, and the darker it is, the softer the level. Dynamics are also indicated by traditional letter markings, below the staff on the left (figure 4). Initially, the score did not include the staff and dynamic markings, but they were added during the testing process to facilitate performing the piece. The purple color of the whole score was also proved to be the most visually clear,

taking into consideration the changing brightness of the background.

The text is placed directly above the mouth, which helps with visual focus. The performer is instructed to sing only when text appears. Rests are indicated by a closed mouth in the center of the score. Just like lines of text, the placement of rests is also fixed. The duration of each line and rest, though, may vary. There are three fixed long rests that mark the endings of sections. Overall, there are four sections to the piece, defined by text grouping. With each section, the text becomes more fragmented.

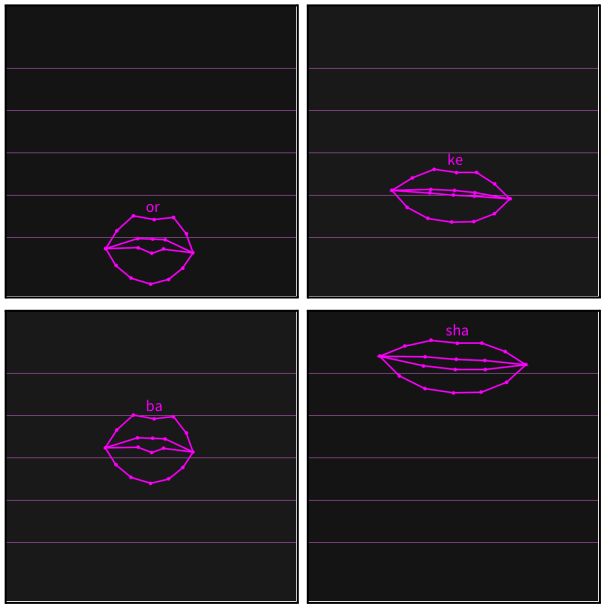

Figure 3. Pitch indication and range, from lowest (top left) to highest (bottom right).

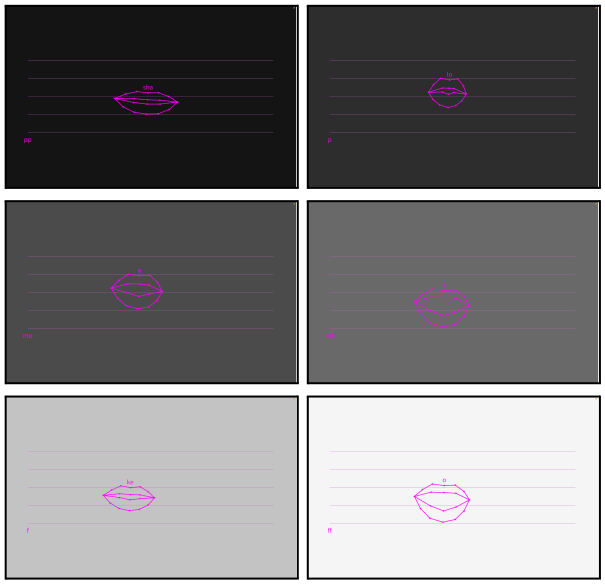

Figure 4. Dynamics indication and range, from *pp* (top left) to *ff* (bottom right).

The score is controlled automatically by an algorithm that selects which vowel shape to display, as well as which synthesized vowel to play, including in what pitch. The selection is done similarly, using first-order Markov chains, but according to different inputs. In the case of visual display and audio playback, vowels are selected according to the text. The selection of pitch is done according to pitch tracking of the performer's input. In addition, an envelope follower is used for smoother control over the electronics' amplitude. However, in various places the automatic selection is overridden in favor of fixed compositional decisions, to better shape the piece.

2.3. Songs of Love and Death in Practice

Bringing the piece onto the stage was done in two phases. The first was a preliminary test performance that took place in a small venue, performed by a soprano. It allowed better preparation for the second phase, which included three performances in concerts of a more traditional, official context by a baritone singer. A video recording of the performance is available online [12].

The graphic score itself proved intuitive and easy to follow, especially in the second phase, after the rounds of testing and improvements during the first phase. In the second performance, the singer also printed the text, to learn and anticipate it better, which further improved the performance.

In both phases, the main challenge was working with the singers on how to approach vocal production. Eventually, the approach that worked the best was to extremely exaggerate the mouth shape given by the score when singing the text, and to consciously prioritize the acoustical result over the text, even up to not attempting to make the text intelligible at all. While both singers naturally understood the morphological aspect of the implied sound instructions, that is the act of sound given by the animated mouth on the screen, applying it into their performance practice required a lot of effort. In this way, the score functions closely to an audio score, rather than a visual one. Not only that the performers needed to embrace possibilities and outcomes that are foreign to their idiomatic vocal practice, but they also had to completely abandon the way they usually present text in song.

3. CONCLUSIONS

Songs of Love and Death was well received by the public, and most importantly, by the performers. Both singers concluded that performing the piece was a valuable experience worth the effort.

The composition successfully shows that an act of sound, or implied sound, approach can be beneficial within the context of visual action-based score. On the one hand, it is intuitive to understand and follow. On the other hand, it embodies audio score notions of achieving unfamiliar acoustic outcomes.

This approach to notation may prove relevant to real-time scores in general, inviting further artistic and research work using similar procedures. Future development possibilities may include having an additional performer performing the act of sound live, directly controlling the score; adding an artificial intelligence vowel detection tool, for more natural algorithmic decision making; and conducting research into audience perception of this type of notation.

4. REFERENCES

1. S. Shafer, "Performer Action Modeling in Real-Time Notation," in *Proc. Int. Conf. on Technologies for Music Notation and Representation – TENOR'17*, Universidade da Coruña, 2017, pp. 117-123.
2. S. Bhagwati, "Elaborate audio scores: Concepts, affordances and tools," in *Proc. Int. Conf. on Technologies for Music Notation and Representation – TENOR'18*, Concordia University, 2018, pp. 24-32.
3. T. Wishart, *On Sonic Art*. Routledge, 1996.
4. G. Dori, "Using gesture data to generate real-time graphic notation: A case study," in Proc. Int. Conf. on Technologies for Music Notation and Representation – TENOR'20, Hamburg University for Music and Theater, 2020, pp. 68-74.
5. S. A. Fulop, *Speech Spectrum Analysis*. Springer Science & Business Media, 2011.
6. H. McGurk, and J. MacDonald, "Hearing lips and seeing voices," *Nature*, 264, pp. 746–748, 1976, doi: https://doi.org/10.1038/264746a0
7. S. Shigeno, "Influence of vowel context on the audio-visual speech perception of voiced stop consonants," *Japanese Psychological Research*, vol. 42, no. 3, pp. 155–167, 2000, doi: https://doi.org/10.1111/1468-5884.00141
8. R. Conrad, "Lip‐reading by deaf and hearing children," *British Journal of Educational Psychology*, vol. 47, no. 1, pp. 60–65, 1977, doi: https://doi.org/10.1111/j.2044-8279.1977.tb03001.x
9. M. Havy, A. Foroud, L. Fais, and J.F. Werker, "The Role of Auditory and Visual Speech in Word Learning at 18 Months and in Adulthood," *Child Development*, vol. 88, no. 6, pp. 2043-2059, 2017, doi: https://doi.org/10.1111/cdev.12715
10. International Phonetic Association, *Interactive IPA Chart in English*. Accessed on: January 15, 2026. [Online]. Available: https://www.internationalphoneticassociation.org/IPAcharts/IPA_charts_TI/IPA_charts_TI.html#eng
11. D. Wilcox, K. McDonald, G. Levin, A. Carabott, D. Moore, and S. Gruber, *FaceOSC Receiver Templates*. Accessed on: January 17, 2026. [Online]. Available: https://github.com/CreativeInquiry/FaceOSC-Templates/tree/master
12. G. Dori, *Songs of Love and Death*, 2025. Accessed on March 15, 2026. [Online]. Available: https://youtu.be/dbAC6nXrnVI