Unsupervised method for Implementing Implication-Realization Model Analyzer on Computer

Kaede Noto[1], Yoshinari Takegawa[1], and Keiji Hirata[1*]

Future University Hakodate
g3120002@fun.ac.jp

Abstract. We propose an implementation method for an implication-realization (I-R) model analyzer that is based on note duration, beat structure, and pitch transition. The proposed method involves two procedures, i.e., symbol-start note estimation based on note duration, beat structure, and pitch transition and symbol assignment by introducing a method for recurrently determining small and large intervals in an I-R model analyzer. With the Narmour's manual I-R analysis, our methof had an F measure of 0.86 symbol-start note estimation.

Keywords: Implication Realization Model, Music theory, Music cognition

1 Introduction

We propose an implementation method for an implication-realization (I-R) model [2],[3] analyzer that is based on note duration, beat structure, and pitch transition. I-R analysis classifies the relationship between adjacent notes in accordance with how implications are satisfied or denied. The method for determining these relationships is based on Gestalt theory. In Gestalt theory, the perceptual elements are grouped and recognized. Narmour claims that there is a similar principle in the perception of melody. The smallest unit of a group by I-R analysis is three notes, which are assigned symbols in accordance with their characteristics. For example, the symbol P (process) is assigned to a sequence of notes that are implicated to be heard at the same interval in the same direction. Even when the same implication occurs, the symbol IP (intervallic process) is assigned when only the implication of interval is satisfied, and the symbol VP (registral process) is assigned when only the implication of direction is satisfied. Thus, in I-R analysis, symbols are assigned in terms of whether the melodic expectation is satisfied or denied concerning interval or direction.

I-R analysis involves two procedures. First, to obtain I-R analysis results, it is necessary to estimate the I-R symbol-start note. This is an operation to discover the cognitive boundary in the melody. According to Narmour, the clue to estimating the symbol-start note is the ”closure”. A closure is a note at which no implication arises from the sequence of notes occurring or where the implication is weakened. In other words, a closure is an event that triggers a grouping boundary. Specifically, it refers to a change in pitch interval, direction, or note value or the occurrence of a strong beat, etc. I-R analysis is conducted for triplets starting from those boundaries.

* Please place acknowledgement here.

Second, the I-R symbols are assigned to a sequence of notes that is longer than three notes, starting with the note estimated in the first procedure. Symbols are assigned mainly on the basis of two principles, principle of intervallic difference (PID) and principle of registral direction (PRD). The PRD states that small (five seminotes or less) intervals imply an interval in the same direction, and large (seven seminotes or more) intervals imply an interval in the registral direction. The PID states that small intervals imply a similarly-sized (plus or minus two seminotes) interval, and large intervals imply a small interval. On the basis of these principles, three particular notes are assigned one of eight symbols. By completing these two procedures, we can obtain the results of I-R analysis.

The purpose of this study was to develop an implementation method for an I-R model analyzer and quantitatively evaluate the results of the analysis. Although several methods have been proposed to implement an I-R model analyzer, there have not been studies that have quantitatively evaluated the accuracy of these methods. Grachten et al. proposed an implementation method for an I-R model analyzer using decision trees [4], and Yazawa et al. proposed one using extended I-R symbols [8], but to evaluate these methods, they used the performance of melodic similarity with the analysis results as features.

Because the performance of these methods are not known. Therefore, the usefulness of the I-R analysis results in the Music information retrieval (MIR) field is not clear. In this study, we evaluated the performance of our proposed method by comparing it with the Narmour's manual analysis.

2 Methodology

The flow of our method is as follows (Figure 1). First, we estimate the symbol-start note. This consists of two steps: closure estimation (Section 2.1, 2.2) and determining the order of symbol assignment (Section 2.3, 2.4, 2.5). Second, after the symbol-start note is estimated, we assign it a symbol (Section 2.6, 2.7).

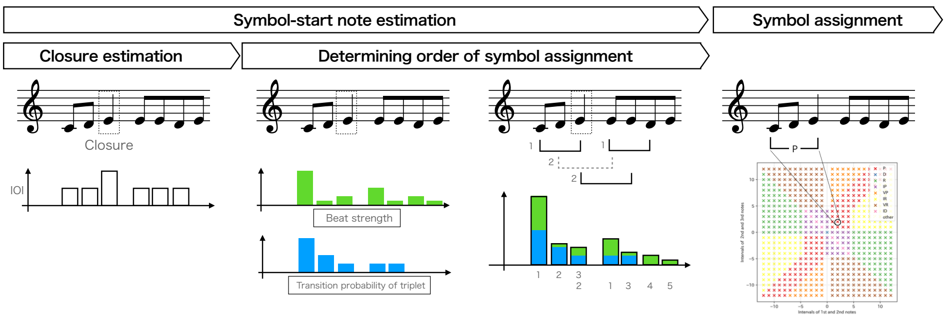

Fig. 1. Flow of proposed method

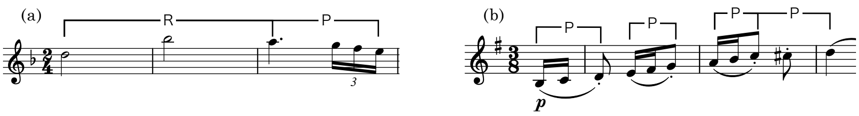

Fig. 2. I-R analysis by Narmour. (a) W. A. Mozart: Piano Concerto No. 19 in F Major, Kv.459, 3rd mov., (b) L. v. Beethoven: Piano Sonata op.14-2, 3rd mov..

2.1 Closure

We first explain a closure to estimate the starting note of the symbol. We give examples of the I-R analysis by Narmour (Figure 2) to explain the relationship between a closure and symbol-start note. In the example in Figure 2 (a), the note considered to be the closure is the first note of the third measure, i.e., ”la,” which has a strong beat and changing note value. In this case, the closure ”la” is the end of symbol R, which starts in the first measure, and the beginning of symbol P, which starts in the third measure.

In the example in Figure 2(b), the first and fourth notes of the first measure, third note of the second measure, and first note of the third measure are considered closures. In this case, the first, fourth, and third notes of the first measure are the end notes of symbol P, and the third note of the second measure is the start and end notes of the symbol. Thus, we define ”closure” as the union set $A \cup B$ when the symbol-start note is represented by set A and the symbol-end note by set B.

2.2 Closure Estimation Based on Inter-onset Interval

Our proposed method uses a closure-estimation method that focuses on the change in the inter-onset interval (IOI). Researchers have attempted to estimate a closure by focusing on note duration. For example, with current closure-estimation methods, the closure is considered to be the point where the note duration increases [6] or rests occur [8]. The problem with these methods is that they estimate many closures for melodies that have alternating notes and rests. However, we integrate the above methods by focusing on the change in the IOI. Because the IOI is the difference between the times at which each note occurs, the IOI of any two notes will not change even if the note value changes or rests are inserted, unless the timing of onset is changed. Because time is handled differently for note value and IOI, it is necessary to develop a method for detecting changes in note duration when targeting the IOI. Current methods are based on the note value (quarter notes, eighth notes) in a score. It is reasonable that an increase in note value is defined as a factor of two or more compared with the previous notes. Therefore, we consider a note at which the IOI increases by a factor of two or more compared with the previous IOI as the closure.

Closure estimation can be used to limit the targets for symbol assignment. As above, because we define a closure to be the union set of the start and end notes of a symbol, we do not assign symbols across closures. However, if the symbols before and after the closure are identical, sometimes they may be regarded as symbols across the closure. The details are given later.

2.3 Transition of Pitch

We introduce transition of pitch on the basis of the hypothesis that the first note of a pitch-transition pattern that frequently occurs in a melody is likely to be the symbol-start note. Because the length of the I-R symbol is three, we calculate the probability of occurrence of a tri-gram in a melody. In addition, the first two notes in the I-R symbol are those that generate an implication. The third note will satisfy or deny the implication. Thus, we use the probability of the $(i+2)$ th note occurring after the $(i+1)$th and i th notes, $P(i+2|i, i+1)$, as a feature for estimating the symbol-start note.

The distribution of $P(i+2|i, i+1)$ changes depending on how the random variable is determined. For example, when determining the transition probability of a melody, the pitch is generally used as a random variable. However, when the pitch is used as a random variable, the probability distribution after learning is likely to be sparse. To avoid this, we consider two random variables, pitch interval and qualitative pitch interval. Because pitch intervals are divided into two values, i.e., S (small) and L (large), in I-R analysis, we define qualitative pitch as a binary expression of n or less seminotes and more than n seminotes.

2.4 Beat Structure

We use the beat structure for symbol assignment. The beat structure is known to affect group formation when listening to a melody. Fraisse reported that when presented with a sequence of sounds that occur in the same time span, people divide these sounds into two or three repetitive groups [7]. We hypothesize that the I-R symbols are also a type of group, and that symbols are assigned on the basis of the beat. For beat strength, we use the value obtained from Music21Object.beatStrength implemented in the Python library music21 [1] as a feature value. In the beatStrength object, beat strength is expressed as a relative value, such as 1.0 for the first beat of a measure, 0.5 for downbeats, and 0.25 for upbeats.

2.5 Feature Integration

We estimate symbol-start notes from closure, pitch-transition pattern, and beat structure. The search range for estimating the symbol-start note is the interval from one closure to the next. We integrate pitch-transition pattern and beat structure within this interval. Integration refers to standardizing each value then calculating the sum. Because the sum of values indicates how likely it is to be a symbol-start note, we assign the symbols in order, starting with the highest value.

3 Symbol Assignment

3.1 Previous Symbol-assignment Method and Actual Data

There are ambiguities with current method of I-R symbol assignment. Basically, we can make rules from the PID and PRD proposed by Narmour on how to assign symbols to the three notes. However, we also need to determine a threshold for determining the S

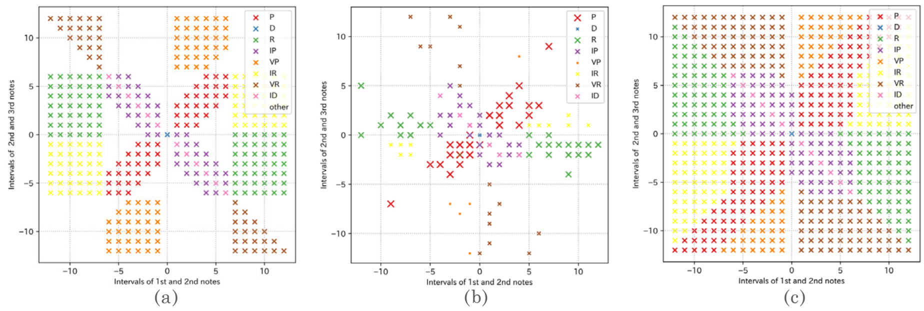

Fig. 3. Distribution of I-R symbols assigned to triplet. Horizontal axis is number of seminotes in first and second notes, and vertical axis is number of seminotes in second and third notes.

or L pitch interval as a hyperparameter. Figure 3(a) shows the distribution of symbols when S is six or less seminotes and L is more than six seminotes, indicating that different symbols are being assigned after the threshold. Nothing is written to the coordinates corresponding to triplets that do not assign the I-R symbols. Figure 3(b) shows the distribution of the I-R symbols observed from Narmour's manual analysis. These figures show the correspondence between the symbols observed from Narmour's analysis and the pitches of the triplets, with the size of each point proportional to the number of times it was observed. We did not observe any example of "other" symbol assigned to a triplet. The intricacy of each symbol's region concerning the axial direction suggests that the threshold for determining the I-R symbol to be assigned is not fixed.

We introduce a symbol-assignment method that recursively changes the threshold. If the threshold is defined as n, then S can be regarded as a pitch within n semitones, and L as a pitch greater than n semitones. The initial threshold is n = 6, following Narmour's rule-based method. After the symbol assignment with n = 6 is completed, we gradually increase the threshold from n = 1. This operation yields a distribution of symbols, as shown in Figure 3(c). We can see that it includes the symbols in Figure 3(a) and many of those in Figure 3(b).

3.2 Detailed Rules for I-R Symbol Assignment

To conduct I-R analysis for an actual melody, we need to determine the number of symbols to be assigned to each note. For example, if we allow three symbols to be assigned to any note, the operation is the same as the I-R analysis for a tri-gram. We conducted I-R analysis with two and three maximum symbol assignments and compared the results.

The object of I-R analysis is a triplet, but four or more notes may be assigned symbols P or D (Duplicate). These symbols have the characteristic of repeating similar pitches in the same direction. Thus, the repetition of symbol P or D is thought to amplify the implication. Therefore, if symbols P and D are superimposed, they are merged and considered one symbol.

However, this is not the case if the implied attenuation occurs within symbols P and D, which consist of four or more notes. As already indicated, even when symbol P is consecutively, they may not be integrated (Figure 2). This is thought to be due to the fact that closures occur between symbols. However, it is difficult to investigate all the possibilities of generating closures. Thus, we introduce a symbol-integration rule that focuses only on the beat structure, which can be understood intuitively.

4 Experiment

We conducted an evaluation experiment to investigate the accuracy of the proposed implementation method in estimating the symbol-start tone and the factors that contributed to the results. There were five evaluation items.

Table 1. Evaluation items

Evaluation items	
1. Features	1-1. Closure estimated from IOI 1-2. Transition probability of triplet 1-3. Beat strength
2. Random variable with transition probability of triplet	2-1. Pitch 2-2. Interval 2-3. Qualitative pitch interval
3. Maximum number of symbols assigned	3-1. Two 3-2. Three
4. Division of symbols (Beat Strength)	4-1. No division 4-2. First beat of measure (1.0) 4-3. Downbeat (0.5) 4-4. Downbeat and upbeat (0.25)
5. Threshold of symbol assignment	5-1. Narmour's method (N = 6) 5-2. Proposed Method

4.1 Evaluation values and dataset

We used the results of the manual analysis by Narmour as the correct data. Because the rules of I-R analysis are often ambiguous and the results are subjective, there is no large data set of I-R analysis results. Thus, we used 61 examples taken from Narmour's analysis examples [3] as the correct data. The melodies used as the correct data were selected on the basis of the following two criteria. The first criterion is the number of notes contained in the melody to be analyzed. If the number of notes is four or less, the results of the I-R analysis can be uniquely determined. Thus, we did not take into account melodies not considered as correct data. The second criterion is whether the three tones to be analyzed are adjacent to each other. In Narmour's analysis, there is an example in which similar sound sequences are considered as one cohesive unit, and the beginning of the unit is extracted and subjected to I-R analysis. In these cases, we did not consider them as correct data because it is necessary to select three tones for I-R analysis, which is beyond the scope of this study. The correct data were all created manually as MusicXML with the following information: pitch, duration, onset time, I-R symbol, and symbol-start note.

The input to the system is pitch, duration, beat strength, and the output is a binary value indicating whether the note is a symbol-start note. Thus, we used recall, precision, and F-measure to evaluate the method.

4.2 Training Data

To calculate Feature 1-2. (Transition probability of triplet), we used 300 melodies from GTTM DataBase [5] as training data. As mentioned above, we did not label the training data because what we want is the conditional probability of three adjacent notes in the melody. Because the conditional probability will be zero if the pitch-transition pattern included in the target melody does not exist in the training data, we included the target melody in the training data.

4.3 Evaluation Results

Figure 4 presents the results for symbol-start note estimation when different features are used for estimation. Cases 1 to 8 on the horizontal axis of each bar graph correspond to the combinations of features in Table 2. The bars located on the left side are the evaluation scores when more features were used. Case 1 is the result of estimation with three features, and Case 8 is that without any features. The highest score was obtained when all the features were used.

We found that the score tended to increase with the number of features used. However, there was no difference in the F-measures between Case 4, which used two features, and Case 5, which used one feature. Because the difference between Cases 4 and 5 is the presence or absence of Feature 1-2, Feature 1-2 is considered to have an effect on the score. However, there was a difference of 0.1 in the F-measure of Cases 2 and 6, which also differed only in the presence and absence of Feature 1-2. This result indicates that it is not only the features used in the estimation but also the combination of features.

Table 2. Feature selection

Feature	Evaluation items							
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8
1-1. Closure estimated form IOI	✓		✓	✓	✓			
1-2. Transition probability	✓	✓		✓			✓	
1-3. Beat strength	✓	✓	✓			✓		

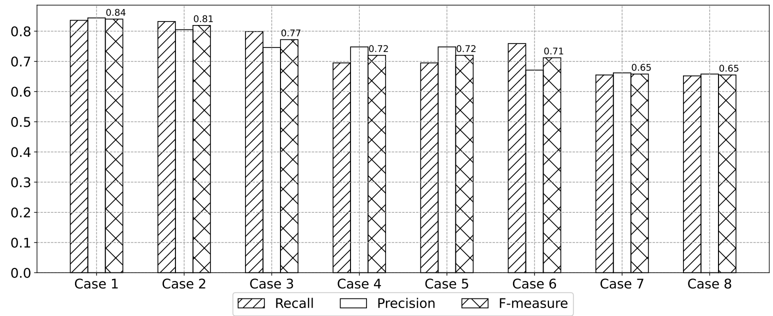

Fig. 4. Experimental results for Cases 1 − 8 regarding 2-3. Qualitative pitch interval, 3-2. Three, 4-2. First beat of measure, and 5-1. Narmour's method (N = 6)

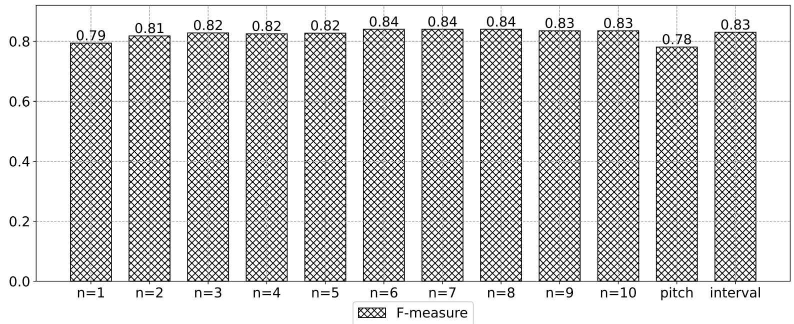

Fig. 5. Experimental results for evaluation item 2 regarding Case 1, 3-2. Three, 4-2. First beat of measure and 5-1. Narmour's method (N = 6))

Figure 5 presents the results for symbol-start note estimation when conditional probabilities are calculated using different random variables. The value of n in the graph represents the number of semitones used to determine the qualitative pitch interval. For example, if n = 3, all intervals appearing in the melody are represented as two values, one for intervals of three semitones or less, and one for intervals of four semitones or more. The highest evaluation values were obtained when n = 6, 7, and 8.

Figure 6 presents the results for symbol-start note estimation when comparing the maximum number of symbols assigned. When the maximum number of symbols assigned is three (3-1.), the analysis results are equivalent to the I-R analysis results for a tri-gram with the symbols that straddle the closure removed. We can see in Figure 6 that when the maximum number of symbols is three (3-1.), recall is higher than when the maximum number of symbols is two (3-2.). This is because when the maximum number of symbols is three, our method estimates more symbol-start notes. However, precision decreased. Therefore, Figure 6 indicates that if we want to achieve a high F-measure, it is better to use the maximum number of symbols of two.

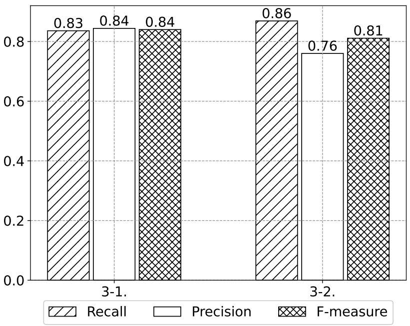

Fig. 6. Experimental results for evaluation item 3 regarding Case 1, 2-3 Qualitative pitch interval, 4-2. First beat of measure and 5-1. Narmour's method (N = 6)

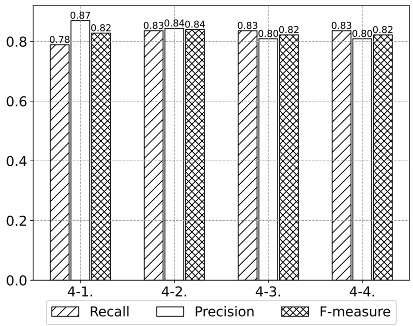

Fig. 7. Experimental results for evaluation item 4 regarding Case 1, 2-3. Qualitative pitch interval, 3-2. Three, and 5-1. Narmour's method (N = 6)

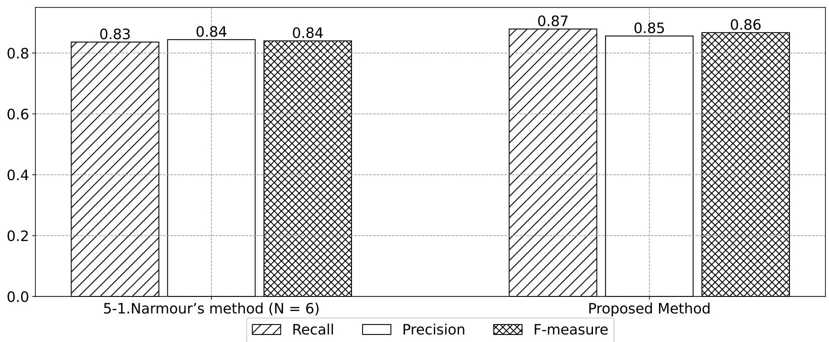

Fig. 8. Experimental results for evaluation item 5. regarding Case 1, 2-3. Qualitative pitch interval, 3-2. Three, and 4-2. First beat of measure

Figure 7 presents the results for symbol-start note estimation when symbols P and D, which consist of four or more notes, are divided in accordance with the beat strength. Thus, the number of notes considered to be symbol-start notes increased. The condition with the highest precision was when no splitting was carried out. However, recall was lowest among the four conditions, which indicates that the coverage in finding the symbol-start note is low. The highest F-measure was obtained when beat strength was 4-1. (beat strength = 1.0), which is when the symbols are split at the beginning beat of the measure. Also, when splitting symbols on smaller beats (downbeat or downbeat and upbeat), precision decreased. Hence, if we want to increase the accuracy of symbol-start note estimation, only splitting symbols at the beginning of the measure is sufficient.

However, the small difference in the evaluation values for 4-1., 4-2., and 4-3. (beat strength = 1.0, 0.5, and 0.25) may be due to a bias in the appearance of symbols P and D. In this experiment, the best score was obtained by estimating the symbol-start note with a beat strength of 1.0, but we do not know whether similar results can be achieved when we conduct I-R analysis on melodies with fast passages.

Figure 8 presents a comparison of the results of symbol-start note estimation with different symbol-assignment methods. The method of assigning these symbols conforms to the symbol distribution shown in Figure 3. Narmour's method (N = 6) corresponds to Figure 3(a), and the proposed method corresponds to Figure 3(c). The proposed method had a better score than Narmour's method.

The difference between the two methods is in the handling of symbols that were considered as ”other” with Narmour's method. With this method, no symbol is assigned to the triplet corresponding to the ”other”, but with the proposed method, a symbol is assigned to the triplet. Thus, more notes will be inferred as symbol-start notes with the proposed method.

5 DISCUSSION AND CONCLUSION

We proposed an implementation method for an I-R analyzer with symbol-start note estimation that is based on note duration, beat structure, and uses pitch transition and a symbol-assignment method for changing the threshold recursively. With the examples of Narmour's I-R analysis, our method had an F measure of 0.86. We also conducted a comparative verification for each feature.

We evaluated the accuracy of our I-R analyzer, but its usefulness in MIR is not clear. Features based on human cognition have been used to get boundary, such as lyrics and syllables [9]. By making comparisons with such studies, we hope to be able to compare the usefulness of I-R model from a cognitive perspective. Future work includes feature design for treating I-R symbols as features and comparison with previous studies.

Acknowledgments This work has been supported by JSPS Kakenhi 16H01744.

References

1. Cuthbert, M. S., Ariza, C.: music21: A toolkit for computer-aided musicology and symbolic music data (2010).
2. Narmour, E.: The Analysis and Cognition of Basic Melodic Structures. The University of Chicago Press (1990)
3. Narmour, E.: The Analysis and Cognition of Melodic Complexity. The University of Chicago Press (1992)
4. Grachten, M., Arcos, J. L., and López de Mántaras: Melody retrieval using the implication/realization model. MIREX (2005).
5. Hamanaka, M., Hirata, K., and Tojo, S.: ”Musical structural analysis database based on GTTM.” (2014)
6. Hatano Y.: Music and Cognition, University of Tokyo Press, pp. 1 – 40 (1989)
7. Paul Fraisse: Rhythm and tempo. The psychology of music, 1, pp.149-180. (1982)
8. Yazawa, S., Hamanaka, M., Utsuro, T.: ”Subjective Melodic Similarity based on Extended Implication-Realization Model.” IJAE 15.3, pp. 249-257 (2016)
9. Bas, C., Zuidema, W., and Burgoyne, A.: ”Mode Classification and Natural Units in Plainchant.” Proceedings of the 21th International Conference on Music Information Retrieval (ISMIR 2020). Montréal, Canada. (2020)