

DESIGN AND IMPLEMENTATION OF AUTOMATIC EVALUATION OF RECORDER PERFORMANCE IN IMUTUS

Erwin Schoonderwaldt Anders Askenfelt Kjetil Falkenberg Hansen

Royal Institute of Technology (KTH)
Dept. of Speech, Music and Hearing
Lindstedsvägen 24, SE-100 44, Stockholm Sweden

ABSTRACT

This paper describes a novel approach towards automatic evaluation of recorder performance. The processes from finding errors to the formulation of feedback are based on analyses of student performances and experience of recorder teachers. The developed algorithms were implemented in IMUTUS, a prototype practising environment for the recorder [3].

1. INTRODUCTION

The last decade showed a growing popularity of software for practicing on musical instruments. Most programs are developed for keyboard and depend on MIDI, but there are also programs available for acoustical instruments, such as the guitar, the flute or the clarinet [1, 2]. However, the quality and the degree of interactivity of these programs still leave room for improvement, especially concerning music performance evaluation. Some programs require the student to synchronize with a metronome or an accompaniment. This strongly limits the freedom of practicing, and presupposes that the student already masters the piece to a certain degree. Also the usefulness of the feedback is in general rather low. The feedback is mostly limited to pitch and note duration, and remains at an abstract level. This makes these programs less suitable for children.

The main goal of IMUTUS [3] was to develop an integrated environment for practicing the recorder for children from 9 to 14 years. Automatic performance evaluation plays a central role in the system, both for enhancing the pedagogical value and the interactivity. The user requirements formed an important basis of the design of the performance evaluation module. In particular, studies were conducted of student performances and teacher evaluation strategies.

In this paper the discussion will be focused on the performance evaluation module in IMUTUS, relying on output of the music recognition module. Performance transcription falls outside the scope of this paper.

2. PERFORMANCE EVALUATION STRATEGY

2.1. Basic performance skills

The performance evaluation module in IMUTUS is based on a model of novice student performance skills, covering different aspects of playing the recorder. Errors or mistakes in a student performance can be classified according to a set of basic performance skill

| No | Basic performances skill category | Average ranking | IC / MP* |
|----|-----------------------------------|-----------------|----------|
| 1 | Airflow | 1.7 | IC |
| 2 | Fingering | 1.7 | IC |
| 3 | Rhythmic performance | 2.0 | MP |
| 4 | Attack | 2.0 | IC |
| 5 | Melodic performance | 4.2 | MP |
| 6 | Tempo | 5.0 | MP |
| 7 | Intonation | 5.3 | IC |
| 8 | Phrasing | 6.0 | MP |
| 9 | Articulation | 7.5 | MP |

* IC: instrument control, MP: musical performance

Table 1: Basic performance skill categories, average ranking, and performance aspects. The ranking refers to the relative importance of the errors/mistakes during the first four terms of playing the recorder. Aspects of instrument control as well as musical performance are represented. Instrument control is considered most important to develop in the early stages of learning.

categories. In this approach, performance errors are closely related to performance skills.

The basic performance skill categories and their relative importance were determined as part of the user requirements. This was done with questionnaires and interviews with recorder teachers. In total 40 music teachers from France, Italy and Sweden responded to the questionnaires. Additional interviews were held with five Swedish recorder teachers.

The results are summarized in Table 1. A set of nine basic performance skill categories was considered to capture the essential characteristics of the students' performances, as judged by recorder teachers. These skills include both aspects of instrument control (IC), which are specific for the recorder and more general aspects of music performance (MP). The (average) ranking shows the relative importance of the skill categories. This reflects also the students' development. For beginning students, who do not master the instrument yet, the main focus lies on aspects of instrument control, while more advanced students pay more attention to aspects of musical performance.

2.2. Typical student performance errors

Specific information of typical performance errors was obtained from an empirical study. For this purpose recordings were made of recorder students at varying level of progress. Five Swedish recorder teachers were

asked to make a structured evaluation of a representative selection of eight of these student performances. They also received the printed scores. The structured analysis included the most important errors, the feedback they would give to the student, and an overall grading of the performance on a scale from 1 to 5.

The performance errors, reported by the teachers were classified according to the previously identified basic performance skill categories. This resulted in a detailed overview of typical student performance errors, and provided the basis for the development of algorithms for automatic performance evaluation.

3. IMPLEMENTATION OF AUTOMATIC PERFORMANCE EVALUATION

3.1. Performance evaluation module

The general structure of the performance evaluation module (PEM) is shown in Figure 1, illustrating the different processes from low-level error detection to generation of performance feedback and grading.

The input – generated by earlier processes of the IMUTUS system – consists of a matched score-performance pair, acoustical features of the performance (a.o. fundamental frequency, sound level, spectral information), structural information from the score (e.g. time signature, tempo) and score annotations. The matched score-performance pair consists of two MTX files,¹ representing score and performance events respectively, and a one-to-one mapping between them. The score annotations are of two kinds: graphical annotations, visible in the score (e.g. staccato dots, slurs, breathing marks) and teacher annotations (invisible), indicating potential difficulties in the student performance. This feature allows the teacher to influence the performance evaluation process by authoring the content (see section 3.4). Graphical annotations are extracted automatically from the XML score representation.²

The output of PEM consists of a complete list of detected errors, as well as a selection of the three top-prioritized errors and an overall grading of the performance. This information is used for providing feedback to the student.

3.2. Performance error detection and processing

Performance errors are identified using a rule-based approach. This is done in two stages: low-level error detection and high-level processing of error information. In the first stage the performance is scanned for performance errors, using criteria based on deviations from the score and acoustical features of the performance (*symptoms*). In the second stage, a high-

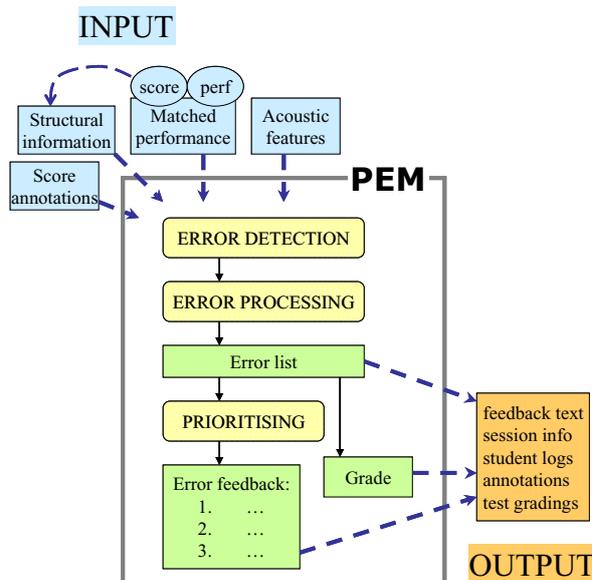


Figure 1: General design of the performance evaluation module.

level explanation of the detected errors is sought by combining low-level errors or looking at the specific context in which they occurred. The processing sheds light on the possible cause of the errors, yielding important information for feedback. Furthermore, a selection is made of errors, which are eligible for feedback. Irrelevant errors are excluded.

Performance errors can have local or a global validity. For example, errors, which are related to pitch or rhythm, are associated with one or a few notes. Tempo and tuning are typical examples of global errors. Some kinds of errors, such as the quality of the attack, are detected locally, but it makes more sense to formulate feedback in global terms. In the last example feedback could be provided if the number of poor attacks surpasses a certain acceptance level.

Also positive comments are generated during the processing, acknowledging that the student performed something well. This can be done when a certain type of error does not occur, in particular when pointed out by teacher annotations as a potential difficulty. Positive feedback is considered to stimulate the motivation of the student.

In the following sections the most important algorithms for identifying performance errors are explained, describing detection criteria and different processing steps in some detail.

3.2.1. Melodic errors

Melodic errors are simply revealed by mismatches in the score-performance pair. A distinction is made between insertions and deletions, where deletions get a higher priority. Insertions are often the result of self correction by the student, and therefore of lower priority. Furthermore, possible causes for pitch errors are

¹ A text-based music representation, compatible with MIDI [4]. See also POCO Web on: <http://www.nici.ru.nl/mmm/>

² Music XML was adopted as score representation standard in IMUTUS [5]

considered, for example the presence of accidentals in the score.

3.2.2. Tempo, timing and rhythm

Detection of performance errors in the time domain is less straight forward. First, the timing of individual notes is separated from the local tempo trend. This is done by splitting the inter-onset intervals (IOIs) of the performed notes into units of equal *nominal* duration. Then a median filter is applied using a centred window with a size of about two bars.

Detection of timing errors is based on tolerance limits for both relative and absolute deviations of IOI, taking into account the local estimate of tempo. High-level information on timing errors is obtained by considering the specific context. For example, long notes, which were played too short, are classified as duration errors. Notes followed by a rest are treated separately.¹ In those cases, the associated feedback refers rather to counting of the rest.

Tempo is regarded as a global performance aspect. The tempo can be too fast or too slow compared to the nominal tempo, but also the tempo stability is evaluated. If the amount of melodic or timing errors exceeds a certain limit, the student is recommended to practise in a slower tempo. This is an example of an indirect tempo error.

Rhythm is regarded as a group property of typically 2-4 notes. Rhythm evaluation is done only for groups of notes which are indicated by teacher annotations as difficult rhythmical groups. In contrast to timing, rhythm evaluation is based on unfiltered IOIs. Proportional IOIs are obtained by dividing the note IOIs by the group duration. For evaluation, the Euclidian distance between the nominal and the performed rhythm is calculated in an n-dimensional space, with proportional note IOIs as orthogonal axes. The default value of the tolerance limit parameter is derived from studies of categorical rhythm perception by Desain and Honing [6]. Rhythmical errors get higher priority in case of overlap with timing errors, because they are considered to be more musically relevant.

3.2.3. Articulation, breathing and hesitations

Articulation, breathing and hesitations are all related to pauses between adjacent notes: the time interval between the offset of a note and the onset of the next one. For evaluation of articulation *relative* pause duration is used, which is obtained by dividing pause duration by IOI.

In Figure 2 the different articulation regions are shown, varying from staccato to legato. The boundaries are based on perceptual tests, using synthesized recorder-like sounds.

It is important for beginning recorder students to learn to sustain the notes. The norm for “neutral”

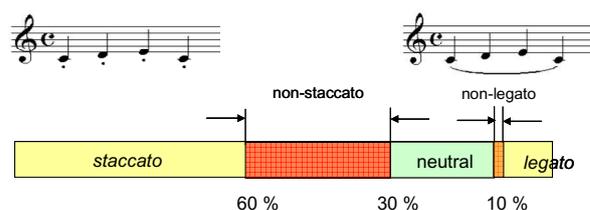


Figure 2: Articulation categories as function of relative pause duration between notes. Between 30% and 60%, there is a grey-zone between legato and staccato articulation.

articulation is therefore assumed to be close to legato. In case the average value of relative pause duration exceeds 30%, an articulation error is added to the error list. Particular evaluation of staccato or legato articulation is done when articulation marks (dots, or slurs) are present in the score.

Breathings and hesitations in the performance have in common that they exhibit a long pause between the notes, typically longer than 300 milliseconds. Hesitations can be distinguished from breathings by the co-occurrence of a prolonged IOI, flagged for as a timing error. Detected breathings are compared with the positions of the breathing marks in the score. Feedback can either be positive, in case the detected breathings coincide with the breathing marks, or correcting when the student takes a breath at the wrong positions. Breathing is strongly related to musical phrasing, and is therefore considered as an important aspect of musical performance.

Hesitations are strong indicators of specific difficulties in the performance. Combination with contextual information, such as other already identified errors or known difficulties, often yields a possible explanation of their occurrence. Some examples are the occurrence of a wrong note, which confused the student (error of another type), a difficult melodic passage, a newly introduced note or a difficult fingering transition (teacher annotations). The test cases showed that hesitations are especially useful as indicators of fingering transition problems. The error processing therefore includes a standard list of difficult fingering transitions, so that evaluation can be done even in the absence of teacher annotations. Hesitations, which remain unexplained after processing are ignored.

3.2.4. Attacks, airflow and intonation

Attacks, airflow and intonation are regarded as aspects of instrument control, and are therefore more instrument-specific, and to a lesser degree related to the score. Both airflow and attacks are considered by teachers as important basic skills (see Table 1), which need to be developed at an early stage of learning.

The attacks in recorder playing should be controlled by the tongue, in such a way that the beginning of a note is marked by a ‘d’ or a ‘t’. Good attacks are characterized by a clear start of the notes and immediate

¹ Rests are included in the IOI of the preceding note, as it is difficult to measure the onset of a rest.

tone stability. Common types of poor attacks in student performance are *glottal* attacks (controlled with the glottis) and *blown* attacks (without the tongue). The detection of poor attacks is based mainly on stability criteria of acoustical features (fundamental frequency sound level, level of the individual partials). Blown attacks can be recognized by an initial pitch rise, resulting from the slow built-up of the airflow. Glottal attacks have less distinctive acoustical features, and are therefore more difficult to detect. Glottal attacks are less controlled and tend to be too hard. However, they cannot be distinguished from too hard attacks produced with the tongue. Too hard attacks can be identified by initial overblowing, revealed by a bump in the level of the second partial or the overall sound level in the beginning of the tone.

For the feedback, attack quality is considered as a global performance property, rather than a property of the individual notes. If the relative number of notes with poor attacks exceeds a certain threshold, attack quality is added to the error list as an item for feedback.

Airflow, intonation and timbre are strongly interrelated in recorder playing. By blowing harder the pitch is raised and the spectrum extends to higher frequencies. For this reason it is difficult to make a distinction between them. On the other hand, this redundancy makes it possible to use for example fundamental frequency as an indicator for airflow. For the production of a stable sustained note, it is important that the airflow is kept constant. An unstable airflow is revealed by fluctuations in the fundamental frequency. Just as attack quality, unstable airflow is treated as a global performance property for feedback.

Blowing too hard on the recorder could lead to overblowing, especially in the lower register. Overblowing is generally characterized by a dominating second partial, which could easily lead to octave errors in pitch recognition. Therefore, in case of octave errors, the feedback relates to airflow, rather than melodic performance.

3.3. Error prioritization, feedback and grading

From the list of performance errors, produced by error detection and processing, a selection of three errors is made to be reported to the student as feedback. This is done to limit the amount of information reported to the student, so that she can focus on the most important. The selection criteria are based on a balance of negative and positive feedback, preference of high-priority errors (for example a missed note) and the order of importance of the performance skills, associated with the errors (see Table 1).

The selected errors are associated with a database of standard feedback messages, which are finally displayed in the score viewer. The feedback messages are subdivided in two levels, for providing hints or more explicit help texts.

Also an overall grading of the performance is calculated by a weighted sum of errors and positive

comments. This gives the student a quick impression of the quality of the performance. The grading is displayed to the student in the form of 1 to 3 blinking stars.

3.4. Guidance of performance evaluation by music teachers

As pointed out earlier, teachers writing the content for IMUTUS have the opportunity to influence the evaluation by PEM. The teacher can add *teacher annotations* marking difficult spots in the score where an error of a certain type is likely to occur. These annotations may concern accidentals, rhythmic patterns, difficult fingering transitions, etc. By means of score annotations, knowledge of the teacher is made explicitly available to PEM. A nice side effect is that it enables PEM to provide well-founded positive feedback when the student has played a difficult passage correctly.

4. CONCLUSIONS

A system for automatic evaluation of recorder performance was developed and implemented in a prototype practice system. The performance errors are interpreted on a high-level, increasing the pedagogical value of the evaluation, and making it suitable for children. Preliminary validation results of the prototype show that students find the system enjoyable and easy to use, and that it has a positive influence on the efficacy of practising [3].

This work was supported by the European Community under the Information Society Technology (IST) RTD programme, contract IST-2001-32270.

5. REFERENCES

- [1] Musicalis (software and online courses for music tuition): <http://www.musicalis.fr/>
- [2] Smart music (practice system for woodwind, brass, string and vocal musicians): <http://www.smartmusic.com>
- [3] Raptis, S. et al. "IMUTUS – an effective practicing environment for music tuition", *Proceedings of the International Computer Music Conference*, Barcelona, 2005
- [4] Honing, H. "POCO: an environment for analyzing, modifying, and generating expression in music", *Proceedings of the International Computer Music Conference*, pp. 364-368, San Francisco: Computer Music Association, 1990.
- [5] Fober, D., Letz, S., Orlarey, Y., Askenfelt, A., Hansen, K.F. and Schoonderwaldt, E. "Imutus, an interactive music tuition system", *Proceedings of Sound and Music Computing*, Paris, France, 2004.
- [6] Desain, P. and Honing, H. "The formation of rhythmic categories and metric priming" *Perception* 32(3), pp. 341-365, 2003.