The Influence of Text in Computer Music Composition: An Approach to Expression Modeling

Juan Reyes(1)
Maria P. Muñoz (2)
(1)MOX- Computación Avanzada en Ingenieria / Departamento de Artes
Universidad de Los Andes
jreyes@uniandes.edu.co
(2)Universidad Javeriana
libroabierto@hotmail.com
Santafe de Bogota - Colombia

Abstract

For centuries there has been a very close relationship between sound and text. Virtually every culture in each epoch has tailored one or more distinctive vocal tradition.This paper journeys on an analytical and rather technical way on how text influences music. It deals with the basic notion of lyric composition and its relationship with phonetics, rhythm, and its transcendence with pitch. We focus on the sound of the voice and how it behaves in syllables and phrases by creating different spectra depending on the emotion and intention of its sound. Various hints about different mixtures and sound blending can be drawn apart by using spectral analysis by means of the FFT. These can be further parameterized into a model for signal processing, synthesis and spectral or physical modeling. Finally, we give substantial weight to text semantics, and approach our subject from a cognitive standpoint in order to take advantage on how we can extract expression models to trigger synthesis out of the relationship among lyrics and music.

Scope about Composing with Text

We assume that phonetically oriented music can benefit from the semantics of preconceptions; words create images or imitations of surroundings and therefore, the listener finds a territory on which the music is portraying. A poem communicates and expresses facts as objects or affections as a result of grammar rules semantics and intonation. In a relationship between poetry and music the only boundaries are given by the common grounds of rhythm and sound, namely phonetics, pitch and durations; This provides composers with the option of working with concepts suggested by the meaning of words and text manipulation in addition to direct harmonic and sound manipulation (Berger 1994). Phonetics give the advantage of having accent and phrasing in a melodic line. It contrasts the sound of vowels by expanding or contracting them with other sounds. Consonants determine a flexibility of positioning these strikes on downbeats or upbeats. As a function of outlining text, the relationship of melody and text is moreover in terms of pitch, intensity, duration, and timbre [2]. This provides hearing words phonetically and hence hearing tones simultaneously. Furthermore, the meaning of words provides a clue to what is symbolized by the music (Meyer 1956). Therefore, if the subject implied in a word suggests feelings of sadness or happiness, the composer might opt for choosing the interval direction, tempo, the harmony or the mode of the piece, upon word suggestions (Hill and Kamenetsky 1996).

COMPUTER MUSIC AND TEXT: ANALYSIS TECHNIQUES

Most of the techniques applied to speech recognition can find their way to computer music composition on which text plays a crucial role on the piece. In this context, speech analysis by means of signal analysis has proven to be the most fruitful method. Sinusoidal models for analysis and synthesis, characterize the amplitudes, frequencies and phases of the component sine waves. These parameters are commonly estimated by means of the Short Time Fourier Transform, in a peak picking algorithm. These peaks are then assigned to a frequency track (McAuley and Quatieri 1986). In this approach, we have found very useful to work in terms of a basic unit proposed by Rapoport in 1997 (Rapoport 1997), referred as the "basic pulse" and used in his research to Opera and Lyric singing. The basic unit is a neural command to the vocal folds , consisting of tightening followed by immediate tension releasing. For our purpose levels of emotion can be expressed in terms of the unit pulse by measuring its degree of excitement in the FFT. We have constrained these results to duration, energy or intensity, legato and vibrato. A tone or a vowel is composed of one or more unit pulses. Computer Music techniques in our experiments as well as compositional work include the Phase Vocoder (Dolson), SMS (Serra 1996), or Spectral Modeling Synthesis (Arcos,Mantaras and Serra 1997), and LPC or Linear Predictive Code (Lansky 1989)s. For sound editing we have used conventional editors like EditSound on NextStep and Pro-Tools on Mac (see Figure - 1).

ANALYSIS OF TEXT

Our work was inspired on the analysis from a poetic standpoint (Muñoz 1998), and in particular the semantic rhyme and phonetic standpoints of poem XCIII of Emily Dickinson (Dickinson). In particular we were looking for clues that will suggest rhythm, sound and rhyme. We were looking for syllables and their accent weak or strong. Also the duration of each syllable which consequently dictates the intonation and intention of the word. Later the phrase and the mental object created from a perceptive or generative grammar standpoint. We thought of rests

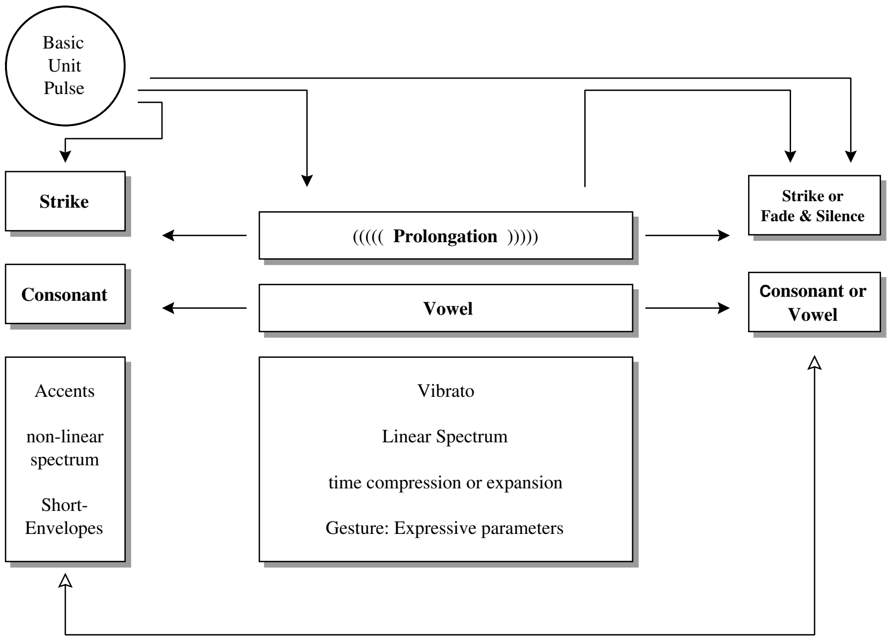

Figure 1 - Text Analysis

as breath points giving contrast to sounds and rhythm. Consonants determine a flexibility of positioning syllables on downbeats or upbeats.Thus, we can say that the syllable is our basic unit and depending on its degree of expressivity it might consist of one or more unit pulses or frames of the FFT. Through phrase analysis by means of spectral analysis and the Fourier Transform we isolated hints that suggest musicality. Poem XCIII reads as follows:

A SEPAL, petal and a thorn
Upon a common summer's morn
A Flash of dew, a bee or two,
A breeze
A caper in the trees,-
And I'm a rose

By isolating the vowels in our analysis, we see that in the first line of the poem, there is a combination of A's and E's and we give more weight to the A's which will contrast at the end with the sound of one O. There are three monosyllabic words which contrast with two words each containing two syllables, therefore there are different durations among all the syllables of the phrase. We suggest that the monosyllabic words have longer durations and thus having more expressive parameters. The second line gives us an accented use of the vowel O in combination with the letter N and a subtle combination with the A and the E. We think that the peak of the line exists at the end of the word "summer's" because of the contrast between the E's and the O's. Line third is more dazzling that the rest of the lines because it contains good combination of most vowels in monosyllabic words. It gives a lot of freedom to the duration and intonation of each word and thus we regard it as the line with more musicality. We can further analyze the rest of the lines of the poem in the same way but paying special attention to the rather asymmetric durations and the off beat contrast of the last line with the rest of the poem. We will like to give more accent to the word "rose" being the essential word of the poem from a semantic point of view although we think is hidden in a weak beat from the rhythm view. In the perception point of view we see that the lines of the poem each one generates a sort of mental image by giving clues to the reader in order to guess what kind of object the writer is suggesting. At the end as readers, we timidly get the image of a flower with the given description.

TEXT AND MUSICALITY

The combinations of vowels give clues for musicality. If a vowel is accented and if it is more emotional, we can suggest that its pitch will go up. The higher the gap between intervals the more excitement the syllable gets. But the longer the duration of the vowel the more expressive parameters it gets. In longer durations more attention to the quality of the tone, namely the use of different formants and timbre, can be paid. In a musical context vibrato can be applied. Also very important are the connections between words. Some might have a silence but some don't. In the later case legato might be applied.If there is some silence a consonant is between silence and vowels. For our rhythmic purpose consonants strike on downbeats or upbeats giving the sense of syncopation or not. The addition of tonality and rhythmic stress to the words of this poem might suggest different states of mind depending on the intervals being used. Better yet by using a major or minor system and time meters, generative grammars might suggests us affections of joy or sadness which were not explicitly outlined in the poem.

Once we get this sort of analysis automatic or manual, we can edit or apply synthesis or re-synthesis parameters to our extracted sounds or vowels. These can be applied by the use of ordinary sound editors in order to change the place and repetitions of the words and consequently changing the semantics of the poem. Likewise by the use of the Phase Vocoder, LPC or Spectral Modeling, intonation, intensity and durations can be altered. At this point our concern is the overall generation of physical objects or concepts that the text might suggest and how they might influence the composition of a musical piece. For this we would start from our FFT analysis followed by editing, signal processing or synthesis techniques.

COMPOSING WITH TEXT

We start from the rules of syntax that specify a particular sequence of phonological sounds in a sentence or a noun and how they might affect the semantics and perception of the sound of the words. Subsequently we outline or hide concepts that the words carry out. Also by changing the the prosodic structure of words, we are transforming the metrical structure of the words. These sort of effects are similar to the different accents from people of different regions in a country speaking the same language which in our case are accounted as performance differences. If we establish repetition of syllables or words in our composition we are establishing a pattern. These sort of repetitions create tensions by concentrating on sound instead semantics and meaning of words.Our aim is that the listener will only conceptualize words when they will be separated from the musical analysis (Lerdahl and Jackendoff 1985). If none of the above work for our text in question, we can begin again from the fact that a word and its semantic meaning provides a clue to the kind of emotion we want to apply to the musical tone (see Figure 2).

It is important to outline that text by means of syllables, words and phrasing suggests different emotion and performance parameters when it is contextually blended with music. Similarly any sort of emotion suggest meanings which can influence the kind of expression that we want to apply to the various unit pulses of a musical phrase. In these sense, music can be used as imitation in its domain of what is being said with spoken or written language. Our Fourier techniques further provide means of detecting parameters previously programmed as expressive. They include dynamic range, duration, spectral content and vibrato of the syllable a word or a phrase. These variables on the time domain are sufficient for a performance of a musical line. In longer phrases the energy value of consonant is required in order to identify the downbeats of a phrase. By transforming these values we can get a musical gesture out of a timbre of the voice which is rich in formant and partials as well as the rhythmic or repetition pattern of our melody. This heuristic acknowledges that an audio text signal carries expressive parameters that are very useful compositionally. The only constrain is that the meaning of text is mandated either by the semantics or by the grammar of the language of the text by the reader, the composer or the listener.

EXPERIMENTS AND RESULTS OF COMBINING TEXT AND MUSIC

Analysis and synthesis experiments on a compositional context were performed by the authors on original material written by Maria P. Muñoz. The parameters to be found, extracted and manipulated included basic unit pulse in syllables, duration, intensity of the sound, excitation formants, pitch change on a single syllable, word transitions and connections, and consonant envelopes. For

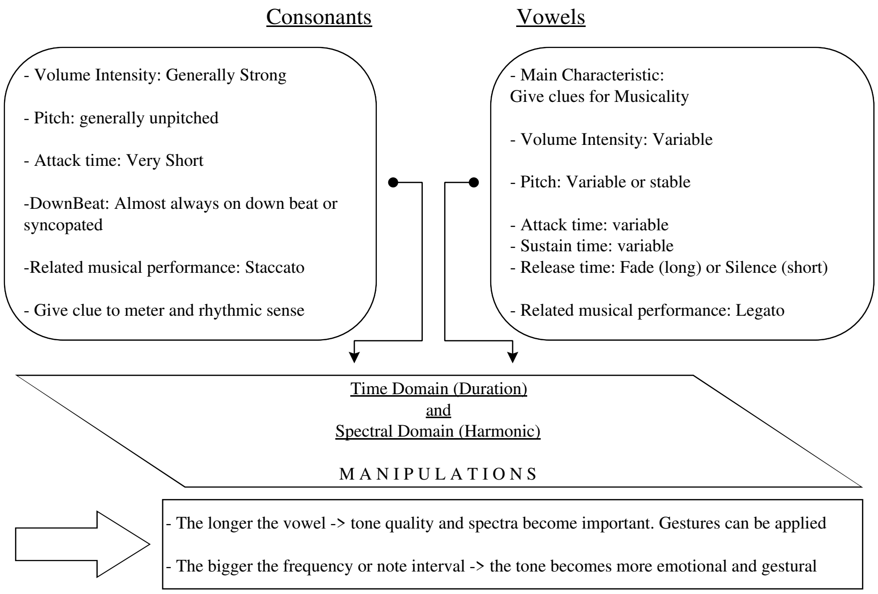

Figure 2 - Composing with Text

synthesis the constrains or rules applied to the gestural line included duration transforms, upward intervals if syllable was accented and if more emotion was to be applied, downward intervals if a phrase or chain of sounds needed to be resolved, more partials and formant region transform if the vowel was accented, vibrato if duration of a vowel exceeds 160 ms, different silence if there were consonant between words, and legato if two vowels are joint together. Most of our experiments were done on the Spectral Modeling Synthesis package written by Xavier Serra. Sound editing was performed manually on conventional sound editors. Some convolution with different impulses such as bells, flutes and strings were performed to the text audio signals. LPC and filter manipulation was also performed in order to harmonized the text. A fraction of the text of the poem in spanish is as follows. (we will provide a table with the different transitions and transformation for transforming these phrases into a musical line).

Debuto Beatriz
la gran actriz
Beatriz es feliz
mete polvo por la nariz
fuma hachis y toma agua de anis.

Most of the expressive manipulation as suggested from the text was performed on the syllable "iz" which in english is composed of the vowel "e" and the enveloped sound of the "z" or the "s". The manipulation of these unpitched sounds were done by the stochastical parameters provided by SMS in which case we can divide the sound of the syllable "iz" by a deterministic part or the vowel "e" and a residual part delimited by the duration and the position in the time domain of the letters "s" or "z". The tempo was originally provided by the poet while digitizing the audio signal. With the constrains referenced above, the results heard showed resolution or downward intervals at the end of each line; upward intervals fills between the syllables in the first words of each line. In the last line there is a further upward interval and accented longer duration in the letter "y" using the vowel "e" therefore getting a spectral change plus vibrato. In the poem, the length of such consonants like "m" or "n" was also prolonged by using a combination of the "o" or "e" spectra with a few higher partials and high energy on the lower partials giving the characteristic nasal sound of those consonants.

Further experiments in the musical context were done from the cognition point of view. By using the sound editor we were able to exchange vowels positions in some cases affecting grammatical rules of the spanish language. By this we were able to change the tense of the verbs. Perception in this cases changes the purpose of the subject by creating a different set of possibilities. We also change the position of the words by subsequently getting from different variations and versions to a totally different poem. And finally we tested different melody and harmonizations over the same text. In every single variation there was a different connotation and a different way of perceiving the melody. Unless the meaning of the text was previously analyzed, the listener will parallel different meanings with the same text. In a more informal environment we extended these experiments to widely known popular songs and we were able to obtain similar results. The meaning of text tends to be masked by melodic or harmonic progressions giving us the opportunity to assert that the emotional content and meaning of a poem phrase with music is encoded inside the musical gesture from a perceptual focus.

CONCLUSIONS

On our research we found that the FFT is stable on working with the vowels of the human voice. Therefore we can safely state that text can be analyzed with the Fourier transform in order to obtain various sets of information apart from spectra and time domain information. By a combination of duration and energy level of the spectra we can obtain the unit pulse. A syllable is composed of one or more unit pulses. The more unit syllables the more expressiveness the syllable might have and the more expressive parameters it might contain. Consonants give attacks and strikes on downbeats or upbeats. Because of its unpitched character we analyzed them as white noise or residual noise with also different levels of energy and duration. Once analyzed, text can be manipulated by proven signal processing techniques. By synthesis of an analyzed sound we can manipulate expressive parameters such as vibrato, intensity, legato or sound with no consonants and durations or even number of unit pulses (sort of granular synthesis). The musical results can be translated into a different way phrasing, accents, time compression or expansion.

Furthermore we now acknowledge that by manipulating text in by the techniques referenced above, the perception of text adds an additional layer above conventional function of phonetics in text. In many cases new processed sounds are perceived distinctly from spoken language and even by lyrics of a song. In this the sound of the word carries a gesture full of expressive information. This sort of result was acknowledged by these authors and several students by experimenting with the expression on dogs and cats where there are not known semantical rules for a communicating language. In a musical context a combination of phonetics of this sort and sound gives the powerful combination that is usually perceived as a song.Our expectection with these techniques suggests that some composers might get into experimenting with their own set of phonological rules in order to achieve a high level of expressivity.

FURTHER WORK

A Joint collaborative work with the Electrical Engineering at La Universidad de Los Andes and these authors has started to do further analysis of text signals as well as bird singing. The aim is to achieve prosodical results in particular to birds from different regions in Colombia. The extensions to our work include signal analysis by means of various wavelet transforms and neural nets. From the synthesis standpoint there will also be some experimenting with chaotic filtering and signals.

REFERENCES

Arcos J.L., de Mantaras R.L., and X. Serra, 1997, *"SaxEx: a Case-Based Reasoning System for Generating Expressive Musical Performances",* in Proceedings of 1997 International Computer Music Conference, pp 329 - 336.

Berger, A. *"Music as Imitation"*, in Perspectives on Musical Aesthetics, New York NY, USA, W.W. Norton & Company Inc.

Hill D.S., S.D. Kamenetsky, 1996, *"Relations Among Text, Mode, and Medium: Historical and Empirical Perspectives"*, Music Perception, Berkeley, CA, USA, University of California Press.

Lansky, P., 1989, *"Compositional Application of Linear Predictive Coding"*, in Current Directions in Computer Music Research, Cambridge Mass, USA, MIT Press.

Lerdahl, F. and R. Jackendoff, 1985, *"A generative Theory of Tonal Music"*, Cambridge Mass, USA, MIT Press.

McAulay, R.J. and Quatieri T.F., 1986. *"Speech Analysis/Synthesis based on a Sinusoidal Representation"*. IEEE Transactions on Acoustics, Speech and Signal Processing. 34(4): 744-754

Meyer, M. 1956, *"Emotion and Meaning in Music"* , Chicago, Illinois, University of Chicago Press.

Muñoz, M.P., 1998, Dissertation Thesis: *"Lo Sonoro en la Poesía de Emily Dickinson"*, Bogota, Colombia, Universidad Javeriana.

Serra, X. 1996. *"Musical Sound Modeling With Sinusoids Plus Noise"*; Phonos Foundation, Pompeu Fabra University.

Rapoport, E., 1997 *"Singing, Mind and Brain - Unit Pulse, Rhythm, Emotion and Expression"* in "Music, Gestalt, and Computing Lehman Marc ed." Berlin, Germany, Spinger-Verlag.

Dolson, M., *"Fourier-Transform-Based Timbral Manipulations"* in Current Directions in Computer Music Research, Cambridge Mass, USA, MIT Press.

Dickinson., E., *"The Collected Poems of Emily Dickinson"*, New York U.S.A. Barnes and Noble