Hochschule
Augsburg University of
Applied Sciences

Project MIDILENS

Report

by
Ruben Schlagowski (Matr.-Nr. 954174)
Martin Stadlmaier (Matr.-Nr. 954084)

for the course
Interaction Engineering (WS 2017/18)
Course Instructor: Prof. Dr. Michael Kipp

University of Applied Research Augsburg

Delivery: 05.02.2018

1) Abstract

This report describes the motivation, concept, implementation and evaluation of “MidiLens”, a system that was developed for the course “Interaction Engineering” at the University of Applied Research Augsburg. MidiLens is a system which teaches piano songs by using Augmented Reality on a head mounted display (Microsoft HoloLens).

This is achieved by enriching a keyboard with hologram, guiding the user through songs and giving him or her feedback on how well he or she played. An intuitive system is used for musical notation so that the user will be able to learn songs without any need for musical knowledge.

As a part of this project, a user study was conducted to compare two distinct application designs for displaying the musical notation and user feedback.

2) Motivation

Music playing is a globally understood way to communicate emotions, but can actually require a huge amount of practicing and expertise. Therefore, many people hesitate to learn an instrument or to play music with other (maybe more experienced) musicians together. We would like to implement a system that enables anyone to play music. This includes improvisation, composition and song learning.

The approach to develop a system that simplifies musical training by showing the user when to play certain notes on a keyboard is not new, as it has been done in various research projects before. One example is the project *Game of Tones* by Linsey Raymaekers, Jo Vermeulen, Kris Luyten and Karin

Coninx which used a beamer to project the musical notation onto a table surface.

A similar style of notation, which is characterized by displaying moving objects representing notes, is used in a variety of applications and instruments such as online piano training videos or computer games such as Guitar Hero.

This project's goals are to evaluate the potential of state-of-the-art augmented reality technology for this approach, such as the Microsoft HoloLens released in late 2016. We also aim to compare different design options for displaying musical notation in augmented reality and therefore gain knowledge on how to implement musical teaching systems using augmented reality which are hopefully more effective, fun and beginner-friendly.

3) Concept

For musical input one of the arguably most precise and well-proven musical interface is used: the classical musical keyboard as it is used in pianos, or more recently midi-controllers such as master-keyboards.

As there are a large amount of compositions in the form of Midi-Files available on the internet, a PC together with a keyboard and a mouse is used for song selection and playback. By connecting a Master-keyboard to the PC and using a sound card for auditory output, it is possible to assign any sound to the keys on the master-keyboard by using Virtual Studio Technology Instruments (VSTis) as standalone software or as Plugins in Digital Audio Work Stations (DAWs).

For the visual output of musical notation and key feedback, the Optical See-Through Displays (OSTs) of Microsoft HoloLens are used.

The song/composition will be displayed by showing moving panels on a two-dimensional GUI-canvas positioned either in a vertical position to the keyboard (mode A, see picture on the left below) or right on top of the keys in a horizontal position (mode B, right picture below). As soon as a panel hits the upper edge of the keys, the user is supposed to press the key until it disappears from the canvas.

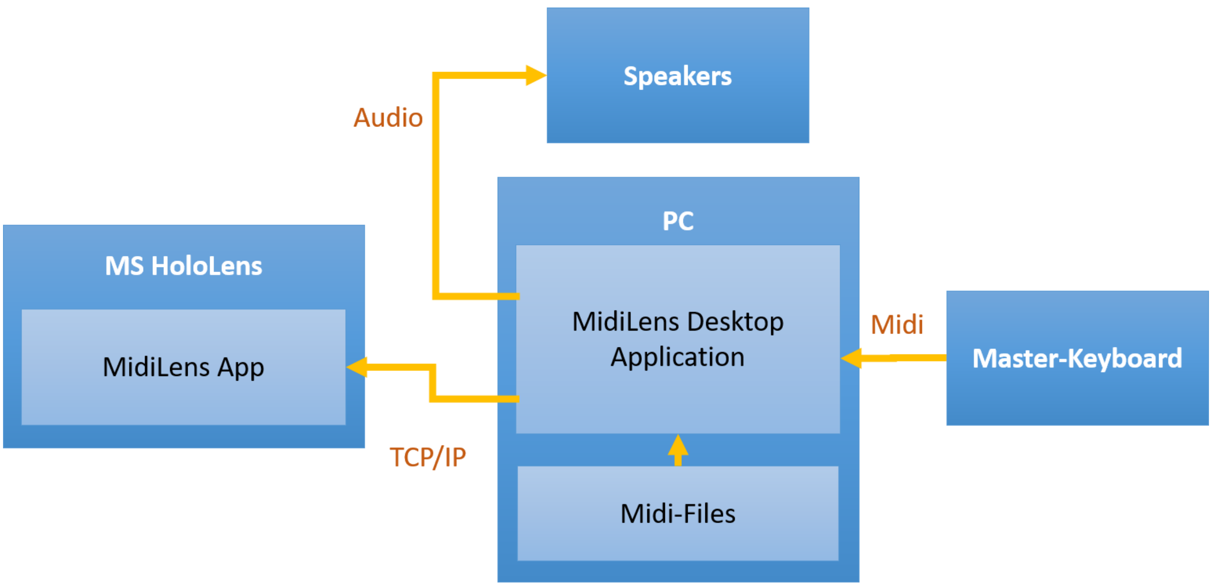
For mode A, visual feedback is given to the user by highlighting the played keys in a colour indicating how precise he played the note (green: good, yellow: medium, red: bad).
For mode B, this feedback is displayed in another position, which is at the upper edge of the keys, as it would overlap with the moving panels indicating notes.

For direct interaction with the HoloLens application the input modalities are speech, gaze and gesture. Gaze and Gesture are mandatory for starting and closing the application while voice commands can be used for optional features such as the playback of hard-coded test songs for testing purposes.

4) Implementation

Two software applications were developed for this project: *MidiLens Desktop* Application and *MidiLens App* for HoloLens (see figure below).
The used communication protocols used are

- Midi for the interface between the Midi-Sources (Midi-Files and Master-keyboard) and Speakers for auditory output and
- TCP/IP for communication between MS HoloLens and the MidiLens Desktop Application.



The MidiLens Desktop Application which was written in C# and mostly serves as an interface between the technical components. Its main functions are:

- Providing a user interface to select midi-files for song playback.
- Generation of auditory output from midi-signals generated by the master-keyboard.
- Translation and transfer of midi-signals from both the master-keyboard and the selected midi-file to Microsoft HoloLens via TCP/IP.

For the midi-interface, *Sanford* midi libraries are used and for communication over TCP/IP *.NET* sockets were used within the Visual Studio IDE.

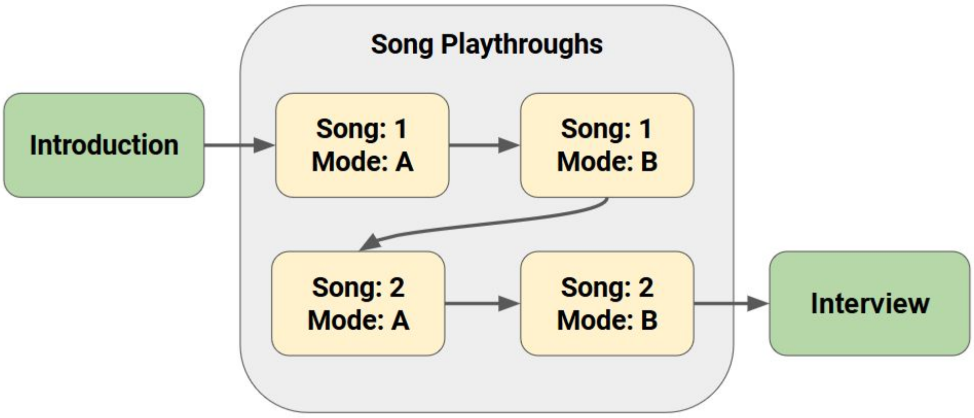
The MidiLens Application for HoloLens was developed in *Unity 3D* together with Prefabs by *Vuforia* for Optical Tracking, *Windows UWP* libraries for communication over TCP/IP and Prefabs from *HoloToolkit* by Microsoft for basic holographic functions and input modalities. The most important Scripts written in C# are listed below.

Script Name	**GameObject**	**Function**
MainApplication	Main	Main Application functions such as startup, calibration, top-level command transfer, mode switches
SongPlayer	Main	playback of hard-coded test songs
SaveValues	Main	Saving timing precision values for score calculation between songs
AnchorBehaviour	Anchor	Function to save optical tracking data as a reference point in 3D space
GuiFunctions	All Canvases in Scene	Basic GUI functions such as hide/show
KeyFeedback	KeysCanvas/KeysCanvas2	Shows highlighted Keys with colour according to precision. Also stores timing values in an array.
PlayNotes	NoteSheetCanvas/ NoteSheetCanvas2	Creates, modifies and destroys panels according to the notes that should be played
UwpTcpSocketFunctions	UwpTcpSocket2	Top-Level communication functions for TCP/IP. Uses functions of a variety of other TCP scripts/objects in Assets/Scripts.
KeywordManager	Managers	Voice command programming (use the inspector). (Author: Microsoft)

5) User Study

In order to test the application, its components and the interaction technique, we conducted an user study. It consisted of the interaction with the MidiLens system itself and a post-test interview with a questionnaire. The test took place on 31. January 2018 at the Fraunhofer Research Institution for Casting, Composite and Processing Technology (IGCV) in Augsburg. In total, we acquired data from nine test subjects which were between 20 and 50 years old.

Approach



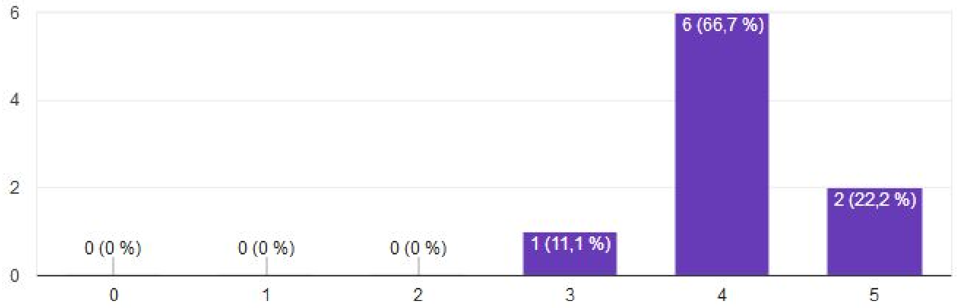
After a short introduction to the MidiLens system, a test subject had to play two songs. Therefore, we decided to take well-known and easy-to-learn songs like *"Alle meine Entchen"* and "*Hänschen klein*". Each of this songs had to be played twice in two different modes (A & B) as described in chapter 3.

After finishing the playthroughs, each test-subject had to do a questionnaire, which consisted of four questions with predefined answers (No 1-4) and two questions with open answers (No 5, 6) in order to collect qualitative feedback. The results are listed on the following pages.

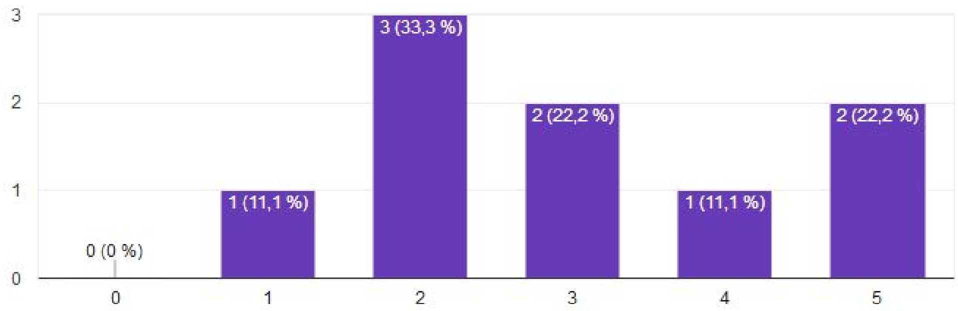
Results

Question 1: How intuitive would you rate the learning processes for both modes?
Answering options: 2 ratings on a scale from 0 (not intuitive) to 5 (very intuitive)

Mode A)



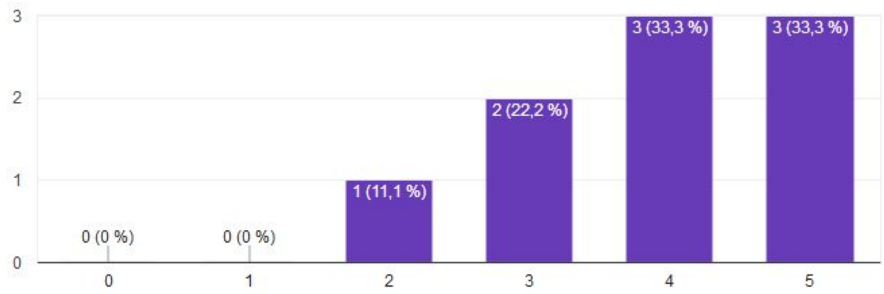
Mode B)



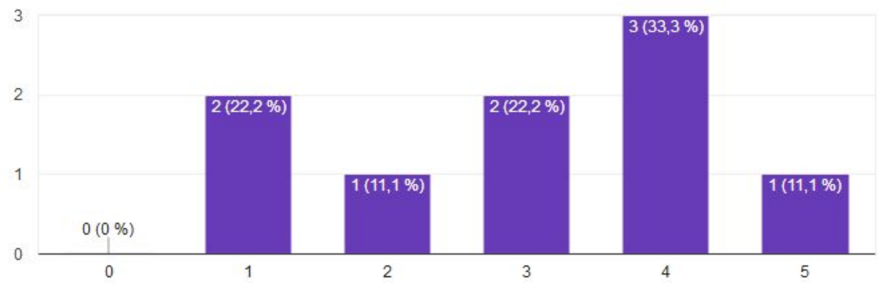
Mode	A	B
Average	4,1	2,9
Median	4	2,5
SD	0,57	1,37
VAR	0,32	1,88

Question 2: How effective would you rate the learning process for both modes?
Answering options: 2 ratings on a scale from 0 (not effective) to 5 (very effective)

Mode A)

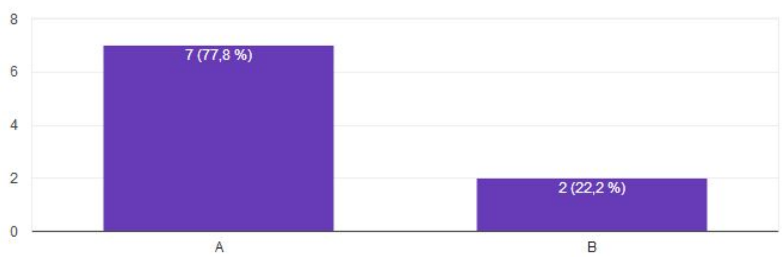


Mode B)



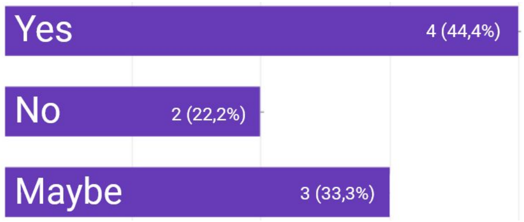
Mode	A	B
Average	3,8	2,9
Median	3	2
SD	1,03	1,37
VAR	1,07	1,88

Question 3: Which mode would you prefer to learn a new song?
Answering options: Mode A or B



With nearly 80%, mode A was chosen as favoured playing mode by the test subjects. For explanation, why two people choose mode B over mode A, see next subchapters.

Question 4: Could you imagine to use the MidiLens system for a longer time-span that may be required to learn a whole song?
Answering options: Yes, No or Unsure.



More than 40% confirmed a long-time appeal of the system, whereas only about 20% denied. One third of the test subjects was unsure whether they would use the system for its intended purpose.

Question 5: What did you find good while using the MidiLens system?
Answering options: Open answer.

In conclusion, the following answers were given:

- **fun** factor (system is very fun to use, reminds of guitar hero and equivalent systems)
- whole system (the approach to use a HoloLens in combination with Midi / a master keyboard is cool)
- **Feedback** (visual feedback helps a lot when instantly playing unknown songs on an unknown instrument)
- **learning effect** (one test subject stated that there has been a learning effect and she would remember on how to play "*Hänschen klein*")

Question 6: What do you think could be improved on the MidiLens system?
Answering options: Open answer.

In conclusion, the following answers were given:

- **technical difficulties** (tracking not precise enough, HoloLens field of view too small, no constant framerate, voice command recognition failed, runtime errors, connection bugs)
- **optical issues** (no or not enough contrast/ difference in colors for white and black keys)
- **Feedback** (more and improved feedback regarding timing, tempo, rhythm and the beat in general)
- **learning effect** (some test subjects stated, that music is played by ear and not by eyes)

6) Evaluation

Evaluation of user study results

In terms of intuitiveness, mode A was averagely rated a bit better than mode B. On the other hand, in terms of effectiveness, mode A and mode B were rated both equivalent. But when asked which mode would they prefered, about 80% of test subjects stated that mode A would be their choice.

This could be explained regarding standard deviation (SD) and variance (VAR). Considering standard deviation and variance from mode B in BOTH results from question 1 and question 2, it is striking that there are the exact same values. This can be explained regarding both graphs, which are close to mirror-inverted in relation to each other. The high standard deviation / variance in both question results for mode B means, that the test subjects did not agree on a common level about, if this mode is intuitive and / or effective.

On the contrary, standard deviation and variance of mode A are very low in terms of intuitiveness and relatively low in terms of effectiveness, which means that the test subjects did agree on a common level about the intuitiveness and / or effectiveness.

We conducted a t-test in order to clarify if the different results for questions 1 and 2 are statistically significant. According to the table below, mode A and B statistically differ significantly in terms of intuitiveness, but in terms of effectiveness, no statistically significant difference could be confirmed.

	Intuitiveness	Effectiveness
paired significance	0,034	0,136
t-value	2,426	1,575
Significant?	YES	NO

Evaluation of open user feedback

In general, feedback from the test subjects was positive. Most liked was the approach to learn to play a real instrument with the HoloLens. Also, mostly every participant liked the MidiLens application because it was "*fun*", "*cool*" and "*easy to understand*" and helped them to play a complete song on a piano.

Most feedback regarding technical issues, like small field of view or inconsistent framerate are due to technical specifications of the HoloLens. Others may / will be solved by time.

Even though the majority of test subjects thinks that songs are learned most effectively by ear, a visual approach in an early stage of learning is reasonable, as there is no need to learn reading notes first. Furthermore, in terms of having fun and gamification aspects, the system could still be improved.

The visual feedback given by the system was pleasing to a certain degree, but should be refined as well. Especially the difference between white (fulltone) and black (halfone) keys should be more clear. Furthermore, the beat (including rhythm, punctuation, accentuation and timing) should be visualized.

Technical Issues/Difficulties

Overall, a variety of technical difficulties either came up during the user study or couldn't be resolved in time due to time constraints. One of the biggest technical issues that we could not fix in time was that midi-commands from File sometimes get delayed or lost when commands for key-feedback are queued in the network. This resulted in some notes being displayed too short or too long when playing songs from a midi file. For the user study we avoided this problem by hard-coding two test-songs which could be started over voice command.

Another important problem that needs more effort to be resolved is that the calculation of delta time values between a key being played and the time when a key should be played is not calculated correctly. The stored values in the prototype right now are therefore faulty. These could later on be used to either conduct more user studies with high precision or calculate a user score for gamification purposes.

Beside the challenges mentioned above, it can be said that optical tracking by using Vuforia together with optical referenciation is borderline precise. Improved or new versions for the used tracking algorithms have been released in the meantime and should be evaluated.

7) Conclusion

In conclusion, a first functional prototype could be developed and the feasibility and practicability of the concept could be confirmed, as the feedback from the test subjects has been better than expected for the overall system.

As this first prototype still contains a variety of bugs and quick-fixes, it still requires some work to get it to alpha-stage. As soon as these technical issues are resolved, more features can be built in, such as gamification aspects (e.g. calculating and displaying user-scores of implementing difficulties).

For future implementations, the AR canvas for notation displaying should be placed vertically (as in mode A, compare pages 3-4) and also be refined, for example by improving the distinction between halftone and fulltone keys (e.g. by better colour contrast).

Furthermore, finger-positioning could be assisted with the system, e.g. by displaying numbers on the keys which are dedicated to the user's fingers. This could improve the development of playing technique and therefore benefit the user long-term, which was often doubted by test subjects.

Finally, an important feature which was initially planned could still be implemented: the general assistance of improvisation. A feature like this could motivate the user to make music together with other people instead of just sitting at home and following instructions. Besides being a unique feature, which is not that common in research, this could have the potential to connect even more people through music, which was one of the primary motivations behind this project to begin with.