Voice Separation in Polyphonic Music: Information Theory Approach

Michele Della Ventura

HAL Id: hal-01821044
https://inria.hal.science/hal-01821044

Submitted on 22 Jun 2018

Voice Separation in Polyphonic Music: Information Theory Approach

Michele Della Ventura

Music Academy "Studio Musica", Department of Technology
dellaventura.michele@tin.it

Abstract. Voice Separation is a delicate stage in a music information retrieval process intended to be used in the automated music analysis processes through textual segmentation or for the indexation of a music score. This article presents a method that is capable of separating polyphonic music, considered in its symbolic aspect, into its individual parts (or voices). This method considers every single note as an individual entity and assigns it to the part (or voice) where the information content that it assumes in relation to the already-existing notes of the same score is maximum. The algorithm may separate the voices identifying them even in the points that intersect. The algorithm was tested against a handful of musical works that were carefully selected from the repertoire of Bach and of Mendelssohn.

1   Introduction

Computerized music analysis is an constantly-changing discipline, both thanks to the consequences originating from technology research and to the ongoing influences deriving from the study of cognitive sciences. Focusing on the comprehension of the brain processes involved in the musical activities [1][2], the latter may assume a decisive role in the analysis of the computerized music analysis.

This kind of approach is important to musicians and to IT developers alike. To musicians it provides an innovative means to obtain the composing elements of a composition (pre-thematic elements, themes/melodies, rhythmic cells,...) [3][4][5][6]. To IT developers, these techniques are used to perfect the text segmentation systems in order to improve the capacities of a search engine in the identification of the information on the web [7]. The amount of information on the web is growing fast as well as the number of new users lacking experience in the art of web searching. The automated search engines based on the correspondence of keywords usually deliver too many low-quality matches. In general, the more precise the collection of data to be analyzed is, the better the results of the elabora-

tions will be. The starting point is therefore the reading of the notes that is carried out from a MIDI file or from the more recent XML file.

In case of a polyphonic composition (figure 1), which entails a simultaneous ensemble of several voices on several pitches and in parallel or opposite directions, data (sound) reading and their proper assignment to a certain voice rather than to another voice becomes vital: both in case of the analysis of the musical text and in case of the information retrieval by the search engine.

Figure 1. Polyphonic composition

The analysis of a score is usually carried out by the main algorithms through a process of segmentation that considers the sounds of the various voices, one after another (figure 2).

Figure 2. Representation of the score for the segmentation process

In polyphonic music and, in particular, in the FUGUE form, there are many examples where a voice intersects another voice (figure 3).

Figure 3. Overlapping musical notes.

In these cases an incorrect reading of the data would lead the segmentation processes (which precede the analyses described above) to rather inaccurate and, in certain cases, even erroneous results. In the example from figure 4 the "subject" of the FUGUE is no longer recognizable.

Figure 4. Representation of the score for the segmentation process

In recent years various methods have been proposed for the separation of the voices of a polyphonic composition.

Some works include algorithms which are capable of separating the voices based on specific preset rules [9][10][11]: the computer does not learn these rules and develop its own knowledge, but rather simply uses the knowledge provided by the human analysis.

In other cases the algorithm is able to autonomously learn the rules [12][13][14]: parameters that are trained on music already labeled with voice information.

Some algorithms [9] take on the problem of overlapping musical notes, even if the results are not always satisfying. These are systems that take into consideration the sound vicinity concept. A marked improvement is the study by Gray and Bunescu [15], a neural model for voice separation in symbolic music that assigns notes to active voices using a greedy ranking approach, or the study by Guiomard-Kagan [16], which taking inspiration from the method of Chew and Wu [9] that defines what contigs to connect and how to connect them, improved these connections by using musical parameters such as the average pitch difference between neighbor contigs.

This article presents a method that, drawing inspiration from the preceding studies, is able to reconstruct the various voices of a polyphonic composition, by reading the sounds from a MIDI file: every single note is considered as an individual entity and inserted into the voice where the information content that it assumes in relation to the already-existing sounds in the same score is maximum. It is a system based on two clearly distinct principles: a musical one, related to the structure of the thematic material (interval structure - distance between various sounds) and a mathematical-statistical one, derived from the Information Theory elaborated by Weaver and Shannon [8], tied to the possibility of transition from one interval to the other.

This paper is organized as follows.

Section 2 describes the information theory. Section 3 describes the analysis of the musical message. Section 4 shows some experimental tests that illustrate the effectiveness of the proposed method. Finally, conclusions are drawn in Section 5.

2 Information Theory

The analysis based on the Information Theory, considers the audio message as a linear process endowed with a syntax formulated not on the basis of preset rules, but on the probability of occurrence of each element of the audio message in relation to the element preceding it [17][18]. From the definition of a "message" as a chain of discontinuous speech "units of meaning" [17], there follows that the speech "units of meaning" coincide with the minimum events of an audio message. Any event of a chain built in this fashion requires a prevision in relation to the event that will follow it [19]. In a communication happening by means of a given alphabet of symbols, the information is associated to every single transmit-

ted symbol [17]. Information, therefore, may be defined as *the reduction of the uncertainty that might have been, a priori, present on the transmitted symbol*. The ampler the range of messages that the source may transmit (and the larger the uncertainty of the receiver in relation to the possible message), the larger the quantity of transmitted information is – and, along with it, its own measure: *the entropy* [19]. In the Information Theory, the *entropy* measures the quantity of uncertainty or of information existing in a random signal. If every message has the probability p_i of being transmitted, the entropy is obtained as the sum of all the set of functions $p_i \log_2 p_i$, each of them being related to a message, i.e.:

$$H(X) = E\big[I(x_i)\big] = \sum_{i=1}^{n} I(x_i) \bullet P(x_i) = \sum_{i=1}^{n} P(x_i) \bullet \log_2 \frac{1}{P(x_i)}$$

3 The voice separation model

To separate the voices of a score starting from a musical input, the sounds must be, first of all, divided into movements, according to the beat of the composition (figure 5a), thus obtaining chords (i.e. the simultaneous overlapping of several sounds). Bear in mind that every voice may be even made up by several sounds within the same movement (figure 5b). The assignment of the notes to the various voices is therefore performed in chronological order, from the left to the right, beginning with the first movement: every note of the first movement is considered in itself as sound of a voice.

Figure 5. Subdivision of the scores into movements.

Subsequently, the algorithm determines all the possible musical segments obtainable by combining the sounds of a voice with the sounds of the other voices of the movement following the analyzed one (figure 6).
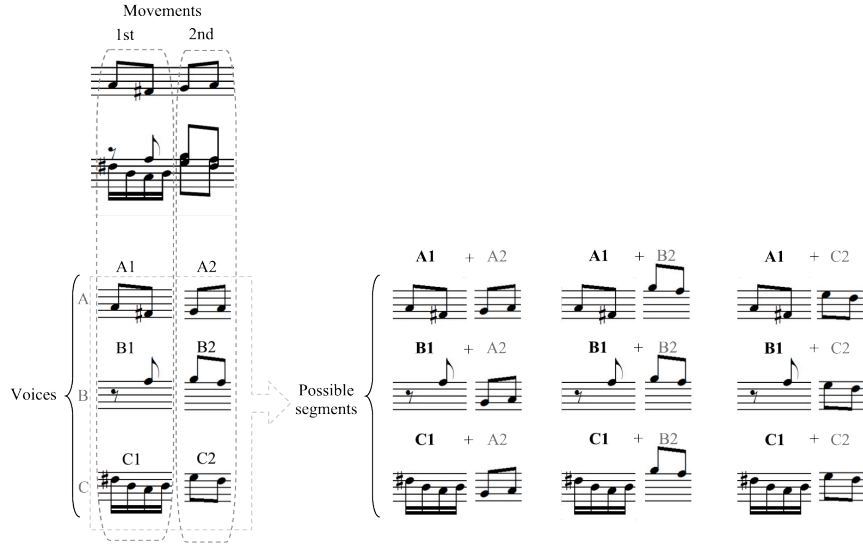
Figure 6. Possible segments between two consecutive movements.

Finally, the information value is calculated for every identified segment: the segment having the larger value will represent the succession of the sounds for the specific voice (figure7).
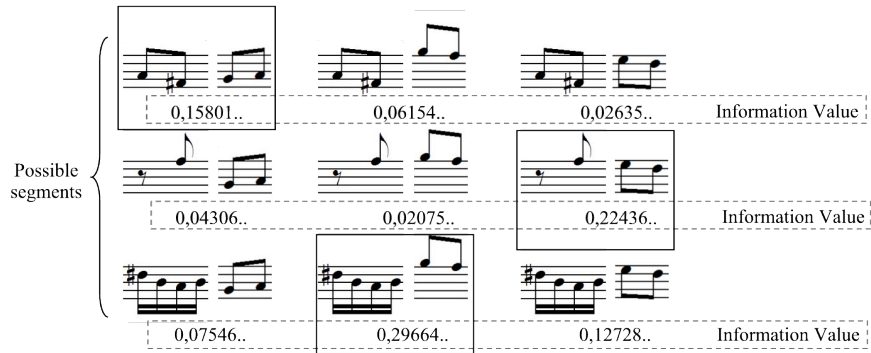
Figure 7. Constitution of the voices.

To compare various segments among them, in order to determine which is more important, each entropy is calculated: the less the entropy value, the greater the information carried by the sound [18]. In order to calculate the entropy it is necessary to take into consideration a specific alphabet: the alphabet is language – specific [19] and, as it may be immediately deduced from the formula (based on the probability of certain symbols rather than other symbols to be transmitted) it

demonstrates to be associated to language.

For the **melodic analysis** the various melodic intervals were classified as symbols of the alphabet [18]. For every single musical piece a table, which represents its own alphabet, is filled in: in the melodic analysis every interval has been considered for its ascending or descending trend (Table 1). A peculiarity of the proposed method is related to the fact that only the sounds of the initial subject of the FUGUE are taken into consideration to define the alphabet table.

Table 1. Example of alphabet.

Sound	Number
Fa	1
Mi	1
Re	1
Do	4
Si	
La	3
Sol	
Fa	2

The definition of the alphabet is not enough in order to calculate the entropy of a musical segment: it is necessary to consider the manner in which the sound succeeds one another inside the musical piece.

In order to do this, the Markov process (or Markov's stochastic process) is used: we chose to deduce the transition probability that determines the passage from a state of the system to the next uniquely from the immediately preceding state [14] [18]. On the base of the above considerations, the transition matrix is created. It consists of the transition probabilities between the states of the system (conditional probability). In our case, the matrix represents the probabilities for a sound to resolve to another sound (Table 2).

Table 2. Transitions matrix drawn about the melodic segment of Table 1.

		Fa		La		Do		Re		Mi		Fa	
		a	d	a	d	a	d	a	d	a	d	a	d
Fa	a					1							
	d				1								
La	a		2										
	d					1							
Do	a			2									
	d								1				
Re	a												
	d									1			
Mi	a												
	d											1	
Fa	a					1							
	d												

4   Obtained results

The method presented in this article was tested by analyzing some musical compositions written in the form of a FUGUE, carefully selected from the repertoire of Johann Sebastian Bach (48 fugues of the well-tempered clavier) and of Felix Mendelssohn (6 fugues op. 35). Both cases regard polyphonic compositions that include from 2 to 5 voices and some of them even have overlapping musical notes. The algorithm, realized on purpose, does not provide any limitation with respect to the dimensions of the table representing the alphabet and the matrix of transitions that will be automatically dimensioned in every single analysis on the basis of the characteristics of the analyzed musical piece. This allows conferring generality to the algorithm and specificity to every single analysis (Strength). Furthermore, the algorithm does not entail any limitation as to the number of possible voices to be extracted (Strength).

The analysis of the time needed for every single elaboration was not taken into consideration inasmuch as the objective of the study was solely to separate the voices of a polyphonic composition to further improve potential subsequent segmentations for analyses of a different nature: musical or web indexation.

The larger the number of decimals used to express the information content value of every segment, the more precise the results of the elaborations. The larger the number of decimals, the smaller the error risk.

Some examples of the obtained results are given below (figure 8 and 9) as well as examples related to the specific moments in which the voices were intersecting. The first stave shows the subject of the FUGUE, presented in the first 2/3 beats; the second stave shows a segment of the FUGUE where the parts intersect; the third stave shows the final result with the various separated voices.

Figure 8. J.S. Bach: fugue No.6 of "The Well-Tempered Clavier" (BWV 851).

Figure 9. F. Mendelssohn: fugue No. 2 op. 35.

5   Discussion and Conclusions

This article described an approach to the separation of the voices of a polyphonic musical composition, considered on their symbolical level, based on the Information concept.

The algorithm was applied to musical compositions in the form of FUGUE by authors such as Bach, Frescobaldi and Mendelssohn.

The results show on the one hand how the musical fabric of a composition is characterized by a strong structural uniqueness; on the other hand how this method represents a solution for the problem of voice separation, even if the voices intersect.

This method represents an alternative approach in the applications of computational methods to the voices separation problem: the high degree of complexity of musical phenomena imposes certain forms of achievement that must be adequate and that, for completeness' sake, must cope with the problems under a sufficiently large number of angles.

References

1. I. Deliège e S. McAdams, 1989, *La musique et le sciences cognitives*, Liegi-Bruxelles, P. Mardaga Editeur.
2. G. De Poli et al., 1991, *Reti neuronali per il controllo delle deviazioni temporali nell'esecuzione musicale*, in Camurri A. e Canepa C. (a cura di), *Atti del IX Colloquio di Informatica Musicale*, Genova, 13-16 novembre 1991, Genova, AIMI/DIST.
3. M. Pancini, *Problematiche e modelli formali per la segmentazione automatica/interattiva di partiture musicali simboliche*, Rapporto Tecnico CNR-PFBC-MUS-TR, 2000.
4. M. Della Ventura, *Analysis of algorithms' implementation for melodic operators in symbolical textual segmentation and connected evaluation of musical entropy*, in *Proc. 1st Models and Methods in Applied Sciences*, Drobeta Turnu Severin, 2011, pp. 66-73.
5. M. Della Ventura, *Rhythm analysis of the "sonorous continuum" and conjoint evaluation of the musical entropy*, in *Proc. Latest Advances in Acoustics and Music*, Iasi, 2012, pp. 16-21.
6. M. Della Ventura, 2015, *The Influence of the Rhythm with the Pitch on Melodic Segmentation*, In: Abraham A., Jiang X., Snášel V., Pan JS. (eds) Intelligent Data Analysis and Applications. Advances in Intelligent Systems and Computing, vol 370. Springer, Cham, pp. 191-201.
7. K. Hui, K. Berberich, 2015, *Selective Labeling and Incomplete Label Mitigation for Low-Cost Evaluation*, In: Iliopoulos C., Puglisi S., Yilmaz E. (eds) String Processing and Information Retrieval. SPIRE 2015. Lecture Notes in Computer Science, vol 9309. Springer, Cham, 137-148.
8. W. Weaver and C. Shannon, *The mathematical theory of information*, Urbana, Illinois Press, 1964.
9. E. Chew and X. Wu, 2005, *Separating Voices in Polyphonic Music: A Contig Mapping Approach*, In *Computer Music Modeling and Re- trieval: Revised Papers*, Esbjerg, Denmark.
10. I. Karydis, A. Nanopoulos, A. Papadopoulos, and E. Cambouropoulos, 2007, *VISA: The Voice Integration/Segregation Algorithm*, In *ISMIR*, Austria, 2007.

11. S. Madsen and G. Widmer, 2006, *Separating voices in MIDI*, In *ISMIR*, Canada, 2006.
12. P. Kirlin and P. Utgoff, 2005, *Voise: Learning to seg- regate voices in explicit and implicit polyphony*, International Conference on Music Information Re- trieval(ISMIR).
13. A. Jordanous. Voice Separation in Polyphonic Music: A Data-Driven Approach. In *Proceedings of the Inter- national Computer Music Conference*, Belfast, Ireland, 2008.
14. R. de Valk, T. Weyde and E. Benetos, 2013, *A Machine Learning Approach to Voice Separation in Lute Tablature*, In Proceedings of the 14th International Soci- ety for Music Information Retrieval Conference, pages 555–560, Curitiba, Brazil.
15. P. Gray, R. Bunescu, 2016, *A neural greedy model for voice separation in symbolic music*, In International Society for Music Information Retrieval Conference +(ISMIR 2016), pp. 782–788.
16. N. Guiomard-Kagan, 2017, *Traitement de la polyphonie pour l'analyse informatique de partitions musicales*, Informatique [cs], Université de Picardie-Jules Verne.
17. E. Angeleri, *Information, meaning and universalit*, UTET, Turin, 2000.
18. F. Lerdhal and R. Jackendoff, *A Grammatical Parallel between Music and Language*, New York, Plenum Press, 1982.
19. J.J. Nattiez, "Fondements d'une sémiologie de la musique", Union Générale d'Éditions, Paris, 1975.