

eMusicStand: an Intelligent Music Stand for Students and Professional Soloists, Ensemble and Orchestra Players

April 24, 2011

1 Problem Definition

There are many different positive aspects in an electronic music stand. An orchestra using such devices would be able to share music sheets automatically among the players, who may add their own notes to the parts, highlight particular musical expressions required by the director, or just ideas on how to interpret a certain piece. Moreover, the device would be able to store thousands of sheets music, bringing the benefits of eReaders and eBooks in the world of music.

Adding Artificial Intelligence techniques to such device would highly improve its functionalities. Using voice recognition it would be possible to accept voice commands; for example, it would be possible to automatically move to the measure indicated by the director - or by the player himself while practicing privately - and with music recognition it would be possible to turn page automatically when needed - very important feature for pianist, for example - and with an accurate recognition it would be possible to turn a simple rehearsal in a more interactive and richer experience, where the players can dynamically see where they currently are in the piece, and see where they made more mistakes and they have to practice more, having access to a variety of statistics such as note accuracy, ability to keep time, and so on. The director could use global statistics to better understand what parts need more time to work on, and may have suggestions for single players, or groups of players.

We propose a project where eMusicStand will be produced and tested. It will include touch screen and wireless communication with other eMusicStands, and will include algorithms of voice and music recognition to permit interaction, as well as other functionalities such as tuner and metronome, using a directional microphone and ambient sound elimination. The final product will be suited for all kind of players,

from students who will be able to practice and keep track of their improvements, to professional players for focusing more on particular aspects, having everything they need in a single product.

2 Related Work

2.1 Speech recognition

Speech recognition is widely studied in the academic world, and boasts a fairly good amount of approaches. One of these methods, consisting in using Hidden Markov Models, is deeply presented in [4] and [8]. Voice is recorded and features are extracted, resulting in a set of vectors, each of them representing a particular phoneme (or set of phonemes). This set of vectors is given as observation to the HMM, which computes the probability that a particular sequence of phonemes represents a spoken word (the “hidden” variables). N-Grams are usually introduced to bias the probability of particular words given the sentence recognized so far, in order to avoid generation of combinations without any sense. Sometimes a so-called “confusion network” is returned, representing all the possible sentences that result from the given observation. A very similar approach is described in [7], where a general Bayesian Model is used instead of a HMM. The probability of a certain phoneme given the evidence is learned. The HMM approach works well, and is able to perform continuous speech recognition.

2.2 Music Recognition

Music recognition, meant not as tune recognition, but as recognition of notes and articulations, such piano vs forte, staccato vs legato, as well as various ornaments and similars, is not subject of considerable research. [2] presents an approach of note and chord recognition using wavelet transform, which performs very well, around 96% on single tune and 90% on chords, with an execution time of around 30ms

on average. They do not try to recognize other attributes, but it is a good starting point for further research. Another approach, expressed in [6], uses Adaptive Template Matching to figure out the instrument and the pitch of a given note. They match the input with a set of templates, which are waveforms of notes previously played and tagged, and then apply Music Stream Networks, a bayesian network representing the stream of a melody, for disambiguation. With this method the authors were able to reach an accuracy of 88.5% in classifying an ensemble of three players. None of these research studies covers recognition of articulation and music expressions.

2.3 Optical Music Recognition

Another required technology is Optical Music Recognition, which is about converting a scanned sheet of music in an electronic format. [1] presents the common challenges in OMR and then presents a general framework, which consists in staff line identification and elimination, object location, musical feature classification (grouping musical objects together) and musical semantic (interpreting the spatial relationship between objects to extract music semantic.) A similar approach is presented in [3], which uses the K-NEAREST-NEIGHBOR algorithm to classify musical objects. The accuracy ratio in the two papers, and in the papers presented by them, reaches 98.6%. A very complete analysis of various techniques for OMR is in [5], which covers staff line elimination, object classification and many different solutions to common problems such as staff line localization and rotation.

3 Proposed Approach

3.1 Speech Recognition

The device does not require a perfect speech recognition algorithm, able to recognize complex dialogues and understand them properly; it has to be able to understand simple commands, such as “Turn page” or “Let’s start over”. HMM will be used for speech recognition, which is the current state of the art for speech recognition algorithms. In order to improve the ability to recognize the commands, we plan to use filtered N-Grams to reduce the number of sentences that can be recognized, and increasing the accuracy. This will not limit the system functionalities, since some words are very unlikely to be used to give commands and can therefore be removed from the used grammar.

3.2 Music Recognition

Probably the most important feature of the device. Since the input will be filtered using ambient microphones and noise reduction algorithms, we can focus on single tune inputs. An approach as in [2] fits perfectly such assumption, and will permit to correctly classify the pitch of the music in input. For what concerns musical attributes classification, such as expressions, we will test a new approach using machine learning classifiers such as Decision Trees, Support Vector Machines and others, extracting features from the input intensity, frequency, and other features that may result to be interesting in categorizing the input correctly. We believe that analyzing the waveform and the frequency space, and partially biasing the result using an electronic version of the music being played, we will be able to correctly classify the input sound, and evaluate the player on various metrics, because the way every note is played will give us information about the original intention in form of particularities in the input wave (for example, a staccato will consist in soft accents at the beginning of every note.)

3.3 Optical Music Recognition

The device will be able to read a new music sheet and convert it in an electronic representation that will be then used in visualization and music recognition. We will implement methods of staff line identification and elimination as in [5], object classification using segmentation and horizontal/vertical projection as in [1], as well as new approaches such as object filtering, SVMs and music semantic. The device will let the user to correct the output of the OMR process, in order to reach the perfection needed to play in an ensemble. It has already been proved that this approach leads to good accuracy, and we believe that using object filtering, SVMs and especially music semantic we will be able to improve the overall quality of the result.

4 Proposed Evaluation

In a first phase the system will be evaluated on a single player, in order to test the correctness of each algorithm. We will start testing the music recognition algorithm on single tunes, measuring the accuracy as number of correctly recognized notes over the total. We will also test the OMR in a similar way, while the speech recognition algorithm will be tested on simple commands. We want also to evaluate the entire system on usability, measuring the level of frustration of

the player while using the new music stand.

In a second phase, more devices will be produced and they will be tested with typical ensembles composed by five to eight players, with different instruments. The device will be tested in an environment with multiple sound origins, and a similar evaluation approach as in phase one will be applied. The first stage of interaction with other devices will be then tested, for sheet sharing as well as command sharing, using a master-server communication approach. More complex tunes will be tested as well, to check the music recognition algorithm.

In the last phase, the device will be tested on an orchestra of medium size, from 40 to 60 players. The evaluation approach will be similar to the previous two, this time with more noise from the environment. The evaluation will be done on pieces of different difficulties, and the users will perform different actions to test completely the device.

After every phase, the previous ones will be evaluated again, in order to check retrocompatibility. Modifications to improve the accuracy while playing in an ensemble do not have to affect the accuracy while playing as a soloist.

We believe that the system will perform very well on a single player, highly improving the effects of training. With an ensemble or orchestra the system will not perform as well as with a soloist, probably, but will still achieve a good accuracy, and the benefits by using this system will be prominent.

5 Research Plan

The project will go through four phases; during the first four months, the basic implementation will be developed, which will consist in the basic speech recognition for commands, music recognition for soloist and OMR, as well as all the other basic technologies such as touch screen interaction. The algorithms will not be complete, but they will be functional, in order to be able to start testing the device.

In the second phase, this first implementation will be tested, evaluated and improved for three months, using a spiral approach between updates and tests with soloists. All the aspects of the device will be updated and tested, in order to achieve good performance with a single player.

In the third phase, lasting three months, the device will be expanded adding network capabilities, and will be used for the first time in an ensemble. The algorithms will be tuned to achieve a good performance level with a group of players. The algorithms will be

almost complete and functional.

In the last two months, the device will be introduced to an orchestra composed by young players, and tested during their weekly rehearsals. The last changes will be applied to the algorithms, to achieve the top level implementation. The first month of testing will be supervised, while during the second month the players will be left to play without any technical supervision. Data from the devices will be still collected weekly, and a survey will be submitted at the end of the test cycle.

References

- [1] BAINBRIDGE, D., AND BELL, T. The challenge of optical music recognition. *Computers and the Humanities* 35 (2001), 95–121. 10.1023/A:1002485918032.
- [2] CAO, Z., GUAN, S., AND WANG, Z. A real-time algorithm for music recognition based on wavelet transform. In *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on (0-0 2006)*, vol. 2, pp. 9926–9929.
- [3] CHOUDHURY, G., DiLAURO, T., DROETTBOOM, M., FUJINAGA, I., HARRINGTON, B., AND MACMILLAN, K. Optical music recognition system within a large-scale digitization project. In *Proceedings ISMIR 00* (2000), Citeseer.
- [4] GALES, M., AND YOUNG, S. The application of hidden Markov models in speech recognition. *Foundations and Trends in Signal Processing* 1, 3 (2008), 195–304.
- [5] JOHANSEN, L. *Optical Music Recognition*. PhD thesis.
- [6] KASHINO, K., AND MURASE, H. Music recognition using note transition context. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on* (may 1998), vol. 6, pp. 3593–3596 vol.6.
- [7] NORRIS, D., AND MCQUEEN, J. M. Shortlist b: A bayesian model of continuous speech recognition. *Psychological Review* 115, 2 (2008), 357–395.
- [8] RABINER, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 2 (1989), 257–286.