A Dataset of Norwegian Hardanger Fiddle Recordings with Precise Annotation of Note and Beat Onsets

TISMIR

OLIVIER LARTILLOT (iD)

MATS SIGVARD JOHANSSON

ANDERS ELOWSSON (iD)

LARS LØBERG MONSTAD (iD)

MATTIAS CYVIN

*Author affiliations can be found in the back matter of this article

ABSTRACT

The Hardanger fiddle is a variety of the violin used in the folk music of the western and central part of southern Norway. This paper presents a dataset of several hours of recordings of Hardanger fiddle music, with note annotations of onsets, offsets and pitches, provided by the performers themselves. A subset has also been annotated with beat onset positions by the performer as well as three expert musicians. The complexity of the music genre—polyphonic, highly ornamented and with a very irregular pulsation, among other aspects—motivated the design of a new annotation software adapted to these particular needs. Beat annotation in MIR is typically recorded as positions in seconds, without explicit connection with actual musical events. In the context of music where the rhythm is carried by the melodic instrument alone, a more reliable definition of beat onsets consists in associating them with the onsets of the notes that represent the start of each beat. This latter definition of beat onsets reflects that beats are generated from within the flow of played melodic-rhythmic events, which implies that the spacing of beats may be shifting and irregular. This motivated the design of a new method for beat annotation in Hardanger fiddle music based on a selection of notes in the note annotation. Comparisons between annotators through alignment—integrated in the interface—enable them to eventually correct their annotations or observe alternative valid interpretations of any given excerpt. After dedicating a part of the note annotation dataset to the training of a machine learning model, for the task of assessing both note pitch and onset time, an F1 score of 87% can be reached. The beat annotation dataset demonstrates the necessity of developing new beat trackers adapted to Hardanger fiddle music. The dataset as well as the annotation software is made publicly available.

1. INTRODUCTION

The work presented in this article is part of a project aimed at making a large collection of folk music recordings available to musicians, musicologists and to the public, augmented with computational music analysis and visualisation capabilities (Lartillot et al., 2022). The focus of the present work is on Hardanger fiddle music, a highlight of Norwegian folk music repertoire, presenting very peculiar and complex musical characteristics.

The first step of a transcription of audio recordings into scores consists of detecting notes, estimating their onset (start) and offset (end) time location in the audio recordings, as well as their pitch height, initially expressed simply in Hertz, then as a pitch height on a musical scale. There has been significant progress in the development of computational models for note detection in polyphony. But in order to properly assess their usefulness for the studied music corpus, it is necessary to establish a set of reference transcriptions, serving as "ground truth" both for systematic evaluation of the state of the art, and as training data for the development of improved or new models.

Despite the existence of available tools for manual note annotation from audio recordings, we found that none were adequate to our particular needs. For that reason, a new annotation tool has been developed, allowing to reach a high level of temporal precision in the assessment of onset time as well as pitch. Also, due to the complexity of the music, we found it more practical to ask expert musicians to record their own music and to annotate themselves what they played using the new annotation software.

Another subsequent step in music transcription concerns the inference of the underlying metrical structure: the beat positions and the grouping of beats into bars. Hardanger fiddle music is particularly challenging in that respect: in some styles, bars are made of beats of unequal duration, and the duration ratio within successive bars is generally variable. In addition, the beat onsets are often not clearly accentuated, and computational beat trackers fail to predict the beat correctly. This justifies the need to collect detailed beat annotation for future research. Even manual beat annotation is challenging, due to the reasons mentioned here, driving us to design an original method for beat annotation and comparison.

The article is structured as follows. Section 2 presents and discusses the peculiar musicological characteristics of Hardanger fiddle music. Section 3 addresses the state of the art concerning the various MIR tasks relating to this study. This includes computer-automated music analysis related to the detection of note and beat and bar positions, the annotation of datasets for polyphonic transcription and the associated annotation interfaces. Section 4 details the proposed method for precise note annotation, while Section 5 presents the original methodology developed for beat annotation. The annotation dataset is further described and its quality analysed in Section 6, discussing the possible ambiguity in the metrical analysis. Section 7 shows how the dataset can be applied to the evaluation and improvement of note and beat onset detection models, and can be used for musical analysis of the asymmetry of Hardanger fiddle metre.

2. HARDANGER FIDDLE MUSIC

2.1 GENERAL PRESENTATION

The Hardanger fiddle is a variety of the violin used in the folk music of the western and central part of southern Norway. It is traditionally played as a solo instrument for couple dancing. Its name comes from the area of Hardanger, from where it spread to many of the valleys in southern Norway in the 18th century.

It features a highly decorated body and fretboard, a short neck and a flat fingerboard and bridge, which supports the characteristic polyphonic playing-style with abundant use of drones, double stops and ornaments, as shown for instance in the transcription in Figure 1. Five sympathetic strings contribute to the richness of the resonating sound. Its impact on the precise note annotation task is discussed in Section 8.

The intricacy of the performance style makes machine transcription difficult, particularly when it comes to determining pitch, note onsets and rhythmic grouping, and discerning between melodic notes, drone notes and ornaments, and between musically essential sounds and noise.

The vast majority of Hardanger fiddle tunes are played in D tonality with the G string tuned up a whole step (A-D-A-E). The music is modal and pitch intonation patterns

Figure 1 Transcription of a Springar tune called *Fossekallen* (Blom et al., 1979).

are characterized by great variability, often deviating considerably from those found in European art music. Rather than being derived from particular scales or chord structures (cf. harmonic tonality), pitch intonations are conditioned by shifting contextual factors, including melodic formulas, local tonal centers (which often equal the open strings) and string resonance.

2.2 BEATS IN HARDANGER FIDDLE MUSIC

Historically, Hardanger fiddle music is used for dancing, recreation and ceremonial functions by rural Norwegian communities (Aksdal and Nyhus, 1993). Within the total repertoire, dance tunes are by far the most abundant. These are divided into types of triple meter (*springar*) and duple meter (gangar/rull/halling) types. Local styles are called dialects, indicating a parallel to the differences in spoken language between different regional districts. This conception concerns small but important (i.e. dialectal) distinctions in repertoire, playing style and dance style. These distinctions are most apparent within the springar genre, which includes differences in the basic meter between the various springar types. Some types have asymmetrical beat cycles following a short-long-medium or long-medium-short pattern, which introduces some additional analytical challenges (see below).

Since this is soloistic melody-driven music, beats are generated from within the flow of played melodic-rhythmic events rather than being represented by an accompanying instrument or similar. Structurally speaking, then, a beat typically consists of a quarter note, two eighth notes or a triplet, but which in reality are subjected to various degrees of temporal stretching and ornamentation. It should also be noted that the meter is not given from the succession of played beats as such, which is particularly evident in the springar genre. More concretely, the start of the metric cycle (the downbeat) and the subsequent accentuation pattern (how beats are experientially weighted) are often not evident from the melodic-rhythmic structuring of the tunes or how they are performed. These features are instead determined by the associated cycle of dance movements, which differs between the various sub-genres (Blom, 1981). For instance, in the so-called Tele-springar the down movement of the dancers' center of gravity corresponding to the downbeat occurs on the long beat (long-medium-short asymmetry), while in the Halling-springar the downbeat is located on the short beat (short-long-medium asymmetry). This means that one and the same tune (melodies often traveled between regions) can be heard as having the first beat on different locations. Notably, the fiddler's foot tapping represents this more generalized sense of "beat" (i.e. meter) that comprises both accentual (e.g. heavy-light-light) and temporal (short and long beats) properties or qualities.

Given that the relevant metrical framework (cf. above) is established, identifying played beats is in principle

relatively straightforward for an experienced listener, since each beat generally is represented by a confined melodic-rhythmic figure (a triplet, two eighth notes etc.). However, Hardanger fiddle music presents some particular challenges in this regard. Firstly, the start of the beat is sometimes ambiguous due to intentionally unclear onsets or multiple competing onsets occurring more or less simultaneously. Secondly, due to how melodic-rhythmic figures and their constituents are timed, phrased (note ties through the bowing) and accentuated, it is not always obvious how notes are grouped, which in turn may imply that it is unclear which note belongs to which beat. Thirdly, at times it is not clear whether a particular rhythmic articulation should be considered as the beat onset or as a syncopation against the beat onset (before or after). The reason why this is an issue is the emergent and context specific nature of rhythmic-temporal reference structures, particularly in the springar material. Concretely, the striking variation in beat duration from one measure to the next that often occurs is generally a product of the melodic-rhythmic structuring of the tune, including the associated stylistically idiomatic phrasing (Johansson, 2017). That is, these variations are generally not syncopations, but simply beats of varying duration. Identifying the exceptions, then, cannot rely on locating onsets against a reference grid. Instead, the particular melodic-rhythmic context in its totality needs to be considered. These ambiguities (microlevel and structural) are intentional and contribute to the desired smooth rhythmic feel of the Hardanger fiddle dance tunes (Johansson, 2022). However, they naturally present some significant challenges to the analyst.

3. RELATED WORK

Due to the complexity and particularity of Hardanger fiddle music, computational analysis is very difficult to undertake. Here we focus on two essential aspects of music analysis, at the core of music transcription: identifying notes and tracking the beats. There is a need to collect ground-truth dataset of note and beat annotations. For this annotation task as well, we need to go beyond the state of the art to fully approach the music style.

3.1 AUTOMATED NOTE DETECTION

Detecting notes from audio recordings is an MIR task, as part of the more general task of Automatic Music Transcription, that can be relatively well fulfilled on particular types of music using deep learning techniques (Elowsson, 2020; Hawthorne et al., 2021). Performance can reach 97% in F1 score metric for joint onset and pitch estimation in piano transcription, and 84% when adding offset estimation. For more challenging types of music,

in particular styles of music that are very different from those considered in the training phase, performance decreases significantly.

Ensuring good transcription results on a specific music style requires therefore to train the machine learning models using annotated data collected from a sufficiently large sample of that music. In addition, systematic evaluation of the models requires to compare the results with a collection of transcriptions carefully established and validated by music experts.

3.2 NOTE ANNOTATION

Researchers have used many different techniques to create annotated datasets for polyphonic transcription in the past. One method is to record individual voices in isolation to facilitate easier annotation. Examples include the four-voiced Bach10 dataset (Duan et al., 2010), the TRIOS dataset (Fritsch and Plumbley, 2013) consisting of musical trios, a five-voiced woodwind recording (Bay et al., 2019), the audio-visual URMP dataset (Li et al., 2018), and the MedleyDB multitracks dataset (Bittner et al., 2014). For polyphonic instruments, the annotation of many simultaneous notes can be cumbersome and time-consuming.

Another method for those kinds of instruments is to generate the sounds and annotations directly from MIDI. The technique has been used for piano datasets (Poliner and Ellis, 2006; Emiya et al., 2010; Hawthorne et al., 2019), but has also been applied across the full range of the general MIDI instrument specification (Elowsson, 2020). To increase the variability and the size of the dataset, researchers can use data augmentation, varying tempo, pitch, dynamics, and timbre during synthesis (Elowsson, 2020).

Although the MIDI generation strategy is appealing because of its efficiency, synthesized MIDI often lacks the full range of variation and complexities found in real performances. Researchers can in this case instead create datasets by synchronizing sheet music with an associated recording (Thickstun et al., 2017).

Sheet music cannot be considered for Hardanger fiddle music due to the improvisatory component of this music, making each performance of a given tune a bit different to the others. So each single performance leads to a specific transcription, which therefore needs to be manually obtained in the training phase. Musicians cannot be asked to play only single monodic voices of a given piece either, so the full polyphony needs to be transcribed altogether.

3.3 NOTE ANNOTATION INTERFACES

Sonic Visualiser (Cannam et al., 2006, 2010) is a popular open-source application that offers functionality for visualizing and annotating audio, as well as the dedicated modules: Sonic Lineup for rapid visualization of repeated performances, Tony (Mauch et al., 2015) designed primarily for solo vocal note transcription, and Sonic Annotator for non-interactive audio feature extraction in batch mode. Of particular relevance is Tony, which has a refined GUI allowing a large range of navigation, playback and editing processes.

Among commercial applications, Melodyne, originally developed for tuning monophonic vocal recordings, has evolved to also enable visualization and correction of polyphonic music. The software can be used as a tool for annotation by dividing the tone curves into MIDI notes, exporting the MIDI and converting it to annotations.

None of the existing note annotation interfaces offer the features we found necessary to quickly and efficiently annotate note onset with a very high level of time and pitch precision.

3.4 AUTOMATED BEAT TRACKING

State-of-the-art approaches in beat tracking are nowadays able to correctly track beats in a large range of music. But performance is generally much lower on musical content that differs from that which is contained in existing annotated datasets used for neural network training, as well as in the presence of challenging musical conditions such as rubato (Pinto et al., 2021). When considering music where the beats are not clearly accentuated and not regularly spaced in time, the estimations are often in total incongruity with the perceived beats.

Concerning the analysis of Hardanger fiddle music, we cannot expect those models to give correct results, without any training on that style of music because of the often non-isochronous and variable beat durations, and unclear or ambiguous note onsets and rhythmic groupings. In fact, for some of the more challenging styles in the Hardanger fiddle repertoire, non-expert (human) listeners themselves would typically not be able to follow the beats. This indicates, here also, the need to collect ground-truth beat annotations of Hardanger fiddle music from expert musicians, both for training and evaluation purposes. And indeed, as we will illustrate in Section 7.1.2, state-of-the-art approaches in beat tracking fail to corroborate with the collected beat annotations.

3.5 BEAT ANNOTATION AND INTERFACES

One way to record humans' judgement of beat location for a given piece of music is by asking them to tap to the music and record the temporal position of those tappings, often using Sonic Visualiser (Holzapfel et al., 2012, 2014; Mauch et al., 2009; McKinney and Moelants, 2006). The temporal precision of this data is questioned by the reliance on the reflexes and on the reactivity of the participants (Repp, 2005; Weiß et al., 2016). The decision of tapping at a given moment is based on many factors that gives some randomness to the process, and we might suppose that a tapping decision at a given instant

might be updated and possibly cancelled due to what has just been heard afterwards, for instance.

One solution is to manually correct the recorded taps in a subsequent step (Holzapfel et al., 2012; 2014). An alternative is to automate this correction step by "snapping" the taps to close-by audio cues (Driedger et al., 2019).

Another annotation method is based on inserting and editing beat positions from scratch. To guide the process, a background canvas shows a graphical representation of the sound (waveform or spectrogram) onto which annotated beat positions are shown with horizontal lines. One common software here also is Sonic Visualiser. Another method is to use an automated beat tracker as starting point, the beat positions being further corrected by human annotators (Eremenko et al., 2018; Marchand and Peeters, 2015; Davies et al., 2009).

3.6 ADDRESSING THE PARTICULARITIES OF HARDANGER FIDDLE MUSIC

Note detection in Hardanger fiddle music recordings cannot generally rely on reference music transcriptions; and as a matter of fact this is one objective of this study to produce music transcriptions of these recordings. The complexity of the playing style requires a versatile note annotation interface with well-thought capabilities to maximise precision, concerning the temporal position of the note onsets in particular. This motivated us to design our own graphical user interface for note annotation.

Concerning beat annotation, one particular challenge is that beats, as experienced by listeners, do not always correspond to particular individual note onsets. Instead

they would be better represented by considering the totality of interacting melodic-rhythmic events by which successive beats are composed. However, the most viable solution in practical terms, as we will discuss in Section 5.1, remains to locate the onsets of the particular notes that represent the start of each beat. This implies an alternative approach for beat annotation, based on selecting and annotating notes detected in the preliminary phase of note annotation.

4. NOTE ANNOTATION

4.1 METHODOLOGY

In our preliminary studies, we learned that it is rather time-consuming for Hardanger fiddle musicians to produce annotations for tunes that they are unfamiliar with, and accuracy may sometimes be lacking. For that reason, the chosen methodology was to ask the musicians to record tunes they are familiar with, and to annotate notes using computer assistance tools as aid. In this way, they could base the transcription not only on what they heard from the recordings, but also on a memory of what they actually played.

In order to avoid any bias, instead of offering the annotators the possibility to correct computer-generated annotations, the annotation (of the normal version, as explained below) is carried out entirely from scratch.

The recording and subsequent note annotations were made by three musicians: two music students, skilled fiddlers, from the Norwegian Academy of Music (S1 and S2) and one professional fiddler, Olav Luksengård Mjelva (P). Tables 1 and 2 list the tunes recorded and annotated

Title	Notes		Duration		Time Signature	Bars
	Normal	All 5	Normal	All 5		1 version
Fuglesangen	550	2778	1:05	6:04	3/4	42
Godvaersdagen	1121	5647	2:23	14:45	3/4	98
GroHolto	657	3389	1:15	7:58	3/4	51
Klunkelåtten	513	2575	0:57	5:49	3/4	37
Kongelåtten	951	4764	1:51	10:16	3/4	76
Langaakern	498	2505	1:00	5:58	3/4	40
Perigarden	648	3283	1:12	7:55	2/4	57
Solmøy	518	2580	1:04	6:21	3/4	44
Spretten	726	3576	1:22	8:12	3/8	125
Strandaspringar	707	3502	1:19	7:44	3/4	56
Tjednbalen	963	4799	2:03	10:58	3/4	83
Toingen	1401	7107	2:45	15:57	3/4	112
Total	**9428**	**46505**	**18:16**	**107:59**		**821**

Table 1 Tunes recorded by the professional musician (P). The number of notes and the duration of each tune are indicated for the Normal version as well as for the 5 expressive versions altogether.

Title	Musician	Notes		Duration	
		Normal	All 5	Normal	All 5
Haslebuskane	S1	566	2828	0:55	4:35
Havbrusen	S1	823	4114	1:45	8:50
Ivar Jorde	S2	334	1665	0:46	3:52
Låtten som bed om noko	S2	369	1819	0:58	4:51
Signe Uladalen	S2	448	2177	0:54	4:30
Silkjegulen	S1	578	2906	1:09	5:38
Valdresspringar	S2	331	1692	0:44	3:49
Vossarull	S1	504	2533	1:17	6:34
Total		**3953**	**19734**	**8:28**	**42:38**

Table 2 Tunes recorded by the students (S1 and S2). The number of notes and the duration of each tune are indicated for the Normal version as well as for the 5 expressive versions altogether.

by each musician. For each tune, five different versions were recorded: first playing in a normal way, and then following four distinct expressions: sad, angry, happy, and tender. For each tune, the normal version is first annotated by the musician; then the note annotations are automatically transferred to the other versions, and further checked and corrected by the musician (Elowsson and Lartillot, 2021).

The onset timing evaluation condition for polyphonic transcription is usually set to 50 ms.[1] Yet according to a study (Friberg and Sundberg, 1993), listeners can notice time displacements of just 10 ms. On the other hand, since fiddle music has rather undefined transients at onsets, a narrower margin than 20 ms is very hard to achieve. We therefore think that 20 ms is a reasonable threshold for very precise onset detection for this music. Hence, we encouraged performers to be very careful regarding onsets, and try to keep errors within 20 ms. This means that we can only allow a very narrow margin of error for the annotations to ensure that they can be reliably used for evaluation.

4.2 NOTE ANNOTATION INTERFACE

We developed a new Graphical User Interface in MATLAB, called *Annotemus*, to allow musicians to annotate each played note as easily and efficiently as possible, while maximising the degree of precision in the annotations. The software was made available to the musicians as a standalone application, which they can install and use on their own computers without MATLAB license nor IT support. Figure 2 shows a screenshot of the software.

The canvas of the annotation interface is a two-dimensional time-frequency representation, the "Pitchogram" (Elowsson, 2020), using a graphical representation of the fundamental frequencies detected in the sound as background image. This representation offers a relatively accurate representation of the temporal evolution of the pitch curves, even in polyphony.

Sporadically there can be local mistakes such as octave errors. The users of the software are advised to rely primarily on the sound as heard, and to use the graphical representation only as a practical—and not necessarily truthworthy—guide.

Each note has to be indicated as a horizontal line starting at its perceived onset and ending at its perceived offset, and located at a particular frequency.[2] The performer could use various key commands as an aid during annotation. This includes audio playback of the current window, playback between the start and end of one or several selected notes, playback that starts prior to a selected annotated note and ends at the annotated onset position, playback with a click at each annotated onset position, and playback with a synthesized version of the annotated score played in one of the stereo channels using an additive synthesis algorithm of our own. The performers were instructed to first try the playback that ends at the annotated onset position for locating the exact onset times for the normal recording and the click and synthesized functionality for verifying annotations, but were free to use whichever method they felt most comfortable with.

All playback functionality is offered with the option of slowing it down to an arbitrary speed selected by the annotator. Since Hardanger fiddle music contains frequent sequences of very fast note successions, the slowdown functionality was used extensively during the annotation process. It is also possible to listen to specific note fundamentals, separated from the other notes in the audio recording.

4.3 NOTE ANNOTATION COMPARISON

Sequences of note annotations are saved as CSV files. The software offers the possibility to automatically compare two sequences of note annotations of the same tune. One sequence is considered as the *reference sequence*, the one annotated by the user; the other sequence will be called the *alternative sequence*, corresponding
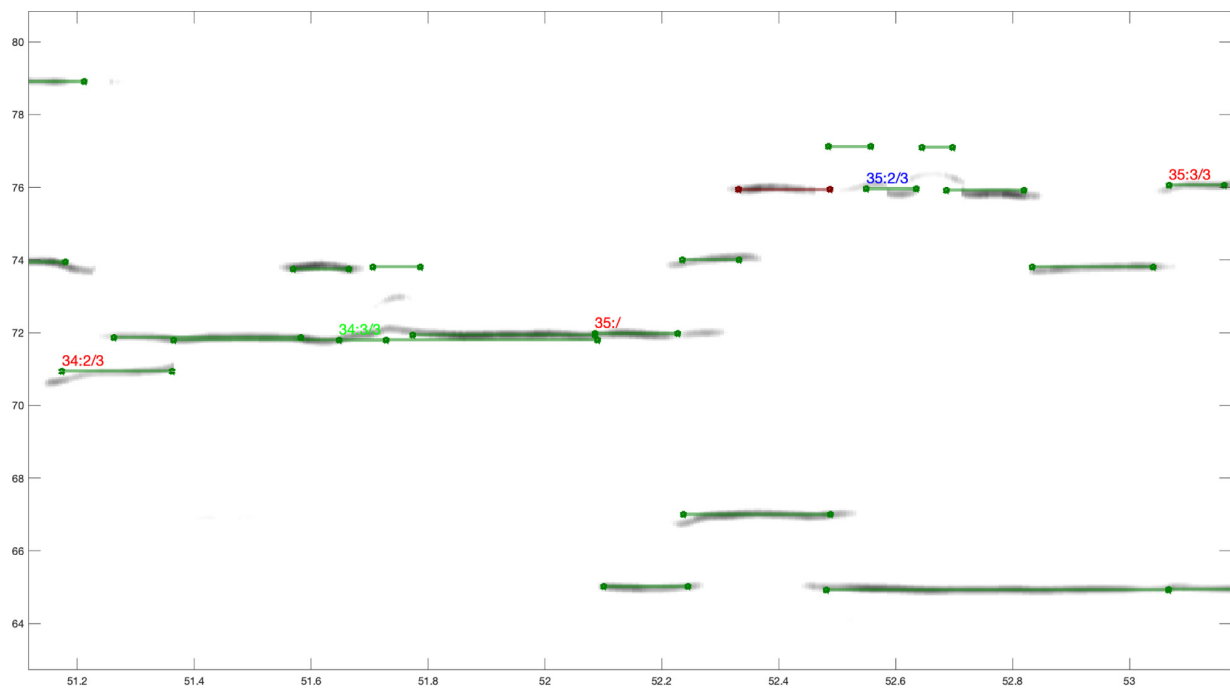
Figure 2 Screenshot of the software. In the background, the fundamental of each note is tracked over time and traced in grey, located as MIDI pitch over time in seconds. Note annotations are shown by green lines and beat annotations indicated above the selected notes, left-aligned with the note onset position. For instance the note starting around 51.2 seconds with pitch around 71 is the second beat of bar 34 (34 : 2/3). The screenshot shows the comparison between the beat annotations of P and S1. A misaligned note (missing in one of the annotations) is shown with a red line. Beat annotations common to both annotators are shown in red, while green and blue indicates annotations specific to one or the other annotator.

to either a sequence automatically generated, or from another user.

To align the notes, two dissimilarity matrices are computed based on onset time and pitch differences between all pairs of notes. The dissimilarity values are turned into similarity values using an inversion function. A dissimilarity of 0 ms or 0 cent corresponds to the maximum similarity, namely 1, while a dissimilarity of or higher than 130 ms or 75 cents is associated to the minimum similarity, i.e. 0, effectively cutting away larger dissimilarities. Dissimilarity values below those limits are transformed into similarity values following the curve of the descending slope of a Hann function of length 130 ms and 75 cents respectively. The two similarity matrices are then summed together. The alignment pairs are detected by considering the problem as a maximum weight bipartite matching problem, picking out elements from the matrix such that each row and column gets only a single non-zero but the sum of all the chosen elements is as large as possible. The resulting alignment consists of this series of alignment pairs, each associating one note from the first sequence with another note from the second sequence.

The comparison between the two sequences is based on the following color convention:

- A couple of note annotations from each sequence that are nearly identical (with an inter-onset interval below 35 ms and an inter-pitch interval below 40 cents) is displayed by only showing, in green, the note defined in the reference sequence, ignoring the variant in the alternative sequence. Thus in the case of nearly identical notes, the reference sequence is considered as the authoritative source of truth.
- A couple of note annotations that are more different are displayed with two blue lines. By manually deleting one of the lines, the other becomes accepted and thus green.
- Any isolated note annotation from one sequence that cannot be aligned to any other annotation from the other sequence is displayed with one red line. The note can be accepted in the annotation (thus turned green) or deleted.

As aforementioned, the initial note annotation of each tune was made by the musician themselves. The last author of this article, also expert in Hardanger fiddle music, checked the complete note annotations once again and made a few minor corrections when necessary. The comparison interface for note annotation is also used as part of the subsequent beat annotation task, as discussed in the next section.

5. BEAT ANNOTATION

5.1 NOTE-BASED BEAT ANNOTATION

As aforementioned, one particularity of Hardanger fiddle music is that beats are not regularly spaced in time.

Performers, listeners and dancers are not necessarily aware of this variability, implicitly feeling that variation as part of the rhythm/groove/beat of the tune (Johansson, 2022). This tolerance for fluctuations in beat durations also implies that the degree of precision in identifying the exact location of beat onsets may vary considerably without compromising musical interactions (Johansson, 2010). Beat tracking by participants tapping to the music is therefore not expected to produce a consistent output. For this reason it is preferable for our purposes to annotate the beats directly by identifying the note onsets that represent the start of each beat.

In this context it should also be noted that the experience of beats in Hardanger fiddle music is not comprehensively represented by their onset points. This is because beats comprise the complete melodic-rhythmic figures by which they are composed, and how beats are experienced is not only dependent on temporal features but also accentual qualities of performed rhythms (Johansson, 2022). In addition, the experienced temporal location of beat onsets does not always align with the played note that is associated with the beat onset. For instance, annotators have reported that there sometimes is a discrepancy between the preferred note-based beat annotation and where they experience the beat onset to be located (cf. Sections 2.2 and 6.2).

This also supports the relevance of the process of music transcription as a core principle for music analysis: considering beat tracking here as a process closely associated with turning the audio recording into a "symbolic" representation of music.

Sometimes beat onsets are not associated with any note onset: in some cases notes are tied across beats, meaning that there is no audible representation of the beat onset (melody note, ornament or bow onset). In our view, this silent beat, at the music discourse level, does not need to be associated with any particular time point, although the beat subdivisions before or after might need to be made explicit to ensure a temporal anchoring of that musical moment. Another related issue is syncopation, where the only note played around that beat onset is heard to be ahead of (or behind) the beat onset position.

The previous discussion justifies the proposed approach, based on collecting beat annotations as selection of particular notes in the annotations for which the onset of each successive note would correspond to the onset of each successive (non-silent) beat. Annotators were therefore asked to associate each successive beat with one annotated note. In the case several notes are played synchronously at a given beat onset, it does not matter which note is selected, as we do not find it necessary, at least for the current study, to address such level of temporal precision.

Annotators have been asked to skip silent beats, but with the possibility to annotate beat subdivisions at the vicinity of the silent beats. This is particularly emphasised

for the tune Spretten with 3/8 time signature, as it features a particular "two against three" rhythm, where the duration of two successive bars is divided into three equal rhythmic values. In such a case, instead of indicating the first beat of the second bar, which remains silent, the annotators are tasked to indicate the ternary sub-beats, i.e., the first beat of the first bar, the third beat of the first bar, and the second beat of the second bar, as illustrated in Figure 3.

It would be even more informative to ask annotators to indicate all possible subbeats, but it would be time expensive, and does not seem necessary, as long as only subbeats in the vicinity of silent beats are indicated. We hypothesise that the other subbeats could be rather easily predicted. But if necessary, annotation on a finer grain can be carried out in future work.

5.2 BEAT ANNOTATION INTERFACE

The interface for annotating beats has been designed to be as efficient and effortless as possible. Only the first beat needs to be associated with an explicit metrical position. For instance, if the first beat to annotate is the first beat of bar 1, and if there are 3 beats per bar, its metrical position would be annotated 1 : 1/3, which can be read as bar 1, beat 1 out of 3. To simplify, a first beat of a bar can be annotated by simply indicating the bar number in the form 1 : /. All subsequent beats are annotated by simply selecting a corresponding note associated to the beat onset. The corresponding metrical annotation (such as 1 : 2/3 for the second beat of bar 1) is inferred and displayed automatically.

Silent beats, as discussed in Section 5.1, are not annotated as such, but subbeats at the vicinity can be annotated instead. For instance, in the musical example displayed in Figure 3, with a peculiar time signature 3/8 indicating one single (ternary) beat per bar, the beat of bar 1 is annotated 1 : /, and the beat of bar 2, being silent, is not annotated. On the other hand, the third subbeat of bar 1 is annotated (1 : 3/3[3]), as well as the second subbeat of bar 2 (2: 2/3).

When playing back the annotation of the tune, the annotated beats are sonified with short burst of noise, with a dynamic accent on the first beat of each bar.

5.3 INTERFACE FOR BEAT ANNOTATION COMPARISON

5.3.1 Comparing own annotation with reference annotation

Similarly as for note annotation, the software offers the possibility to compare two sequences of beat annotations related to the same tune. Since beat annotation is based

Figure 3 Schematic realisation of a "two against three" rhythm, spanning two bars, repeated twice.

on note annotation, we need first of all to make sure that the sequences of note annotations are aligned too. In principle, all beat annotators were given as input the same sequence of note annotations of the tune, namely, the version manually annotated by the performer themselves.[4] But if necessary, the beat annotators are free to modify or add any note. It happened also that notes were added by mistake while annotating beats due to an imperfection in the interface. The alignment of the sequences of note annotations is performed using the method presented in Section 4.

Similarly to note annotation, the comparison of the two sequences of beat annotations is displayed graphically:

- If the same aligned note in both sequences is associated with the same beat annotation, it is shown with its default red color.
- When a note is beat-annotated on one annotation sequence only, it is shown either in green or blue, depending on the sequence it comes from. This color differentiation enables the beat annotator to precisely see which beat comes from their own annotation, and which from the performer's annotation.
- As an exception to the previous rule, if the two notes in the two sequences associated with a beat annotation are temporally distant by less than 40 ms, they are considered synchronous, and therefore not in conflict. They are both shown in the default red color, with the later of the two notes' beat annotations in a lighter shade.

It is possible to listen to either of the two beat sequences (i.e., the one annoted in red and green, or the one in red and blue).

The beat annotators' task is to check whether they would still consider each of their own beat annotations as congruent to their own understanding of the music (implying therefore that the other annotation from the performer gives another understanding), or if on the contrary they made any mistake which they need to correct. To make the process as time-efficient as possible, the annotators only need to correct any mistake by modifying one of their annotations or replacing it with the one from the performer. For all the other conflicting beat annotations left untouched, the beat annotator's version will remain and the performer's one is ignored.

5.3.2 Superposing multiple beat annotations

For the final check of all beat annotations of the same tune, we added the possibility to load each of them and superpose the beat annotations altogether on a single note annotation of the tune. An automated check is carried out to test whether note annotations have been modified by beat annotators, and resolution of these conflicts is made available.

5.4 CAMPAIGN FOR BEAT ANNOTATION AND COMPARISON

We have so far focused on the 12 tunes (totalling 18 minutes) played by the professional musician, as shown in Table 1. To study the degree of agreement among music experts concerning beat annotations, each tune has been annotated by three experts in addition to the professional musician (P):

- a Scandinavian folk music scholar and fiddle music expert, Mats Sigvard Johansson (M),
- two music students from the University of South-Eastern Norway (S3 and S4), expert in Hardanger fiddle music.

These experts were subsequently asked to compare their annotation with the version by the musician. For each successive beat:

1. If the annotated notes are different but considered by the expert as synchronised, they are asked to ignore that divergence.
2. If the annotated notes are different and considered as non-synchronised, if the expert thinks that both note onsets offer plausible alternative beat positions, they can leave the divergence unchanged.
3. If the expert thinks that they made a mistake and that the musician's annotation is correct, they can delete their own annotation.
4. Inversely, they can delete the musician's annotation.

This enables to obtain beat annotations for which the variability of the expert (with respect to the musician) has been fully reflected by the expert themselves, so that divergence can be considered as valid alternative beat grids.

One of the annotators, S4, compared all the beat annotations of each tune. The superposition of multiple beat annotations, made possible by the tool presented in Section 5.3.2, helps to establish a synthetic overview of the comparison throughout the whole corpus, as presented in Section 6.2. Besides, S4 has combined the multiple versions of beat annotations of each tune into one single authoritative version, using the pairwise comparison tool, presented in Section 5.3.1, observing the following methodology:

- Two versions (for instance P and M) are compared. Any divergence in note annotations is corrected. Any divergence in beat annotations is resolved using the four rules (1–4) stated above. Rule 2 allows multiple annotations of the same beat.
- The resulting beat annotation is then further compared with another version (for instance S3) in the same way.
- Same for the fourth version (for instance S4).

6. DESCRIPTION OF THE DATASET

The dataset consists so far of the 12 tunes (totalling 18 minutes) recorded five times with distinct expressions by the professional musician (P), as shown in Table 1, with the corresponding note annotation (initially made by the musician P himself, but further checked by S4). In addition, each tune is associated with its beat annotation (initially by P, M, S3 and S4, and further checked and fused into an authoritative version by S4). This is completed with 8 tunes recorded by students S1 and S2, each note-annotated by its corresponding performer.

The dataset is published[5] in the form of a repository in the Open Science Framework (OSF). The dataset contains:

- the recordings as audio files;
- the note and beat annotations as CSV files, which can be read also using the annotation software;
- documentation.

The audio recordings are in stereo, in WAV format, with a sampling rate of 44100 Hz and a bit depth of 16 bits. P's recordings were carried out in a studio in an old log building with natural "wooden house acoustics", using WA84 stereo mics, U67 replica room mics, an Audient asp880 preamp and the Presonus Firepod interface. S1's recordings were carried out in a room with wooden walls, with relatively normal acoustics, using a Zoom H6 recorder. S2's recordings were carried out in a small and relatively dry room using a Zoom recorder. The musicians were not specifically asked to tune their fiddle to a specific diapason, leaving them to proceed as more natural for them.

The annotation software *Annotemus* is released as a free standalone program,[6] developed and compiled using MATLAB. Installing the software also installs the free MATLAB Runtime libraries. The source code of the annotation software is not made public.

6.1 QUALITY OF THE NOTE ANNOTATIONS

The musicians have been encouraged to try to achieve a precise annotation of the note onsets, aiming at a precision around 20 ms, which seems to be the maximum possible precision due to the inherent imprecision of Hardanger fiddle attacks in general. The precision of the annotations can be qualitatively observed using the slowdown and multiple sonification capabilities provided by the annotation interface. Observations made by the other beat annotators confirm the high quality of the note annotations. Just a few notes needed to be corrected, mostly due to some possible bugs in preliminary versions of the note annotation interface. These corrections have been carried out as part of the beat annotation task itself.

Concerning offset annotation, due to the longer and richer resonance caused by the sympathetic strings, as discussed in Section 2.1, we did not aim for very high precision.

6.2 AGREEMENT BETWEEN BEAT ANNOTATORS

The superposition of alternative beat annotations, presented in Section 5.3.2, is used to compare between the different annotators. Generally, there is strong agreement between annotators on a structural or macro level, which includes the location of the downbeat (the start of the rhythmic cycle) and determining which notes belong to which beat. However, there are some exceptions. First, there was one instance where one annotator located the downbeat (first beat) on the second beat, thereby skewing all the subsequent beat annotations for that tune (Figure 4). This may be due to unfamiliarity with the particular style in question (Halling-springar) and since the annotator agreed that it was a mistake we do not consider this to be an example of disagreement between annotators. It should be noted, however, that the wrong interpretation makes perfect sense from the point of view of melodic/motivic organization, which highlights the need for expertise in determining the correct metrical framework.

In Section 2.2 we also anticipated that there might be discrepancies between annotators due to ambiguous rhythmic grouping. We have recorded several instances of this, one of which is illustrated in Figure 5. In this case, there is no correct alternative as both versions are

Correct placement of downbeat   Wrong placement of downbeat

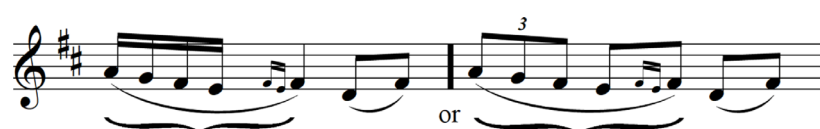**Figure 4** Klunkelåtten, the start of the tune.

or

Figure 5 Fuglesangen 00:26 (bar 17) and 00:32 (bar 21).

stylistically viable. What can be noted is that the two interpretations may be associated with two different interpretational rationales. When phrasing and melodic organization are considered, the interpretation to the left is arguably preferable. But this results in a very long first beat (the first beat in Halling-springar is generally short), meaning that the interpretation to the right makes more sense from a metrical perspective.

On a microlevel we found a number of discrepancies between the annotators concerning the exact location of the beat onset. Most of these fall into the following two categories: beat onsets defined by or surrounded by ornamentation (third beat in Figure 6); beat onsets consisting of a double stop where the two note onsets are asynchronous, producing an effect resembling sliding into a note (second beat in Figure 6). The material contains numerous versions of ornamented beat onsets as well as various configurations of associated discrepancies between annotators that will be explored in later work.

7. APPLICATION OF THE DATASET

7.1 GROUND TRUTH FOR COMPUTATIONAL MODELS

7.1.1 Note detection

The 12 tunes recorded by P have been used to test a set of commercially available polyphonic note detection tools:

- Celemony Melodyne 5.3.0
- Logic Pro 10.7 Flex Pitch
- ScoreCloud
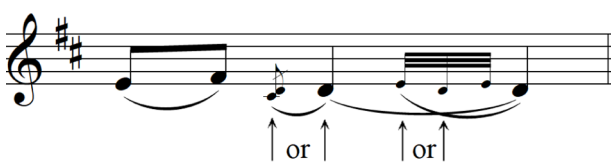
Figure 6 Solmøy 00:56 (bar 40).

as well as the "deep layered learning" (DLL) model presented by Elowsson (2020). We also had the opportunity to train the DLL model on the rest of our dataset—i.e., the tunes performed and annotated by S1 and S2. We have tested both the original version of the DLL model from Elowsson (2020), trained on classical music, as well as the version we trained on S1 and S2's annotations.

Both onset and pitch of each note are estimated. The output from each method, represented as a list of notes with their corresponding onset and pitch, is evaluated by aligning it with the list of notes from the ground-truth sequence, and then assessing the number of notes that can be successfully aligned. The alignment is based on the method presented in Section 4.3. All aligned notes are considered as correct, or true positives, the rest defining the false positives and negatives, leading to precision, recall and F1 scores.

Figure 7 shows the evaluation results for each algorithm, for each separate tune. Statistics across tunes are shown in Table 3. The Logic Pro module is clearly irrelevant for this task, while Melodyne gives more satisfying results, although the three remaining models are significantly more successful. Scorecloud and the original DLL model from Elowsson (2020) have relatively similar F1 values, although ScoreCloud excels in precision while DLL ensures a high recall, at the expense of precision. On the other hand, when training DLL on a subset (S) of the dataset, we can achieve much better results on the other dataset (P), reaching an F1 score of 87%.

7.1.2 Beat tracking

The collected beat annotations have been used to test the latest available version 0.16.1 of the reference beat tracking software Madmom (Böck et al., 2016). For all the tunes, the predicted beat positions do not correspond to the annotated beats. Annotated and predicted beat onsets do coincide at isolated places, but because they do not extend along successive beat onsets, these coincidences cannot be considered as metrical convergence. Figure 8 illustrates the characteristic behavior of the beat tracker. We might hypothesise that this kind of mistake could
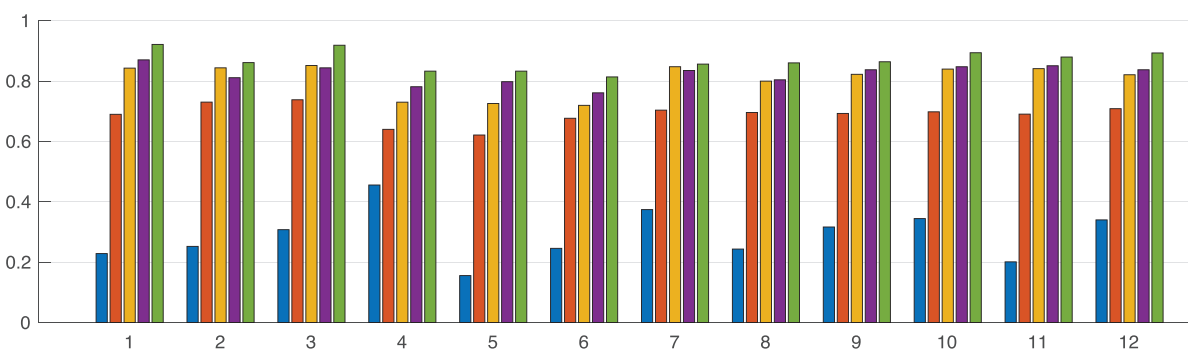

Figure 7 F1 statistics of the evaluation of the following note annotation models, from left (blue) to right (green): Logic Pro 10.7 Flex Pitch, Celemony Melodyne 5.3.0, ScoreCloud and DLL, both its original version and the one trained on the S dataset. The evaluation was made on each separate tune in the P dataset, following the same order as in Table 1, represented here from left to right.

relate somehow to the way non-expert listeners would tap beats to Hardanger fiddle music.

The collected beat annotations will therefore prove useful when training machine learning models on that particular music genre, and more generally when designing and testing beat trackers dedicated to this music.

## 7.2 MUSICAL ANALYSIS OF ASYMMETRICAL METRE

The beat annotations of the performances by the professional musician (P) enable us to study timing characteristics of Hardanger fiddle music.

The most conspicuous observation is that tempo remains relatively stable across all tunes, with a mean bar duration of 1.399 seconds, a standard deviation of 96 ms, and a mean absolute difference between successive bars of 84 ms. Indeed this is dance music and the music is expected to be played in a certain standard tempo. There is variability, still, corroborating the musician's indication that he felt he played some tunes faster or slower than others.

A more detailed look at the temporal evolution of bar duration across each tune (Figure 9) shows that the first bar is always the longest, and much longer than any other bar, exceeding sometimes 2 s. The rest of the bars oscillate

Model	F1	Precision	Recall
Logic Pro 10.7 Flex Pitch	29	43	22
Celemony Melodyne 5.3.0	69	82	60
ScoreCloud	81	95	71
DLL	82	83	82
DLL trained on S	87	91	83

Table 3 Evaluation of the note annotation models on the P dataset, in percentage.
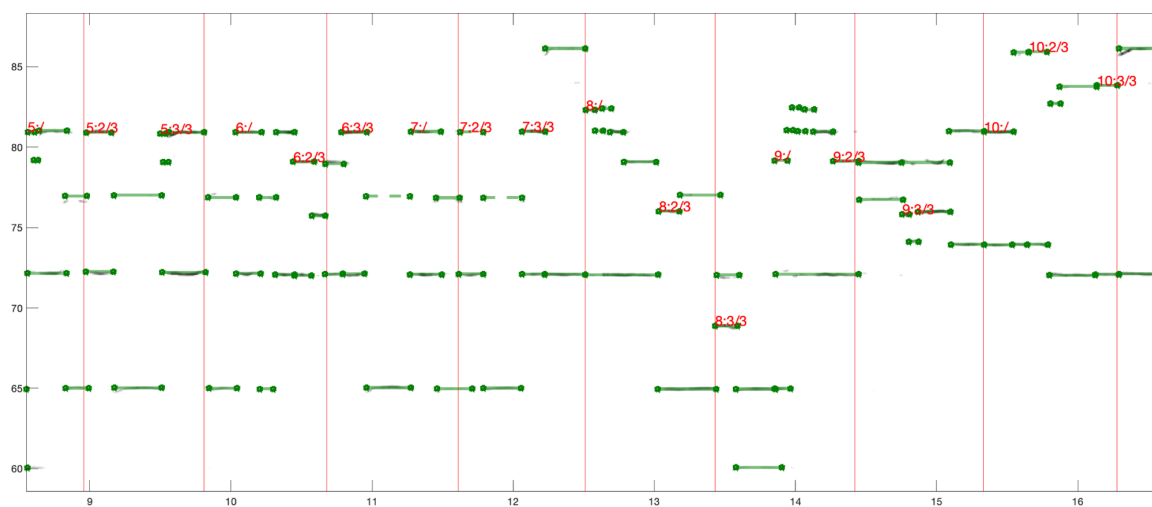
Figure 8 Screenshot of the Annotemus software (cf. Figure 2) comparing the beat annotations of an excerpt of Gro Holto provided by the musician P himself (notes annotated with the red text) with the beat predicted by Madmom 0.16.1 (Böck et al., 2016) (indicated with red vertical lines).
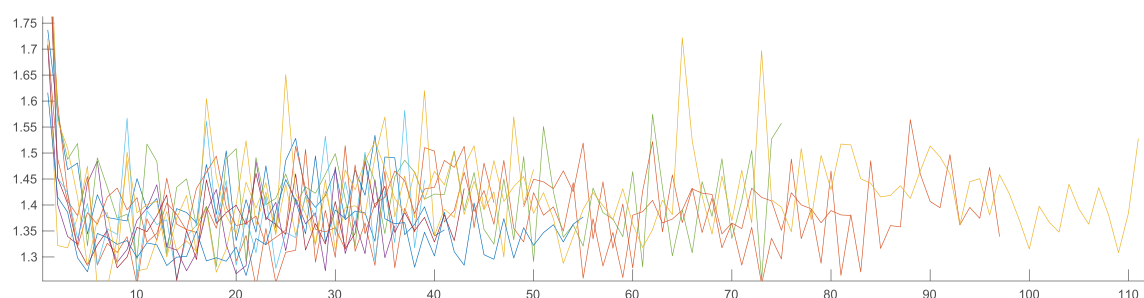
Figure 9 Bar duration, in seconds on the Y axis, along each tune played by P. Each curve corresponds to a different tune, with the bars spanning from left to right, and numbered on the X axis.

between 1.3 and 1.5 s, occasionally with an isolated bar with larger duration, up to 1.7 s. This invariance allows us to study beat duration both absolutely and relatively to bar duration, which is convenient due to the respective pros and cons of these two viewpoints.

The other, and rather contrastive, prominent observation is that beat duration is overall very irregular. The average beat duration[7] is 466 ms — i.e., around 128 BPM — but with a standard deviation of 80 ms, and a mean absolute difference between successive beats of 103 ms, which is more than 22% of the mean beat duration. However some invariance can be observed when distinguishing the first, second and third beat of each bar. Table 4 shows their mean and standard deviation across all tunes, as well as the mean absolute difference for the same beat between successive bars, comparing it to the same statistics across all beats. With respect to overall standard deviation, the first beat of bar has much higher deviation (72 ms, compared to 65 and 51 ms for the second and third beats). But actually when considering mean absolute difference between successive bars, the second beat has slightly more variability (71 ms, compared to 69 and 56 ms for the first and third beats). This is due to the fact that the first beat of the first bar is often very long.

Figure 10 shows the histogram of beat duration, also distinguishing between first, second and third

beats in bars. We notice that the distribution of first beat duration is rather multimodal, with a first cluster around 25% and a last cluster around 35% of the bar duration. This phenomenon can be analysed in more detail by observing, beyond those global statistics, the temporal evolution of beat duration ratio along single tunes. Figure 11 shows this for one particular tune, namely Godvaersdagen. The first beats of bars 1 and 2, respectively short and long, correspond to distinct modes. They actually form one 2-bar phrase, which is subsequently repeated, leading to the same short-long pattern in bars 3 and 4. Other oscillations between short and long first bars can be observed throughout the piece. In addition, we notice progressive duration changes for each separate mode across successive bars.

Beat-level variations in the asymmetrical timing patterns of springar performances seem to be related to "melodic-rhythmic" structures, in the sense that particular motivic segments are associated with particular timing profiles, suggesting that structural and other expressive features influence beat duration patterns (Johansson, 2017). Inspired by this perspective, we are conceiving software prototypes in order to offer structural and multidimensional perspectives on the complex rhythmical structuring of HF springar performances.

Beat in bar	1st	2nd	3rd
Mean	399	489	512
Standard deviation	72	65	51
Mean absolute difference between successive bars	69	71	56

Table 4 Beat duration statistics, in milliseconds, on the P dataset, where beats are categorised with respect to their position in bar.
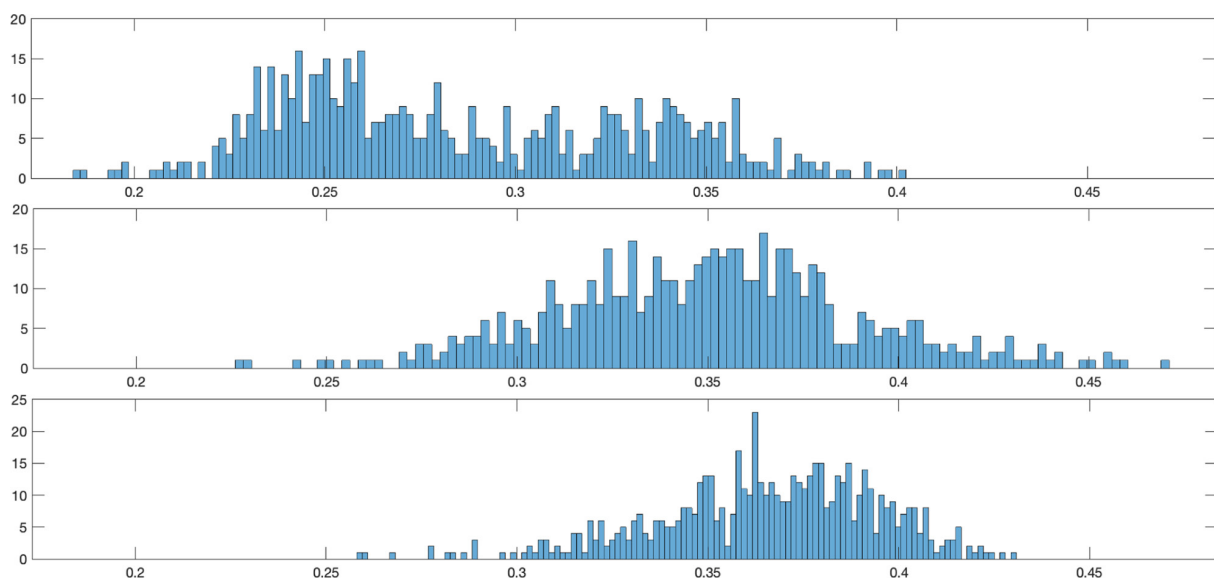
Figure 10 Histograms of beat duration (expressed as a ratio with respect to bar duration), distinguishing between first (top), second (middle) and third (bottom) beats in each bar.
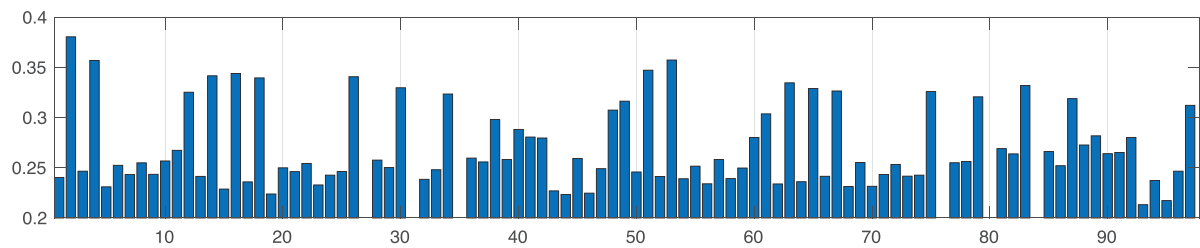
Figure 11 Duration of the first beat of each successive bar of the tune Godvaersdagen. Beat duration is normalised with respect to bar duration. The second beat onset of bars 27, 31, 35, 76, 80 and 84 being silent, the duration of the first beat for those bars is undefined and thus ignored.

8. CONCLUSIONS

Towards the automated transcription of a challenging type of music such as Hardanger fiddle music, a series of challenging steps needs to be carefully addressed. This article focused on two fundamental steps: 1. detecting the notes from the audio recording and precisely characterising their timing; 2. inferring the metrical grid by estimating the temporal position of bar and beat onsets.

Concerning the first step, we demonstrate the possibility of obtaining a very detailed note onset annotation of one hour of Hardanger fiddle music. The high level of precision has been ensured by asking the musicians themselves to annotate the notes from their own recorded performances, and more importantly by designing new annotation software specifically dedicated to this aim. Although existing annotation software such as Sonic Visualiser allows the annotation of notes, represented on top of a spectrogram, our proposed software Annotemus enables very accurate indication of note onset and offset timing thanks to a panoply of sonification tools, providing a kind of audio magnifier, while also offering ways to isolate notes within a polyphony. The software enables us also to compare and resolve conflicts between alternative annotations of the same piece. The note annotation dataset is shown to allow precise evaluation of state-of-the-art note annotation software, and can be used to train machine learning models to improve performance, reaching in our experiment an F1 score of 87%. The resulting model can be improved even further by using the whole dataset as training data.

The amplification of the resonance by the addition of the five sympathetic strings did not impede note onset estimation. On one hand, the resonating sound might blur the graphical representation of fundamental frequency over time, as discussed in Section 4.2. On another hand, expert musicians did not manifest any disturbance when asked to annotate note onset positions from the audio recording. Similarly, machine-learning-based models for note onset detection can reach a rather high F1 score on this type of sound,

after being trained on manual annotations of similar performances.

The second main step studied in the paper, namely beat and bar onset tracking, is a particularly challenging problem in the case of Hardanger fiddle music, due to the rhythmical peculiarities of this music. To answer this problem, we developed a new method for beat onset formalisation of potential interest for a large range of music outside our specific corpus. This starts from a distinction between beat onsets estimated by tapping—which correspond to the way beats are traditionally considered in MIR research—and beat onsets determined by identifying played note onsets. The latter allows temporal annotations with low variance, as can be seen by the large degree of consensus between the four expert annotators who participated in the study. It is thus particularly adapted to our underlying objective of music transcription. Beat and bar annotation capabilities have been integrated into the annotation interface, demonstrating here also the interest of developing our own annotation solution. The beat annotation dataset is of high interest for musicological analysis as well as, potentially, in the establishment of beat trackers adapted to this music.

We are investigating the subsequent steps toward automated music transcription. Pitch height, initially expressed in Hertz, needs to be expressed as a degree within a musical scale. The metrical position of each note needs to be determined. Finally, the polyphony needs to be structured into a superposition of two melodic lines. From the obtained music transcription, higher-level musicological analysis, such as modal and motivic analysis are considered. One objective of the project is to apply these tools to a large collection of audio recordings of Norwegian folk music (Lartillot et al., 2022).

By providing the data to the research community, we hope that it will also encourage MIR research to take into account the specificity of this particular repertoire. At the same time, the methodological and technological framework that has been developed in the context of this project can be reused for the establishment of other music transcription datasets. The annotation software can be directly used for annotating notes from

recordings of polyphonic music of a large range of music genres and cultures. Additional musical characteristics not yet taken into account in the current version of the software, such as glissandi and portamenti, can be further added in collaboration with potential users of this extended software. Addressing the rhythmical particularities of specific music repertoires might require extensive investigation. In addition, if there is a motivated need for it, annotation of felt beats can be implemented. Besides sharing the software tools, we plan to also share the insights we gained while addressing the fusion of multiple and fallible annotation sources into one single, consensual and more reliable source of authority.

NOTES

1 MIREX 2020: Multiple Fundamental Frequency Estimation & Tracking. https://www.music-ir.org/mirex/wiki/2020:Multiple_Fundamental_Frequency_Estimation_%26_Tracking.

2 Glissandi and portamenti are uncommon in the Hardanger fiddle style; their annotation has therefore not been taken into consideration.

3 While in the previous paragraph the denominator 3 related to beats, because there were 3 beats per bar, the denominator 3 relates here to the ternary subbeat, because there is one single beat per bar, itself decomposed into 3. The ambiguity is resolved with the indication of the time signature in the annotation file.

4 The verification of the note annotations, as discussed in Section 4.3, was performed after the beat annotation campaign.

5 https://doi.org/10.17605/OSF.IO/935DQ.

6 https://www.uio.no/ritmo/english/projects/mirage/software/annotemus/.

7 In the statistical analysis of beat duration presented in this section, silent beats, as discussed in Section 5.1, are ignored. Hence when a beat onset is silent, the duration of both that beat and the previous one is ignored. In addition, both tunes Spretten and Perigarden are excluded from the statistics, as they do not have 3 beats per bar.

ACKNOWLEDGEMENTS

FUNDING INFORMATION

COMPETING INTERESTS

Olivier Lartillot is an Editor of the Transactions of the International Society for Music Information Retrieval. He was removed completely from all editorial processing. The authors have no other competing interests to declare.

AUTHOR AFFILIATIONS

Olivier Lartillot orcid.org/0000-0001-6129-686X
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway

Mats Sigvard Johansson
University of South-Eastern Norway, Norway

Anders Elowsson orcid.org/0000-0002-4957-2128
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway

Lars Løberg Monstad orcid.org/0009-0003-5675-2033
RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway

Mattias Cyvin
University of South-Eastern Norway, Norway

REFERENCES

Aksdal, B., and **Nyhus, S.** (eds.) (1993). *Fanitullen. Innføring i Norsk og Samisk Folkemusikk*, Oslo, Universitetsforlaget.

Bay, M., Ehmann, A. F., and **Downie, J. S.** (2019). Evaluation of multiple-F0 estimation and tracking systems, *Proceedings of the International Society for Music Information Retrieval Conference*, pages 315–320.

Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and **Bello, J. P.** (2014). MedleyDB: A multitrack dataset for annotation-intensive MIR research, *Proceedings of the International Society for Music Information Retrieval Conference*, pages 155–160.

Blom, J.-P., Nyhus, S., and **Sevåg, R.** (eds.) (1979). *Hardingfeleslåttar Springar i 3/4 Takt = Slåttar for the Harding Fiddle Springar in 3/4 Time*, 6, Oslo, Universitetsforlaget.

Blom, J.-P. The dancing fiddle, in Blom, J.-P., Nyhus, S., and Sevåg, R. (eds.) (1981). *Slåttar for the Harding Fiddle: Norwegian Folk Music*, 7:305–312, Oslo, Universitetsforlaget.

Böck, S., Korzeniowski, F., Schlüter, J., Krebs, F., and **Widmer, G.** (2016). Madmom: A new Python audio and music signal processing library, *Proceedings of the 24th ACM International Conference on Multimedia*, pages 1174–1178. DOI: https://doi.org/10.1145/2964284.2973795

Cannam, C., Landone, C., and **Sandler, M. B.** (2010). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files, *Proceedings of the 18th ACM International Conference*

on Multimedia, pages 1467–1468. DOI: https://doi.org/10.1145/1873951.1874248

Cannam, C., Landone, C., Sandler, M. B., and **Bello, J. P.** (2006). The Sonic Visualiser: A visualisation platform for semantic descriptors from musical signals, *Proceedings of the International Conference on Music Information Retrieval,* pages 324–327.

Davies, M., Degara, N., and **Plumbley, M.D.** (2009). Evaluation Methods for Musical Audio Beat Tracking Algorithms, *Technical Report C4DM-TR-09-06, Queen Mary University of London.*

Driedger, J., Schreiber, H., Bas de Haas, W., and **Müller, M.** (2019). Towards automatically correcting tapped beat annotations for music recordings, *Proceedings of the International Society for Music Information Retrieval Conference.*

Duan, Z., Pardo, B., and **Zhang, C.** (2010). Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions, *IEEE Transactions on Audio, Speech and Language Processing,* 18(8):2121–2133. DOI: https://doi.org/10.1109/TASL.2010.2042119

Elowsson, A. (2020). Polyphonic pitch tracking with deep layered learning, *Journal of the Acoustical Society of America,* 148(1),446–468. DOI: https://doi.org/10.1121/10.0001468

Elowsson, A., and **Lartillot, O.** (2021). A Hardanger fiddle dataset with performances spanning emotional expressions and annotations aligned using image registration, *Proceedings of the International Society for Music Information Retrieval Conference.*

Emiya, V., Badeau, R., and **David, B.** (2010). Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle, *IEEE Transactions on Audio, Speech, and Language Processing,* 18(6):1643–1654. DOI: https://doi.org/10.1109/TASL.2009.2038819

Eremenko, V., Demirel, E., Bozkurt, B., and **Serra, X.** (2018). Audio-aligned jazz harmony dataset for automatic chord transcription and corpus-based research, *Proceedings of the International Society for Music Information Retrieval Conference,* pages 483–490.

Friberg, A., and **Sundberg, J.** (1993). Perception of just-noticeable time displacement of a tone presented in a metrical sequence at different tempos, *Journal of The Acoustical Society of America,* 94(3):1859–1859. DOI: https://doi.org/10.1121/1.407650

Fritsch, J., and **Plumbley, M. D.** (2013). Score informed audio source separation using constrained nonnegative matrix factorization and score synthesis, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing,* pages 888–891. DOI: https://doi.org/10.1109/ICASSP.2013.6637776

Hawthorne, C., Simon, I., Swavely, R., Manilow, E., and **Engel, J.** (2021). Sequence-to-sequence piano transcription with transformers, *Proceedings of the*

International Society for Music Information Retrieval Conference, pages 246–253.

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., Elsen, E., Engel, J., and **Eck, D.** (2019). Enabling factorized piano music modeling and generation with the MAESTRO dataset, *Proceedings of the International Conference on Learning Representations.*

Holzapfel, A., Davies, M. E. P., Zapata, J. R., Oliveira, J., and **Gouyon, F.** (2012). Selective sampling for beat tracking evaluation, *IEEE Transactions on Audio, Speech and Language Processing (TASLP),* 20(9). DOI: https://doi.org/10.1109/TASL.2012.2205244

Holzapfel, A., Krebs, F., and **Srinivasamurthy, A.** (2014). Tracking the "odd": Meter inference in a culturally diverse music corpus, *Proceedings of the International Society for Music Information Retrieval Conference,* pages 425–430.

Johansson, M. (2010). *Rhythm into Style: Studying Assymmetrical Grooves in Norwegian Folk Music.* PhD Thesis, University of Oslo.

Johansson, M. (2017). Empirical research on asymmetrical rhythms in Scandinavian folk music: A critical review, *Studia Musicologica Norvegica,* 43(1):58–89. DOI: https://doi.org/10.18261/issn.1504-2960-2017-01-05

Johansson, M. (2022). Timing-sound interactions: Groove-forming elements in traditional Scandinavian fiddle music, *Puls – Journal for Ethnomusicology and Ethnochoreology,* 7.

Lartillot, O., Elowsson, A., Johansson, M., Thedens, H.-H., and **Monstad, L.** (2022). Segmentation, transcription, analysis and visualisation of the Norwegian folk music archive, *Proceedings of DLfM '22: 9th International Conference on Digital Libraries for Musicology.* DOI: https://doi.org/10.1145/3543882.3543883

Li, B., Liu, X., Dinesh, K., Duan, Z., and **Sharma, G.** (2018), Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications, *IEEE Transactions on Multimedia,* 21(2):522–535. DOI: https://doi.org/10.1109/TMM.2018.2856090

Marchand, U., and **Peeters, G.** (2015). Swing ratio estimation, *Proceedings of the International Conference on Digital Audio Effects (DAFx).*

Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., Bello, J.P., and **Dixon, S.** (2015). Computer-aided melody note transcription using the Tony software: Accuracy and efficiency, *Proceedings of the First International Conference on Technologies for Music Notation and Representation (TENOR).*

Mauch, M., Cannam, C., Davies, M. E. P., Dixon, S., Harte, C., Kolozali, S., Tidhar, D., and **Sandler, M.** (2009). OMRAS2 metadata project 2009, Late-breaking session at the *International Society for Music Information Retrieval Conference.*

McKinney, M. F., and **Moelants, D.** (2006). Ambiguity in tempo perception: What draws listeners to different metrical levels?, *Music Perception,* 24(2):155–166. DOI: https://doi.org/10.1525/mp.2006.24.2.155

Pinto, A.S., Böck, S., Cardoso, J.S., and **Davies, M.E.P.** (2021). User-driven fine-tuning for beat tracking, *Electronics*, 10(13):1518. DOI: https://doi.org/10.3390/electronics10131518

Poliner, G. E., and **Ellis, D. P.** (2006). A discriminative model for polyphonic piano transcription, *EURASIP Journal on Advances in Signal Processing.* DOI: https://doi.org/10.1155/2007/48317

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature, *Psychonomic Bulletin &*

Review, 12(6):969–992. DOI: https://doi.org/10.3758/BF03206433

Thickstun, J., Harchaoui, Z., and **Kakade, S.** (2017). Learning features of music from scratch, *Proceedings of the International Conference on Learning Representations.*

Weiß, C., Arifi-Müller, V., Prätzlich, T., Kleinertz, R., and **Müller, M.** (2016). Analyzing measure annotations for Western classical music recordings, *Proceedings of the International Society for Music Information Retrieval Conference*, pages 517—523.