Merged Pitch Histograms and Pitch-duration Histograms

Hui Liu[a], Tingting Xue[b] and Tanja Schultz[c]
Cognitive Systems Lab, University of Bremen, Germany

Keywords: Pitch Statistics, Pitch Histogram, Merged Pitch Histogram, Pitch-duration Histogram, Pitch-related Features, Music Computing, Music Information Retrieval.

Abstract: The traditional pitch histogram and various features extracted from it play a pivotal role in music information retrieval. In the research on songs, especially applying pitch statistics to investigate the main melody, we found that the pitch histogram may not necessarily reflect the notes' pitch characteristic of the whole song perfectly. Therefore, we took the note duration into account to propose two advanced versions of pitch histograms and validated their applicability. This paper introduces these two novel histograms: the merged pitch histogram by merging consecutively repeated pitches and the pitch-duration histogram by utilizing each pitch's duration information. Complemented by the description of their calculation algorithms, the discussion of their advantages and limitations, the analysis of their application to songs from various languages and cultures, and the demonstration of their use cases in state-of-the-art research works, the proposed histograms' characteristics and usefulness are intuitively revealed.

1 INTRODUCTION

In (Tzanetakis et al., 2003), the authors first propose the concept of *pitch histogram* that reflects the pitch content of notes in music pieces. Due to the normalization, pitch histogram is also called *pitch-frequency histogram* (Gedik and Bozkurt, 2010), but the essence is the same. A variation of the basic pitch histogram based on music theory is the *pitch class histogram*, which counts only twelve pitch classes, and the various octaves of each pitch are grouped into the same bin. Derived from the pitch class histogram, a *folded fifths pitch class histogram* is calculated by reordering the bins of the original unordered histogram such that perfect fifths rather than semitones separate adjacent bins. The pitch histogram also has successors for different occasions, such as a *melodic interval histogram* that counts the distances between two consecutive pitches rather than the individual pitches themselves. A variety of features can be extracted from the pitch histogram and its derivatives, such as range, mean, variability, skewness, and kurtosis (McKay, 2010), applied to various machine learning-based music research aspects like genre classification and cultural diversity analysis.

[a] https://orcid.org/0000-0002-6850-9570
[b] https://orcid.org/0000-0002-5815-7217
[c] https://orcid.org/0000-0002-9809-7028

As another musical characteristic alongside the pitch, the note duration and the rhythm resulting from it are also important research material. A *rhythmic value histogram* is a normalized histogram, where the value of each bin specifies the fraction of all notes in the piece with a quantized rhythmic value corresponding to that of the given bin (McKay, 2010). Not similarly, a *beat histogram*, first applied to *Music Information Retrieval* (MIR) research in (Brown, 1993), emphasizes note onsets rather than durations, and is also utilized for genre classification (Tzanetakis et al., 2001) (Tzanetakis and Cook, 2002) (Tzanetakis, 2002) (Lykartsis and Lerch, 2015). From a simplistic perspective, rhythm can be perceived as the number of notes played at a specific tempo within a bin, initiating a *(normalized) duration histogram* in (Karydis, 2006).

Although the pitch sequence and the duration sequence of music are widely used to generate various kinds of histograms and features, each note's pitch and duration information are rarely used together. For example, (Karydis, 2006) used both pitch histograms and (normalized) note duration histograms to generate their series of features, respectively, for the accuracy improvement of symbolic music genre classification, but each note's pitch and duration information was not co-analyzed. As stated in (Adams et al., 2004), while a transcription into a sequence

Liu, H., Xue, T. and Schultz, T.
Merged Pitch Histograms and Pitch-duration Histograms.
DOI: 10.5220/0011310300003289
In *Proceedings of the 19th International Conference on Signal Processing and Multimedia Applications (SIGMAP 2022)*, pages 32-39
ISBN: 978-989-758-591-3; ISSN: 2184-9471

of $(pitch, duration)$ pairs is convenient and musically intuitive, no evidence shows that it is an optimal representation (at that time).

In the study of songs' main melody scores, we found that the pitch histogram and the features it produces do n¡t necessarily depict the pitch distribution of the whole song perfectly. For instance, the center pitch of the whole song's main melody, also called the *melody center* in the automatic singing key estimation research (Liu et al., 2022b), is often not perfectly reflected by the *mean pitch* or the *median pitch* features extracted from the pitch histogram. More reasonable descriptions can be achieved by applying each note's pitch and duration information together. Another typical example is that jointly considering consecutive identical pitches can, to some extent, provide a better representation of the particular pitch distribution of traditional Han anhemitonic pentatonic folk songs (Liu et al., 2022a). Based on a series of experimental verification, we put forward and instantiate two novel types of pitch histograms, the *merged pitch histogram* and the *pitch-duration histogram*, combined with note duration in some or full measure, hoping that they can benefit various aspects of music research.

The vertical axes of the three pitch histograms (basic, merged, pitch-duration) in all figures of this paper are normalized, respectively, for a better comparison. As introduced above, because of the normalization, pitch histograms can be called pitch-frequency histograms (Gedik and Bozkurt, 2010). Therefore, with normalization, the two proposed histograms can be similarly named *merged pitch-frequency histograms* and *pitch-duration-frequency histograms*.

The instances illustrated in this paper apply the notation-based *symbolic* pitch statistics instead of the calculation of audio data. For the latter, (Tolonen and Karjalainen, 2000) proposed a multiple pitch detection algorithm, on which the two novel histograms proposed in this paper are also applicable without obstacles.

2 TWO NOVEL VARIANTS OF PITCH HISTOGRAMS

2.1 Merged Pitch Histogram

Consecutive notes of the same pitch in instrumental performance should be counted reasonably repeatedly, as the player needs to use the organ repeatedly and proficiently to play these notes. For the performance or research of songs, whether the pitch's continuous repetition should be counted repeatedly deserves further investigation.

Figure 1: The first phrase of *The Sound of Silence*'s main melody.

For songs in polysyllabic languages, a word is sometimes split into consecutive identical pitches. The words "hello" and "darkness" in the first phase of *The Sound of Silence* are examples (see Figure 1). However, this is not the case for songs in monosyllabic languages. Therefore, it may cause problems in calculating the pitch histogram-based features for a music research work containing songs of both language types.

In addition, the basic pitch histogram is not well compatible with differences in melodic details, including symbolic-symbolic, symbolic-audio, and audio-audio divergences, which will be detailed in Section 4.1.

As a modification of the basic pitch histogram that considers pitches' temporal information in a simple way, a merged pitch histogram counts consecutive notes of the same pitch only once. If there is a rest among a series of consecutive identical pitches, the first pitch after the rest should be counted again.

The red histogram in Figure 2 illustrates an example of a merged pitch histogram representing the first phrase of the song *That's Why (You Go Away)*.

2.2 Pitch-Duration Histogram

The merged pitch histogram considers pitches' temporal information in some measure (from the perspective of continuity) to make decrements on the corresponding bins; In contrast, the pitch-duration histogram makes information increments on the bins utilizing each pitch's temporal information in full measures (from the perspective of weighting). Each bin in the pitch-duration histogram responds to the total duration of the corresponding pitch in the music piece, instead of its number of occurrences.

The green histogram in Figure 2 displays a pitch-duration histogram containing only the first phrase in the song *That's Why (You Go Away)*.

2.3 Comparison

Figure 2 clearly demonstrates the differences between the two proposed novel histograms and the basic pitch histograms:

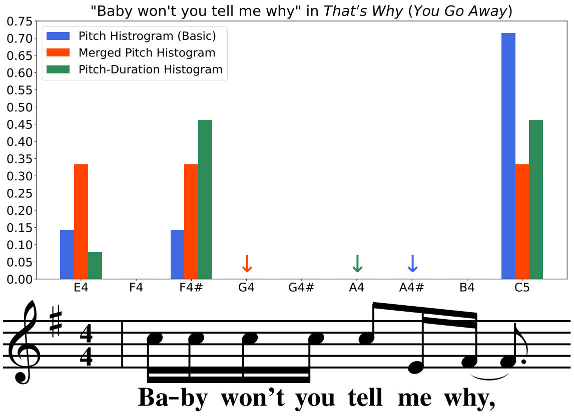

Figure 2: The basic pitch histogram and its two novel variants for the first phrase of *That's Why (You Go Away)*'s main melody. Colored down arrows mark the *mean pitch* features ("melody centers") for each type of histograms.

- The basic pitch histogram in blue emphasizes the note C5 due to considering only the number of occurrences. The corresponding *mean pitch* (A4#) is relatively high, close to C5.
- By merging consecutive identical pitches, the merged pitch histogram in red presents the phrase as an even distribution, while its *mean pitch* (G4) is skewed towards the lower two pitches.
- The pitch-duration histogram in green elevates the importance of F4# to the same level as C5 regarding the entire note duration. The calculated *mean pitch* (A4) is between the above two.

As a side note, the *median pitch* calculated by the average of the entire phrase's highest and lowest pitches is G4#, different from all the mean pitches.

2.4 Histogram Calculation

For the calculation of the basic symbolic pitch histogram using notation-based formats, as (Tzanetakis et al., 2003) explains, the algorithm increments the corresponding pitch's counter. The value in each histogram bin is normalized in the last stage of the calculation by dividing it by the total number of pitches of the whole piece, in order to account the variability in the average number of pitches per unit time between different pieces of music. Because of the normalization step, in some literature, like (Gedik and Bozkurt, 2010), a pitch histogram is also called a pitch-frequency histogram.

The merged pitch histogram is generated based on the above basic pitch histogram calculation with a slight modification: if a series of consecutive notes have an identical pitch, and there are no rests between them, the corresponding pitch's counter is added only by one. Such a calculation is straightforward to implement with a standard loop: the current note's pitch is compared with the previous one (a rest is also counted as a note, but only for comparison). If the same, the pitch corresponding to the note is not counted; otherwise, add one to the pitch's counter. Note that the denominator in the final normalization process is not the same as in the basic pitch histogram.

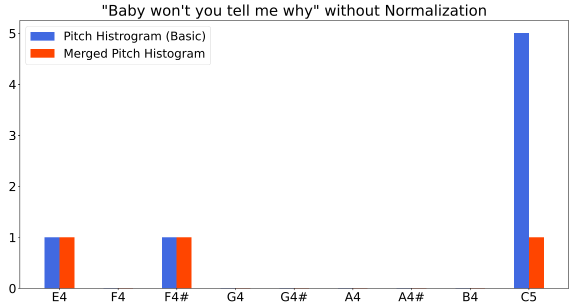

Figure 3: The basic pitch histogram and the merged pitch histogram for the first phrase of *That's Why (You Go Away)*'s main melody without normalization.

It is worth pointing out that without normalization, we can directly observe which pitches are merged and how much is the quantity of notes being merged by comparing the merged pitch histogram to the basic pitch histogram, as Figure 3 evidences. After normalization, the conclusion drawn by height comparison changes slightly. First, we exclude the easy case where no merging happens in the whole melody. The pitch bins that do not have merged notes must be higher in the normalized merged pitch histogram than in the normalized basic pitch histogram (see the pitches E4 and F4# in Figure 2) due to the fact that the numerators of the two calculations are identical and the former's denominator is less than the latter's. The contrapositive conclusion is that a pitch must contain merged notes if its bin's normalized height in the merged pitch histogram is shorter than or equal to its corresponding normalized height in the basic pitch histogram (see the pitch C5 in Figure 2).

The construction process of the pitch-duration histogram can go simply through a traversal of the notes in the piece. For each note, its duration is accumulated over the bin of the corresponding pitch. After processing all notes, the histogram is normalized by dividing by the sum of the durations of all pitches.

3 INSTANCES AND ANALYSIS

For intuitive analysis, we apply the basic pitch histogram, the merged pitch histogram, and the pitch-duration histogram to a set of folk songs from differ-

Table 1: Information on all songs analyzed in this paper.

Song	Lyricist	Composer	Language/Country	Pitch range
The Sound of Silence	Paul Simon		English/USA	18 semitones
That's Why (You Go Away)	Jascha Richter		English/Denmark	21 semitones
Auld Lang Syne	Robert Burns		Scottish/UK	18 semitones
Vo Luzern/Luzärn auf/uf Wäggis zue	Johann Lüthi		German/Switzerland	18 semitones
Bella Ciao			Italian	14 semitones
Arirang			Korean	13 semitones
Anile Anile Vaa Vaa Vaa			Tamil/India	13 semitones
Troika Pochtovaya			Russian	16 semitones
Sakura Sakura			Japanese	14 semitones
The Green Poplar and Willow			Mandarin	15 semitones

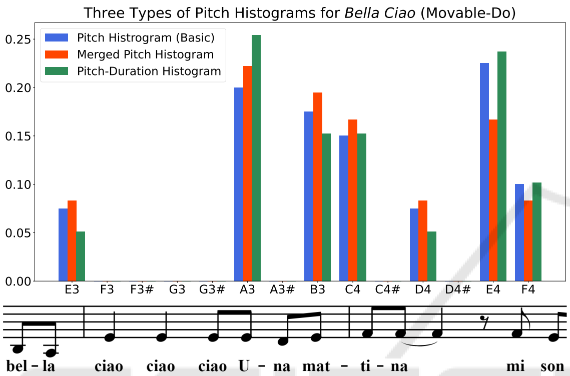

Figure 4: Three types of pitch histograms for *Bella ciao*'s main melody with Movable-Do. The score segment below indicates the phrases where E4 and F4 are repeated.

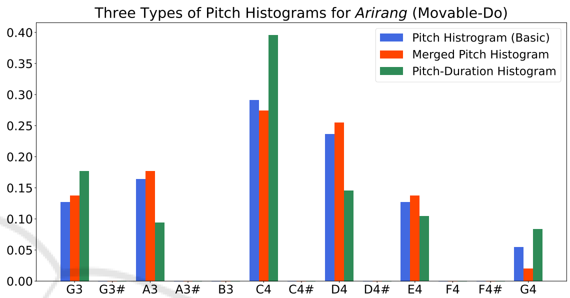

Figure 5: Three types of pitch histograms for *Arirang*'s main melody with Movable-Do.

ent languages and cultures. Since most of them do not have original definite keys, they are all notated using Movable-Do. Table 1 lists the information on the songs, including the three exemplified in Sections 2 and 4.

Bella Ciao

The merged pitch histogram of the Italian folk song *Bella Ciao* illustrated in Figure 4 has shorter bins for E4 and F4 (in Movable-Do) compared to the basic pitch histogram, resulting in correspondingly longer bins for all other pitches. Such a phenomenon mirrors the fact that only E4 and F4 are repeated successively in *Bella Ciao*'s main melody, the former four times and the latter twice (see the score segment below in Figure 4). In contrast, the importance of E4 and F4, together with A3, is lifted in the pitch-duration histogram, highlighting the fact that these pitches are sung for extended periods throughout the whole song.

It is well understood that the pitch-duration histogram generates the same *mean pitch* (C4) as the basic pitch histogram, while the merged pitch histogram gives one-semitone lower *mean pitch* (B3). The median pitch, generated by the average of the highest and lowest pitches of the song, lies between A3# and B3 and is evidently not comprehensive in reflecting the melody center.

Arirang

As shown in Figure 5, the mergeability of the pitches C4 and G4 (Movable-Do) is witnessed in the Korean anhemitonic pentatonic folk song *Arirang*. Meanwhile, the pitch-duration histogram exchanges the importance of G3 and A3 while emphasizing C4's significance, strongly reflecting this song's major pentatonic scale mode.

Auld Lang Syne

Among the main melody histograms for the Scottish ballad *Auld Lang Syne* exhibited in Figure 6, the merged pitch histogram reverses the importance of G3 and A3 compared to the basic pitch histogram, while the pitch-duration histogram acts significantly on A4. This anhemitonic pentatonic song's basic and merged histograms have been employed recently in a new research topic of computational ethnomusicology, of which more introduction will be provided in Section 4.2.

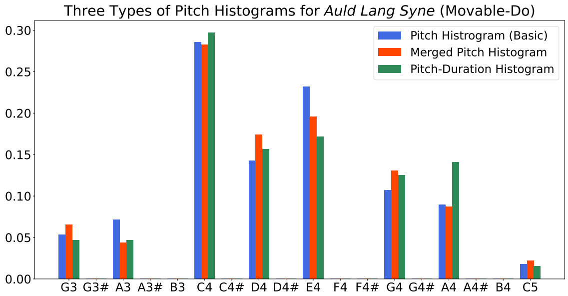

Figure 6: Three types of pitch histograms for *Auld Lang Syne*'s main melody with Movable-Do.

More Instances

Further examples of folk songs in various languages and cultures are given in Figure 7. Most folk songs lack the original scores and are more likely to have divergent details of symbolic melodies than modern songs, which will be detailed in Section 4.1. Applying the merged pitch histogram and the pitch-duration histogram will help dilute these divergences' effects.

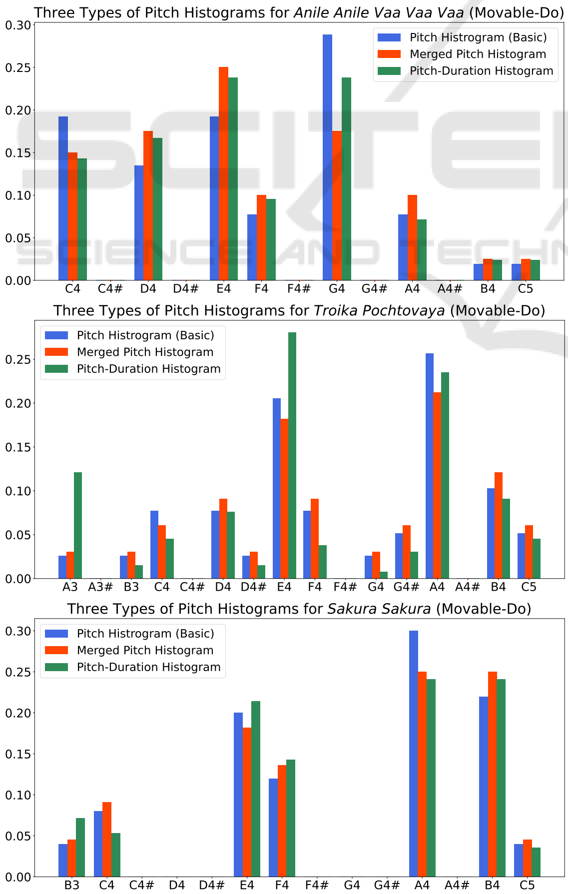

Figure 7: Three types of pitch histograms for *Anile Anile Vaa Vaa Vaa* (top), *Troika Pochtovaya* (middle), and *Sakura Sakura* (bottom) with Movable-Do.

4 DISCUSSION AND USE CASES

4.1 Advantages and Limitations

The potential advantages of merged pitch histograms and pitch-duration histograms are (1) bridging the influence of different linguistic characteristics, which is already explained in Section 2.1; (2) helping downplay differences in melodic details of a song's various versions, which, in MIR studies, can be divided into three types of divergence:

- *Symbolic-symbolic divergences* usually occur when the song or music does not have an original score. Many traditional folk songs have been created by local people's improvisation and handed down orally among the people, then notated by several individuals. Hence, the scores of each version may differ in many details.
- Singers sing, and musicians play the same piece with their characteristics, leading to *symbolic-audio divergences*.
- For the same reasons as the previous one, the melodies that different singers and musicians eventually produce differ from each other, causing *audio-audio divergences*.

A typical instance is the yodels, where it can be noticed that many songs' scores differ in melodic details (e.g., whether some parts use syncopation or not) in different collections, causing a symbolic-symbolic divergence. When sung, the yodeling part often uses a series of successive identical pitches, which usually do not appear on the score, creating a symbolic-audio divergence. The audio-audio divergence is easy to imagine and does not need to be elaborated on. Both merged pitch histograms and pitch-duration pitch histograms eliminate the effect that makes the yodeling pitches' bins in unnecessarily high counts produced by repeatedly being sung, thus providing a reasonable de-peaking of the song's overall pitch distribution.

Figure 8 conveys an example of the symbolic-symbolic divergence (marked in red). Its symbolic-audio divergence and audio-audio divergence can be reflected in each singer's yodeling parts. All the divergences can be addressed by the merged pitch histogram and the pitch-duration histogram.

The pitch-duration histogram is an improvement that reflects as comprehensively as possible the importance of each pitch in the piece. However, from a melodic divergence perspective, the basic pitch histogram and the merged pitch histogram are sometimes more advantageous for pitch statistics than the pitch-duration histogram, especially in cases where the pitch duration can be arbitrarily prolonged. For

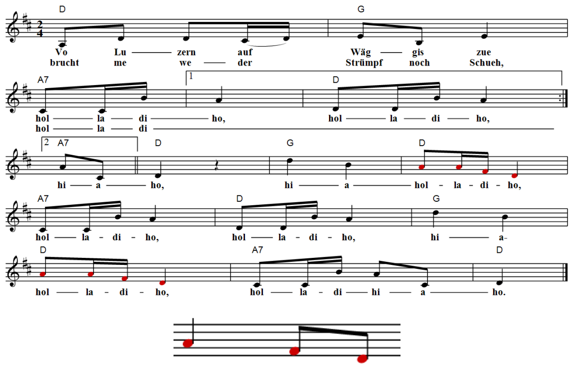

Figure 8: Top: one version of the scores for the Swiss yodel *Vo Luzern/Luzärn auf/uf Wäggis zue* generated using the XML file downloaded from the *Alojado Lieder Archiv* website[1]. The score fragment below shows the divergence marked in red that often occurs in other sources. The merged pitch histogram can eliminate the divergence.

example, for the same position of one phrase's end, some scores extend the beat to the end of the bar, while some give a rest. The basic and the merged pitch histograms count such a pitch once, making more sense than fully considering its duration in the pitch-duration histogram.

4.2 Progressing Cultural Diversity in Computational Ethnomusicology/ MIR with Merged Pitch Histograms

Recent studies have found that the vast majority of traditional Han anhemitonic pentatonic folk songs can be identified intuitively according to their distinctive bell-shaped pitch distribution in pitch histograms, reflecting the Chinese characteristics of Zhongyong (the doctrine of the mean) and following the trend from an ethnocultural perspective (Liu et al., 2022a). The bell-shaped basic/merged pitch histograms are also witnessed in some anhemitonic pentatonic folk songs of other East Asian ethnic groups, such as Japanese and Korean (exemplified in Figure 5), indicating the exchange between countries, peoples, and cultures throughout history, although they also have unique musical systems of their own.

As one of the study's results, the bins produced in the basic pitch histogram that violate the bell shape can be improved by the merged pitch histogram to a great extent.

Take Figure 9 as an example. This song's basic pitch histogram generates significant exceptions at the lowest pitch that violate the bell-shaped curve. It is interesting to note that the "Do Do Do" is frequently repeated at several phrase ends of this song to express an upbeat rhythm, causing this pitch to have a higher count number (Liu et al., 2022a). By utilizing the merged pitch histogram, the pitch statistics show a perfect bell shape.

[1]https://www.lieder-archiv.de/

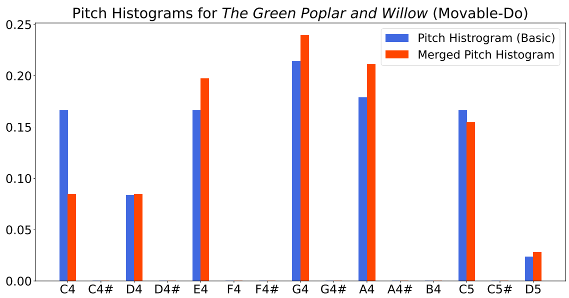

Figure 9: Three types of pitch histograms for *The Green Poplar and Willow*'s main melody with Movable-Do.

For comparison, the Scottish folk song *Auld Lang Syne* is also anhemitonic pentatonic (see Figure 6), but its pitch histogram is far from perfect bell-shaped. Its merged pitch histogram is even less bell-shaped.

The merged pitch histogram were applied together with the *anhemitonic pentatonic pitch histogram* proposed in (Liu et al., 2022a) to create a *merged anhemitonic pentatonic pitch histogram*, playing a helpful role in the feature design and genre classification research. In addition, an existing feature from the *Time Series Feature Extraction Library* (TSFEL) (Barandas et al., 2020), *Negative Turning*, and a designed novel feature, *Degree of Bell Shape* (DoBell), are extracted from various types of histograms to describe to which extent a pitch distribution approximates a bell shape. The preliminary classifiers built with these features performed well, indicating that lightweight machine learning applying only pitch histograms and merged pitch histograms can promote cultural diversity in MIR.

4.3 Automatic Singing Key Estimation Applying New Histogram Variants

The *mean pitch* features derived from the three types of histograms involved in this paper play an essential role in the recent study of automatic singing key estimation (Liu et al., 2022b).

Compared to the simple *median pitch* feature that is calculated as the average of the lowest and highest pitches, the *mean pitch* features generated from the three pitch histograms are more amenable for suggesting the appropriate keys for individuals, among which the *mean pitch* of the pitch-duration histogram is the most reliable.

Figure 10 gives a running interface screenshot of the software *Automatic Singing Key Estimator* (ASKE, pronounced as "Ask-key") presented in (Liu

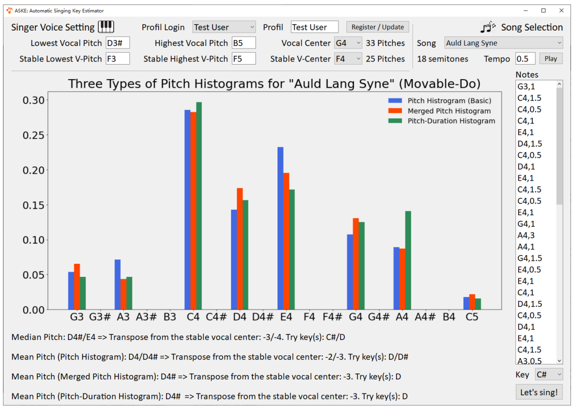

Figure 10: Screenshot of the software *Automatic Singing Key Estimator* (ASKE).

et al., 2022b). For *Auld Lang Syne*, only one key was automatically calculated for each amateur singer using a merged pitch histogram or a pitch-duration histogram, and the resulted key was scored satisfactorily by the subjects in the singing experiment.

5 CONCLUSION AND OUTLOOK

This paper introduces two novel variants of the basic pitch histogram, the merged pitch histogram and the pitch-duration histogram, which consider each pitch's temporal information in the entire melody, the former in some measure (from the perspective of continuity) and the latter in full measure (from the perspective of weighting). Complemented by the analysis of song examples and the comparison between histograms, the characteristics of the proposed histograms are visually expounded. Furthermore, the proposed histograms' computational algorithms, advantages, limitations, and use cases in the latest research works are exhibited in detail.

We believe that the merged pitch histogram and the pitch-duration histogram are meaningful and helpful for music research, especially for the information retrieval of songs' main melody, as exemplified in Sections 4.2 and 4.3. We expect the two introduced histograms to be extensively applied in various research aspects. Future work includes investigating abundant new features extracted from the two novel pitch histograms and applying these features to various machine learning-based music research fields. Besides, diverse transformation approaches of the basic pitch histogram, such as pitch class, folded fifths, and melodic interval, can be practiced on the two novel variants to generate more varied histograms and new features, like the *kurtosis* of the *pitch-class-duration histogram*.

Moving from overall pitch statistics to segment-based features, windowing/framing the song's $(pair, duration)$ sequence to generate time series of merged pitch histograms and pitch-duration histograms deserves further investigating, e.g., utilizing the *Time Series Subsequence Search Library* (TSSEARCH) (Folgado et al., 2022) to conduct a subsequence similarity analysis on the melody.

REFERENCES

Adams, N. H., Bartsch, M. A., Shifrin, J., and Wakefield, G. H. (2004). Time series alignment for music information retrieval. In *ISMIR 2004 – 5th International Conference on Music Information Retrieval*.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). Tsfel: Time series feature extraction library. *SoftwareX*, 11:100456.

Brown, J. C. (1993). Determination of the meter of musical scores by autocorrelation. *The Journal of the Acoustical Society of America*, 94(4):1953–1957.

Folgado, D., Barandas, M., Antunes, M., Nunes, M. L., Liu, H., Hartmann, Y., Schultz, T., and Gamboa, H. (2022). Tssearch: Time series subsequence search library. *SoftwareX*, 18:101049.

Gedik, A. C. and Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for turkish music. *Signal Processing*, 90(4):1049–1063.

Karydis, I. (2006). Symbolic music genre classification based on note pitch and duration. In *ADBIS 2006 – 10th East European Conference on Advances in Databases and Information Systems*, pages 329–338. Springer.

Liu, H., Jiang, K., Gamboa, H., Xue, T., and Schultz, T. (2022a). Bell shapes: Exploring the pitch distribution of traditional han anhemitonic pentatonic folk songs. (submitted).

Liu, H., Xue, T., Gamboa, H., Hertenstein, V., Xu, P., and Schultz, T. (2022b). Automatic singing key estimation for amateur singers using novel pitch-related features: A singer-song-feature model. (submitted).

Lykartsis, A. and Lerch, A. (2015). Beat histogram features for rhythm-based musical genre classification using multiple novelty functions. In *DAFx15 – 18th International Conference on Digital Audio Effects*.

McKay, C. (2010). *Automatic music classification with jMIR*. PhD thesis, McGill University.

Tolonen, T. and Karjalainen, M. (2000). A computationally efficient multipitch analysis model. *IEEE transactions on speech and audio processing*, 8(6):708–716.

Tzanetakis, G. (2002). *Manipulation, analysis and retrieval systems for audio signals*. Princeton University.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302.

Tzanetakis, G., Ermolinskyi, A., and Cook, P. (2003). Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research*, 32(2):143–152.

Tzanetakis, G., Essl, G., and Cook, P. (2001). Automatic musical genre classification of audio signals. In *Proceedings of the 2nd international symposium on music information retrieval, Indiana*, volume 144.