Symbolic Textural Features and Melody/Accompaniment Detection in String Quartets

Louis Soum-Fontez, Mathieu Giraud, Nicolas Guiomard-Kagan, Florence Levé

Symbolic Textural Features and Melody/Accompaniment Detection in String Quartets

Louis Soum-Fontez[1], Mathieu Giraud[2], Nicolas Guiomard-Kagan[1], and Florence Levé[1,2]

[1] Université de Picardie Jules Verne, MIS, F-80000 Amiens, France
[2] Université de Lille, UMR CNRS 9189 CRIStAL, F-59000 Lille, France
florence.leve@algomus.fr ⋆

Abstract. Music is often described as melody and accompaniment, and several MIR studies try to identify melodies. But the organization of voices is not limited to such a distinction between melody and accompaniment: *Textural effects* – such as repeated notes, syncopes, homorhythmy, parallel moves or imitation – underline the melody/accompaniment layout, and changes in texture usually mark structural transitions in music. We investigate how textural and other characteristics can help to identify melodic voices in polyphonic music. We select *measure-level features* to analyze symbolic scores of string quartets, including new *textural features*, and propose models to predict, on each measure, *melodic and accompaniment layers* in such scores, each layer possibly including several instruments. We evaluate these sets of features and the models on 12 movements in Haydn and Mozart string quartets. The best models have an average accuracy of more than 85%, taking into account both statistical and textural features.

1 Introduction

Melody, as the foreground of a musical material, is complex to define and characterize. In *string quartets*, the first violin (Vln1), as a leading instrument, often plays the main melody. However, the three other instruments of the quartet – second violin, viola, and cello – can also join the melody or play it alone over time (Figure 1).

Melody Detection. Research on melody extraction is an active field in the audio domain [23, 8, 14]. In the symbolic domain, studies investigated the melodic content of monophonic phrases and patterns through the lens of melodic similarity [31], melodic segmentation [1, 19, 29, 30], and contour analysis [25, 24].

Concerning the particular question of identifying the melody in a polyphonic score, Uitdenbogerd and Zobel [28] proposed several algorithms identifying the melodic line in polyphonic MIDI files, including the simple *skyline algorithm* that labels as melody the highest pitch at each onset. Rizo et al. proposed a set of statistical descriptors extracted from each track of MIDI files of different music styles (classical, jazz and pop) and trained a random forest classifier to identify melody tracks in these pieces [22].

⋆ This work is partially funded by French CPER MAuVE (Région Hauts-de-France).

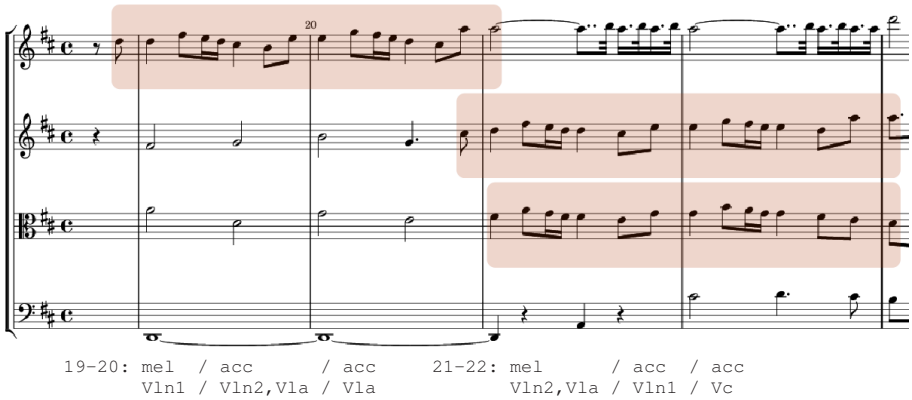

Fig. 1. Haydn, Quatuor op. 33/1, I, mes. 19-22. The texture is described for each measure – even if the melodies are not stricty aligned on measures boundaries. On mesures 21-22, the role of the first violin may be debated, but the main melody is played on the second violin and on the viola, mostly in parallel move in sixths.

Madsen et al. proposed an algorithm for predicting melody notes at any point of the piece, based on a sliding window rendering the complexity of the musical lines [17]. One limitation is that they assume that there is only one melodic line at a time and they reported that the skyline algorithm was still better performing on one Haydn string quartet. They also use this method to identify the melody track in two datasets of popular music [16]. These first results show that assessing complexity may help the recognition of the melody. Friberg et al. tried to recognize the main melody in a polyphonic symbolic score on ringtones of popular music [7], using Huron's perceptual principles [11]. Some of the features they use are derived from symbolic data but intend to model audio features, such as timbre, staccato/legato, or sound level.

Texture and Melody. The role of *texture* has long been recognized in music theories [15, 13], but systematic, formalized, or computed analyses of texture remain few. In 1960, Nordgren quantified some aspects of the orchestral texture [20]. In 1982, Rahn, discussing the melody identification in polyphonies, argues that *a melody "stands out" from its accompanying parts largely on the basis of its complexity* [21]. In 1989, Huron discussed the semantics of the term "texture" and proposed measures to evaluate the textural diversity of music [10].

Several studies focused on the segregation of polyphonic music voices or *streams*, as a listener might perceive them [2, 3, 26]. Duane's thesis [6] further proposed to characterize texture in string quartets by grouping the notes in streams perceived by listeners and by characterizing the *role* of these streams. He described three roles: main lines (including melodies), secondary lines and accompaniment. He established by statistical methods that the perception of textural flows was mainly related to note synchronicity, coordinated pitch modulation (especially parallel movements), as well as the presence

of certain harmonic intervals (metricity, rhythmic repetition, rhythmic patterns, melodic contour, and harmony do not seem to have a significant role).

We previously proposed to describe texture with *layers*, each one determined according to its role and qualified according to its composition [9]. At the first level, layers are mainly qualified as melody, accompaniment, or other minor roles. One melodic layer may include several instruments, and there can be two melodic layers. At the second level, layers can be tagged as repeated notes, syncopation, sustained notes, imitation, and homorhythmy, which can be refined by a complementary descriptor in the case of parallelism, unison, or octave.

Outline. Several textures are more specifically used for rhythmic or accompaniment parts (as repetitions, homorhythmies, or syncopes) or are indicators of some relationship between two voices or more. However, no MIR studies have linked such textures to the analysis of melodic and accompaniment layers. Our goal here is to improve the analysis of textures in symbolic scores, notably the melody/accompaniment detection, focusing on *string quartets*, where melody is often taken by other instruments than the first violin. We aim to improve the understanding of interactions between instruments and the changes in texture by determining melodic and accompaniment layers more precisely. We select measure-level features to analyze symbolic scores of string quartets, gathering existing features [22], and new *textural features* (Section 2). We introduce models to predict melodic and accompaniment layers based on such features (Section 3). We evaluate these sets of features and the models on a set of 12 movements in Haydn and Mozart string quartets and discuss these results (Section 4).

2 Measure-level Features for Melody Detection

To predict whether a measure is melody or accompaniment, we use the following set of features computed on each measure.

2.1 Voice Name (4)

These features enable the (baseline) skyline algorithm, considering that the top voice is the melody.

– (voice-name): 4 binary features, activated depending on the voice (first violin, second violin, viola, cello)

2.2 Statistical Features (20)

The features introduced by Rizo et. al were used to predict, *on the whole piece*, which track is the melody between all tracks [22]. They are linked to music properties that can make what is a melody or what is an accompaniment (see Section 4.2). We computed here these features *on each measure*. They are grouped in 5 categories: track information (normalized duration, number of notes, occupation rate, polyphony rate),

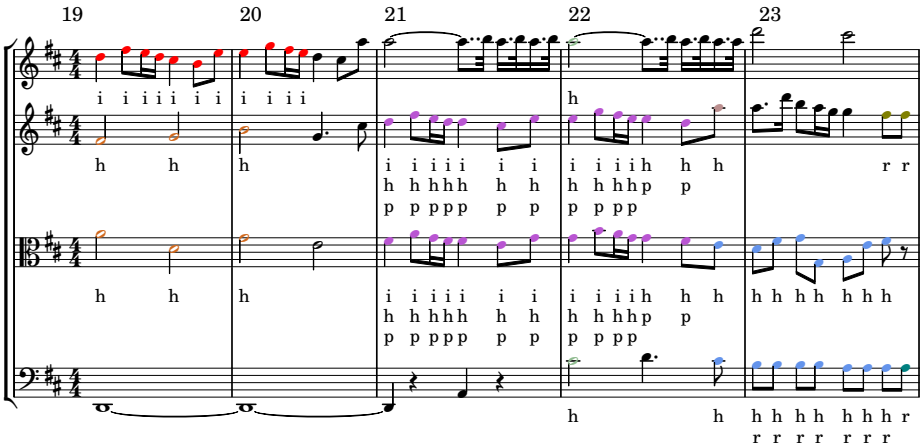

Fig. 2. Detection of individual textures in measures 19-23 of String Quartet op. 33-1, 1st movement by Haydn (see also Figure 1). The colors show the related voices for textures i/h/p.

pitch (highest, lowest, mean, standard deviation), pitch intervals (number of different intervals, largest, smallest, mean, standard deviation), note durations (longest, shortest, mean, standard deviation), and syncopation (number of syncopated notes). We also added the number of repeated notes.

2.3 Textural Features ($7 \times 16 + 1$)

To further describe the music texture, we propose a new set of high-level features describing the organization of notes and voices. The taxonomy of [9] introduced several textures but proposed an algorithm only for homorhythmic layers. Inspired by this taxonomy, we design here the following binary features, that can be computed *on every note*:

- repeated notes (r): We consider as repeated notes sequences of at least three successives notes of the same pitch and the same duration, for a total duration of at least one beat and a half, possibly spaced with rests of at most one beat.
- syncopes (s): A note is considered as a syncope if it starts on a weak beat or on a second half of any beat and continues on at least the next beat.
- homorhythmy (h): Two voices are considered as homorhythmic when they play only notes starting and ending at the same time during at least three beats.
- parallel moves (p): Two homorhythmic voices are considered in a parallel move when at least three close pairs of notes have the same diatonic interval – generally thirds, sixths, or octaves or unison.
- imitation (i): We consider as imitation the repetition of a pattern on some voice, called the original pattern, by another voice with some delay. This can be seen as a parallel move delayed in time. This is computed by a simplification of the Mongeau-Sankoff algorithm [18] requiring here five exact notes or more approximate matches.

– rest (rest): There is no note, but a rest.
– The feature (none) is added when none of the previous six features is activated with the given sixteenth.

On a given measure, these 7 features are actually computed on *each of the 16 sixteenth notes* – the considered corpus being in 4/4, see next section. For each sixteenth note, a vector gives the features that are activated, taking into account an expansion rule – that is including notes that are still sounding but not attacked there. Moreover, the following summarizing feature is added on each measure:

– texture ratio: number of sixteenths on which at least one of the textures is activated, represented as a ratio between 0 and 1

Figure 2 shows an example of the heuristic detection of these textures. The prediction of repeated notes (r) (such as some eights in measure 23) and homorhythmy (h) is very reliable and would be here close to a manual annotation. The parallel move (p) is correctly identified at measures 21-22. The imitation pattern (i) at measures 19-20 on the first violin, that is later taken on the second violin and viola at measure 21-22, is also correctly detected. However, a manual annotation would probably not set the same boundaries for such annotations, for example by ending the homorythmy one note later on the measure 20.

3 Learning Models: Melody/Accompaniment Prediction as a Measure Classification Task

We see the melody/accompaniment prediction as a binary classification problem, given the features presented in the previous section. We choose the measure granularity to be consistent with the reference annotations. The statistical features were normalized into a gaussian distribution (0 ± 1) and almost all the textural features are binary. Theses vectors are gathered into a vector of maximal size $4 + 20 + (7 \times 16 + 1)$ for each measure, and a given melody or accompaniment class for the reference annotation.

3.1 Model Architecture

Two models were tested:

– A random forest (RF) classifier as used by [22], taking the average of a set of 200 decision trees trained on random subsets of features, where data are weighted to account for the unbalanceness of categories.
– A simple neural network (NN) with an initial dropout layers with a rate of 0.5 to reduce overfitting [27], 2 hidden fully connected linear layers (64 then 32), separated by *relu* activation layers and batch normalization layers, and a last layer, composed of a unique neuron, with a *sigmoid* activation function and a threshold of 0.5 between melody/accompaniment prediction. Weights were initialized uniformly. Batch size is 32 and the learning rate is 10^{-3}, with early stopping after 50 iterations without improvement. To estimate errors at each iteration, the loss function used is binary cross-entropy. The optimization of the gradients is done with Adam [12].

Haydn			m	% mel	% Vln1 mel
17.1	i	E Major	111	29.1	97.1
17.2	i	F Major	101	30.9	81.6
17.3	iv	Eb Major	70	34.1	73.1
33.1	i	B minor	91	31.2	72.6
33.2	i	Eb Major	90	34.9	93.5
33.3	i	C Major	165	32.2	100.0
33.4	i	Bb Major	83	35.7	98.3
50.1	i	Bb Major	164	31.4	76.2

Mozart			m	% mel	% Vln1 mel
No. 2, K. 155	i	D Major	119	37.0	96.5
No. 4, K. 157	i	C Major	126	45.6	80.8
No. 6, K. 159	i	Bb Major	71	53.2	98.2
No. 14 K. 387	i	G Major	171	29.5	86.7

Table 1. The corpus contains 12 movements of Haydn and Mozart string quartets, all in 4/4. The last three columns give the number of measures (m), the ratio of measures (on the 4 voices) labeled as melody, and the ratio of measures on the first violin labeled as melody.

4 Results

4.1 Corpus, Implementation, and Availability

The corpus includes 12 movements of string quartets by Haydn and Mozart (Table 1), totaling 1362 measures, as `.krn` files. We extended the corpus of our previous study [9], and we distribute the complete set of annotations as open data at `www.algomus.fr/data`. Each measure on each of the four instruments was labeled as *melody*, *accompaniment*, or *other*. Only measures with *melody* and *accompaniment* were taken into account, totaling 4791 measures. The files were processed with music21 [5], using actual pitch spelling (for example to compute intervals), and the learning models were implemented with keras [4]. The code is available at `www.algomus.fr/code`.

4.2 Statistics on the Features

Figure 3 shows the distribution of some features over the measures labeled as melody or accompaniment in the corpus.

As expected, the features on the pitch (highest, mean, and lowest), being very similar to (voice-name), are very significant to tell apart the melodic and accompaniment parts. More interestingly, other features play a significant role, such as (num-notes) (melody tends to have more notes) or (num-diff-int) (melody tends to use conjoint intervals). In textural features, imitation and syncopes are significantly associated to melody, whereas repeated notes are significantly associated with accompaniment. Homorhythmy and its subset parallel moves are found in both roles, but parallel moves are more used for melody.

4.3 Accuracy over a Leave-one-piece-out Strategy

We did not split these 4791 measures into a training set and a validation/test set: Indeed, having different measures of the same piece in different sets would bring overfitting due to repeats or similar sections inside each piece. Considering the relatively small size of

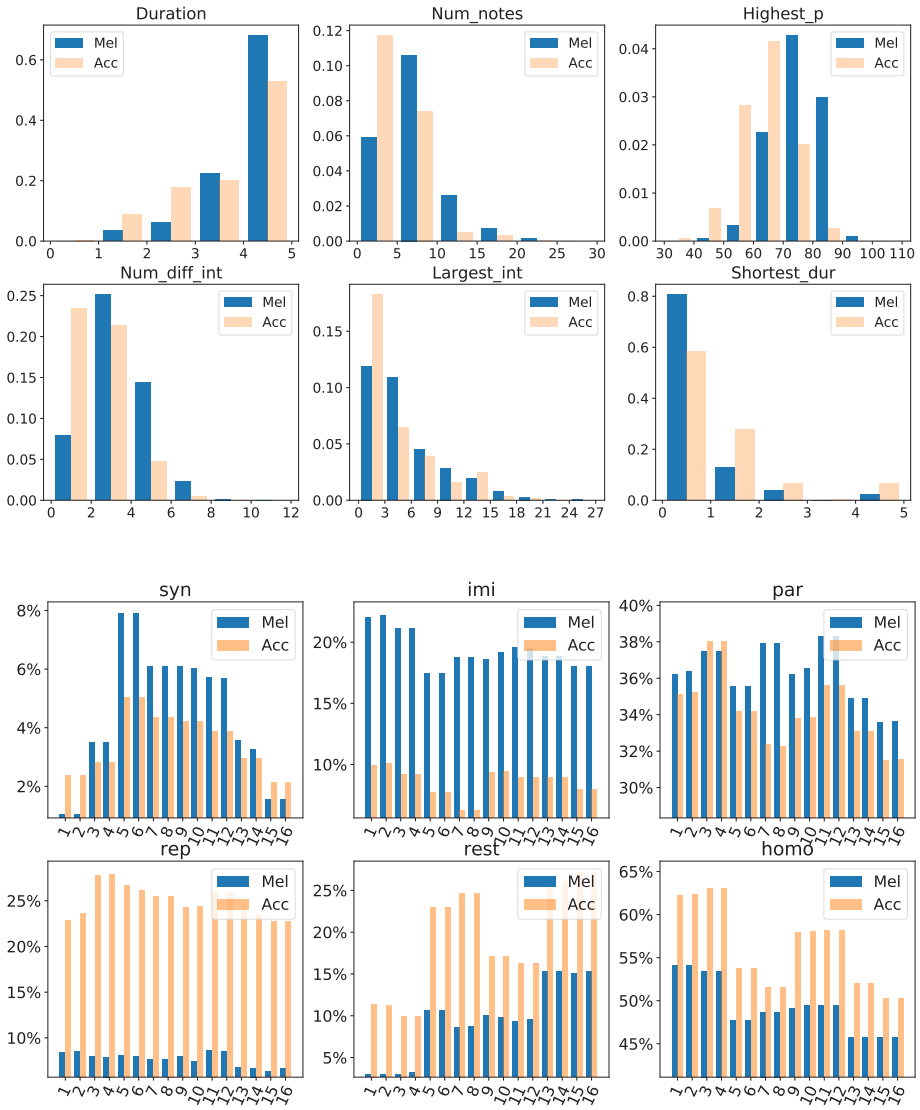

Fig. 3. Distribution of some features in the 12 movements of the corpus, split into the measures labeled as melody (Mel) or accompaniement (Acc) in the reference analysis. The data was normalized in order that each of the area equals to 1. Top two lines: Some of the statistical features introduced by [22]. Bottom two lines: Textural features on each sixteenth on each measure, temporal barplot representing the proportions of a given sixteenth having one of these textures.

	base		RF	NN	
Features	all	dm	all	all	dm
Majority	65.4	19.0	–	–	–
Top (Vln1)	**84.1**	0.0	–	–	–
Statistics	–	–	78.9	81.9	51.9
Texture	–	–	69.1	72.3	**52.8**
Statistics + Texture	–	–	80.3	**82.4**	49.9
V. Name	–	–	84.3	84.2	0.0
V. Name + Texture	–	–	83.1	**86.8**	26.6
V. Name + Statistics	–	–	83.3	84.1	25.4
V. Name + Statistics + Texture	–	–	84.3	85.3	**32.0**

Table 2. Mean accuracy of the baseline models (base), as well of the Random Forest (RF) and the neural network (MLP) on the leave-one-piece-out strategy and various sets of features, evaluated on *all* measures but also on *difficult* measures (dm) i.e. where the first violin is not playing the melody or another voice is playing it.

the corpus, we rather opted for a *leave-one-piece-out* strategy: Each piece is separately considered as a validation set of a model trained on all other pieces. We iterate and report the average accuracy over all the pieces.

As baseline models, we consider *Majority* (all measures are predicted as accompaniment) and *Top* (the first violin, as the top voice, is predicted as the melody – this is equivalent than only considering the (voice-name) feature). Table 2 shows that the best model is the NN taking into account the voice names and our proposed textural features, with about 86.8% of correct predictions. As expected, the (voice-name) alone has significant results – and this is confirmed by the baseline Top(Vln1), 84.1%. The statistical features, even alone, have a good performance, but include features on pitch that are very similar to (voice-name). Conversely, the textural features, without any feature related to pitch, manage alone to identify 72.3% of the measures, including 52.8% on difficult measures where the first violin is the voice playing the melody. Adding these textural features to (voice-name) improves the accuracy.

We call *difficult* measures the 15.8% measures where melody is not at the first violin or shared between several instruments. The best model here correctly predicts the melody in 32% of such measures.

4.4 Focus on Specific Cases

With the best model, the best results are on the first movement of Haydn 33.4 (96.1%), this movement having melodies almost always played on the first violin (see Table 1).

Conversely, Figure 4 details the prediction on five difficult measures in a Mozart quartet. On measures 27-29, the three voices are predicted as accompaniment, whereas the reference annotation labels as melody the second violin. Although the melodic patterns are about the same in measures 27, 30, and 31, it is worth noting that the prediction on measure 27 is wrong, whereas on the measures 30 and 31, the model correctly predicts the melody in parallel moves (p) between one of the violins and the viola: The textural information here helps in predicting the melody.

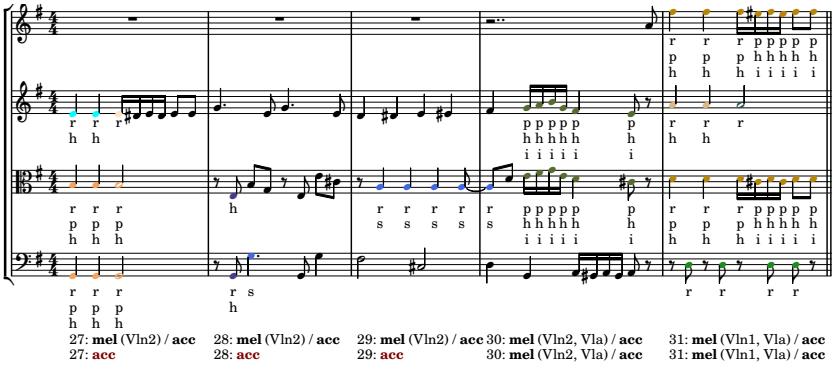

Fig. 4. Textural features, reference annotation (top), and mel/acc prediction by the model NN (bottom) on measures 27 to 31 from quatuor K387 by Mozart, first movement.

5 Conclusion and Perspectives

We evaluated sets of features and models to predict, on each measure, which instrument(s) is playing a melodic content. Experiments on string quartets by Haydn and Mozart show that some textural features are distributed differently in melodic and accompaniment parts, and that the best models detect some of the melodies beyond the first violin or distributed among several instruments. This brings a new step towards a general characterization of melody and texture in polyphonic pieces. Further studies could improve the features and the learning model, generalize such approaches to more complex polyphonic works such as orchestral music, and study the correlation of texture with other parameters such as harmony or form.

References

1. Emilios Cambouropoulos. Musical parallelism and melodic segmentation. *Music Perception*, 23(3):249–268, 2006.
2. Emilios Cambouropoulos. Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception*, 26(1):75–94, 09 2008.
3. Elaine Chew and Xiaodan Wu. Separating voices in polyphonic music: A contig mapping approach. In *International Symposium on Computer Music Modeling and Retrieval (CMMR 2005)*, pages 1–20, 2005.
4. Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2017.
5. Michael Scott Cuthbert and Christopher Ariza. music21: A toolkit for computer-aided musicology and symbolic music data. In *International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 637–642, 2010.
6. Ben Duane. *Texture in Eighteenth- and Early Nineteenth-Century String-Quartet Expositions*. PhD thesis, Northwestern University, 2012.
7. Anders Friberg and Sven Ahlbäck. Recognition of the main melody in a polyphonic symbolic score using perceptual knowledge. *Journal of New Music Research*, 38(2):155–169, 2009.
8. Klaus Frieler et al., Don't hide in the frames: Note-and pattern-based evaluation of automated melody extraction algorithms. In *Digital Libraries for Musicology (DLfM 2019)*, pages 25–32, 2019.

9. Mathieu Giraud, Florence Levé, Florent Mercier, Marc Rigaudière, and Donatien Thorez. Towards modeling texture in symbolic data. In *International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 59–64, 2014.
10. David Huron. Characterizing musical textures. In *International Computer Music Conference (ICMC 1989)*, pages 131–134, 1989.
11. David Huron. Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception*, 19(1):1–64, 2001.
12. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR 2015)*, 2015.
13. Katalin Komlós. Haydn's keyboard trios Hob. XV: 5-17: Interaction between texture and form. *Studia Musicologica Academiae Scientiarum Hungaricae*, 28(1/4):351–400, 1986.
14. Ranjeet Kumar, Anupam Biswas, and Pinki Roy. Melody extraction from music: A comprehensive study. In *Applications of Machine Learning*, pages 141–155. Springer, 2020.
15. Janet M. Levy. Texture as a sign in classic and early romantic music. *Journal of the American Musicological Society*, 35(3):482–531, 1982.
16. Søren Tjagvad Madsen and Gerhard Widmer. A complexity-based approach to melody track identification in MIDI files. In *Artificial Intelligence and Music (IWAIM 2007)*, 2007.
17. Søren Tjagvad Madsen and Gerhard Widmer. Towards a computational model of melody identification in polyphonic music. In *International Joint Conference on Artifical Intelligence (IJCAI 07)*, page 459–464, 2007.
18. Marcel Mongeau and David Sankoff. Comparison of musical sequences. *Computers and the Humanities*, 24(3):161–175, 1990.
19. Daniel Muellensiefen, Marcus Pearce, and Geraint Wiggins. A comparison of statistical and rule-based models of melodic segmentation. In *International Conference on Music Information Retrieval (ISMIR 2008)*, pages 89–94, 2008.
20. Quentin R. Nordgren. A measure of textural patterns and strengths. *Journal of Music Theory*, 4(1):19–31, 1960.
21. Jay Rahn. Where is the melody? *In Theory Only: Journal of the Michigan Music Theory Society*, 6:3–19, 01 1982.
22. David Rizo, Pedro J. Ponce De León, Antonio Pertusa, and Jose M. Iñesta. Melody track identification in music symbolic files. In *International Florida Artificial Intelligence Research Society Conference (FLAIRS 2006)*, 2006.
23. Justin Salamon. *Melody Extraction from Polyphonic Music Signals*. PhD thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2013.
24. Marcos Sampaio. Contour similarity algorithms. *MusMat*, 2(2):58–78, 2018.
25. Mark Schmuckler. *Tonality and Contour in Melodic Processing*, pages 143–165. Oxford University Press, 2016.
26. Federico Simonetta, Carlos Cancino-Chacón, Stavros Ntalampiras, and Gerhard Widmer. A convolutional approach to melody line identification in symbolic scores. In *International Society for Music Information Retrieval Conference (ISMIR 2019)*, pages 924–931, 2019.
27. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.
28. Alexandra Uitdenbogerd and Justin Zobel. Melodic matching techniques for large music databases. In *International Conference on Multimedia (Multimedia 99)*, pages 57–66, 1999.
29. Gissel Velarde, Tillman Weyde, and David Meredith. An approach to melodic segmentation and classification. *Journal of New Music Research*, 42(4):325–345, 2013.
30. Valerio Velardo, Mauro Vallati, and Steven Jan. Symbolic melodic similarity: State of the art and future challenges. *Computer Music Journal*, 40(2):70–83, 2016.
31. Anja Volk and Peter Van Kranenburg. Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, 16(3):317–339, 2012.