From Body Movements to Music
A New Device for Movement Therapies

Christian BRÄUER-BURCHARDT*[1], Matthias HEINZE[1], Roland RAMM[1],
Robert WECHSLER[2], Ursula MÜLLER[3], Peter KÜHMSTEDT[1], Gunther NOTNI[1,4]
[1] Fraunhofer Institute IOF, Jena, Germany; [2] Palindrome Dance Company e.V., Weimar, Germany;
[3] Grenzenlos e.V., Jena, Germany; [4] Technical University Ilmenau, Germany

https://doi.org/10.15221/19.158

Abstract

A device for music generation through movement is introduced which may be a new tool for therapies especially of persons with disabilities. The system connects fast 3D capturing of acting person(s) with high-resolution 2D images and specialized sound/music modules, which translate the detected movements into sound and musical phrases. In this paper, the main challenges and solutions of the project are presented. Experiments and applications with different groups of patients are described and evaluated.

Keywords: 3D human motion scanning, music generation, motion tracking, disabilities

1. Introduction

Dancers know that music is a stimulator for body movements, and musicians may confirm that self-made music can be a highly satisfying experience. Combined, movement and music can be a stimulator for movement, joyful expression and improved quality of life. However, persons with other abilities (also “persons with disabilities”) are often unable to learn and play a musical instrument either because of cognitive or physical limitations. In order to overcome some of these restrictions, we have developed a new device which translates body movements into own music compositions. This device should help to keep and improve vitality and movement abilities in the context of physiological therapy.

In a scientific research project, partners with various backgrounds (a research institute, a dance company specialized in interactive performance, an industrial design company and a registered social-support society) worked together in order to realize a scanning device that translates body movements into own music compositions. The origin of this project was a commercially available device called “MotionComposer” [1, 2] with a number of shortcomings to be improved.

The goal of our work was to develop a system, which could improve the mobility of patients with motoric handicaps as well as generally improve quality of life for persons with other abilities. The principle of the system is to create music that is dependent on the partial movement of the acting person(s), such that a certain series of movements causes a pre-defined acoustic result. Such pre-determined settings are defined into a variety of “Musical Environments” (MEs). Some MEs are characterized by melody and rhythm, while others involve triggering sounds from nature, such as animals, which are likewise triggered and modulated through movement.

In order to achieve the goals of the project and to realize such a music-through-movement system, a series of problems must be solved. The following tasks are parts of the development:

- Capturing the acting person(s) and separation of acting persons (if more than one)
- Tracking of acting person
- Determination of the 3D position in the considered space
- Determination of relevant physical properties of acting persons such as arm stretching, jumping, kicking, head-shaking, and eye blinking
- Translation of the detected movements into music using a so-called movement alphabet

Paramount in the realization of these tasks is the requirement that all aspects of the system take place with very short latency times. This is necessary in order to have the sense of causality in the patient’s experience. In other words, the time between the physical action, and the resulting sound coming out of the loudspeakers must be low enough that the patient has a Sense of Agency (SoA) [3, 4], i.e. “I caused the sound to happen.”

The following technical principles were used to reach the goal:

- 3D-capturing of the scene using a time-of-flight (ToF) camera [3]
- Mapping the 3D data onto a 2D image of an additional 2D-camera with higher resolution
- Segmentation of the moving persons in the scene
- Image processing for detection of certain body movements

*christian.braeuer-burchardt@iof.fraunhofer.de; +49-3641-807235

In this work, we begin with a description of the task-solving process, including a brief review of the state-of-the-art. Next, the solution of the technical challenges is described. A presentation of the technical features of the developed system is then given and application examples of the work with several groups of patients with different abilities is presented. Finally, a discussion of the achieved results and an outlook to future work conclude this paper.

2. State of the art

Recognition of human bodies and estimation of location, pose, gestures, and movement dynamic (activity) in image sequences is a matter of research for many years (see e.g. [4]). Three-dimensional movement data are mainly used for computer games. Typical features are fast sampling rates and low spatial resolution. Human movements can be captured quickly using additional tracking means. A short latency is one of the crucial criteria for the approval of the users.

There are only a few products that combine movement and music with the focus on therapy of people with handicaps. Many products such as the Kinect sensor [5] are capturing movements, but data processing typically does not reach a high technical level, which is necessary for applications like therapies, nor do such devices include content - musical environments designed for therapeutic applications.

Currently, there are three products available on the market which translate movements into music for therapeutic purposes.

The system “Orgue Sensoriel” [6] uses a series of mechanical input devices, e.g. presses or rollers, generating music by using a certain software module. Certain movement patterns are guided to the user by the input devices. Persons with other abilities sometimes cannot fulfill the requirements concerning these necessary movements. The requirement of physical controllers limits the range of possible users.

The system “Soundbeam” [7] captures movements in the space using an ultrasound sensor. The acquisition of free movements is possible, but the high sensitivity of optical sensors is lacking. The captured space is very limited. The user must be very close to the sensor. At most, two persons can be tracked simultaneously. Additionally, mechanical input devices are available.

The system “MotionComposer MC2.0” [1, 2], which is the basis of our development, detects movements using 2D camera and shapes and gestures using a TOF-type sensor. It is intolerant of room-lighting conditions. The data processing is relatively slow leading to latency time higher as desired.

A big disadvantage of all three systems is the complex handling and controlling. The usage of the devices requires relatively high qualification of the user, also during therapy sessions.

3. Approach

3.1. Situation and goal

The task of the system is the translation of movements of the acting persons into creative music. The features are the same as of the preceding system “MotionComposer”, but with several improvements. Additionally, a certain target audience was in focus, namely, persons with other abilities and several groups of patients with certain diseases for example muscle weakness, blindness or other constraints of locomotor system, dementia and others. The purpose of application is to improve health condition and quality of life by stimulation of the mobility.

The main requirements to the system were the coverage of rooms of a size of approximately 5 m width, 3 m height, and 5 m depth. Single persons should be captured as well as small groups of persons. The developed system consists of several modules and components, respectively. The main components are

- 3D-scanner, consisting of a ToF camera in near infrared range (NIR)
- 2D camera in visible light range (VIS)
- Mini PC
- Tablet
- Software modules
    - Control of the hardware components
    - User interface
    - Image-data storing and processing (3D-2D data mapping, blob detection, tracking)
    - Translation module (movement alphabet element recognition)
    - Music environments
- Loudspeakers

Figure 1 shows an acting person in a room and the measured 3D data of the scene.

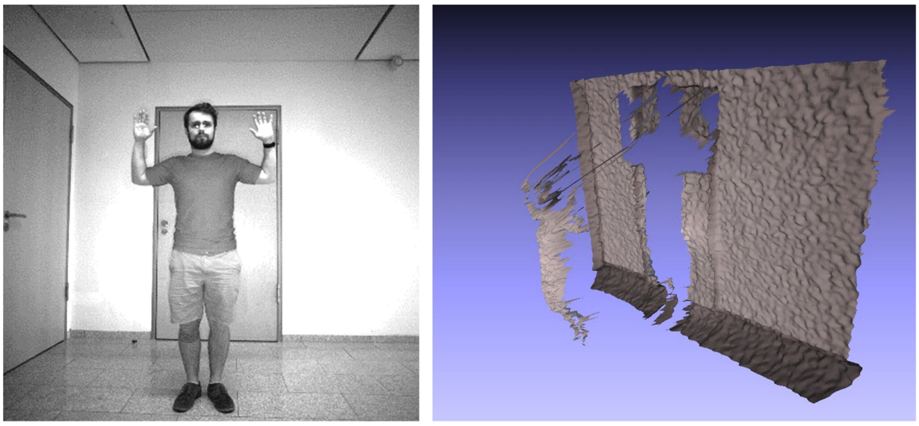


Fig. 1. Acting person (left) and 3D representation of the measured data (right)

3.2. Technical features

3.2.1. Measurement principles

The measurement principles used for detection of the moving person(s) are 3D-scanning and 2D observation of the scene. 3D scanning is necessary for the robust detection of the person and for scaling, whereas 2D camera images are used for detailed movement interpretation. The 3D scanning of the scene is realized by a commercially available ToF-camera. Initially arranged principle of structured light illumination and stereo observation was proved to be too time consuming and faint. Already realized systems with high 3D data rate cover only much smaller rooms [8] or need structured illumination in the visible range [9], which cannot be accepted because of the visual irritation to the acting persons. As alternative, the ToF-camera provides sufficient spatial resolution (approximately 10 mm) and depth measurement accuracy (maximum depth error: about 50 mm), which is acceptable for the purpose of rough detection of the position of the moving persons. Detected 3D data was then mapped onto the high resolution 2D camera image of the moving persons, and different gestures and body movements were detected using these 2D data.

3.2.2. Challenges and approaches

The main challenges to be solved in the development of the system were a robust detection of the acting persons and an exact identification of the moving body parts in a very short time, because the generated music should appear immediately after the movement. Additionally, light conditions in the observation room should be tolerant, e.g. direct sunlight or sharp contrasts should not influence the scene.

The ToF-camera used has a frame rate of 20 Hz and the 2D camera of 60 Hz. 3D data are processed obtaining the so-called blob of the person (see Fig. 2). Current 2D image with higher spatial resolution is masked by blob(s), i.e. only 2D data in the blob(s) are processed. The maximum processing time of the 2D data is less than 50 ms. Depending on the soundcard used in the embedded mini-PC, there is a time delay between the finish of data processing and the ring out of sound from the speakers. Currently. in our system this time takes 80 ms. The total latency time between the movement of the person and the corresponding music generation is hence less than 180 (50 + 50 + 80) ms.

Different illumination conditions of the room are managed by automated control of exposure times and gain of the 2D camera. In case of changing illumination conditions these settings can be changed by a reset of the application.

3.3. Data processing and music generation

From movement to sound several processing steps have to be realized. The scheme in Fig. 3 illustrates the data processing steps.

The 3D data are used for detection of the location and position (form) of the acting person. When detected, a so-called blob is generated. The 2D data are mapped onto the blob. Consecutive 2D images are used to calculate a difference image which tells something about the so-called activity in the image. Additionally, gestures and location of extremities and head are detected. This task was using a software module by our external partner FusionSystems GmbH [10].

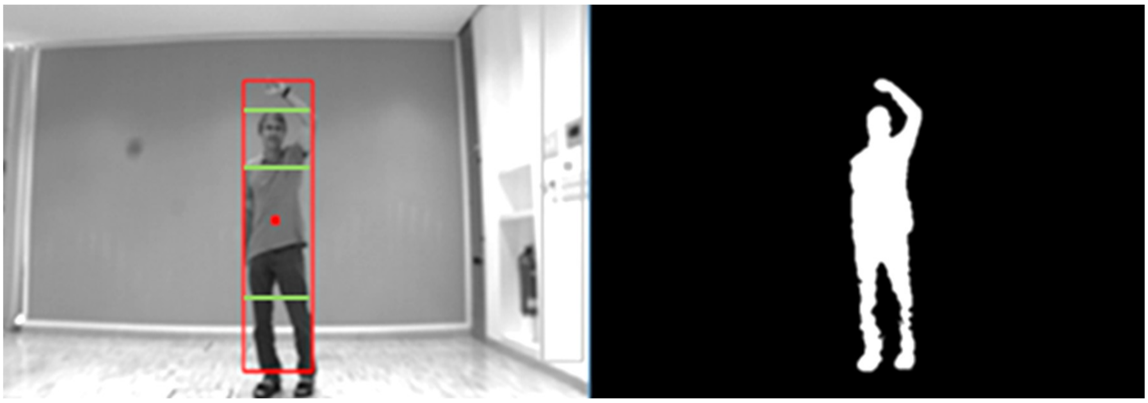


Fig. 2. Acting person with detected body areas (left) and extracted blobs (white region - right)

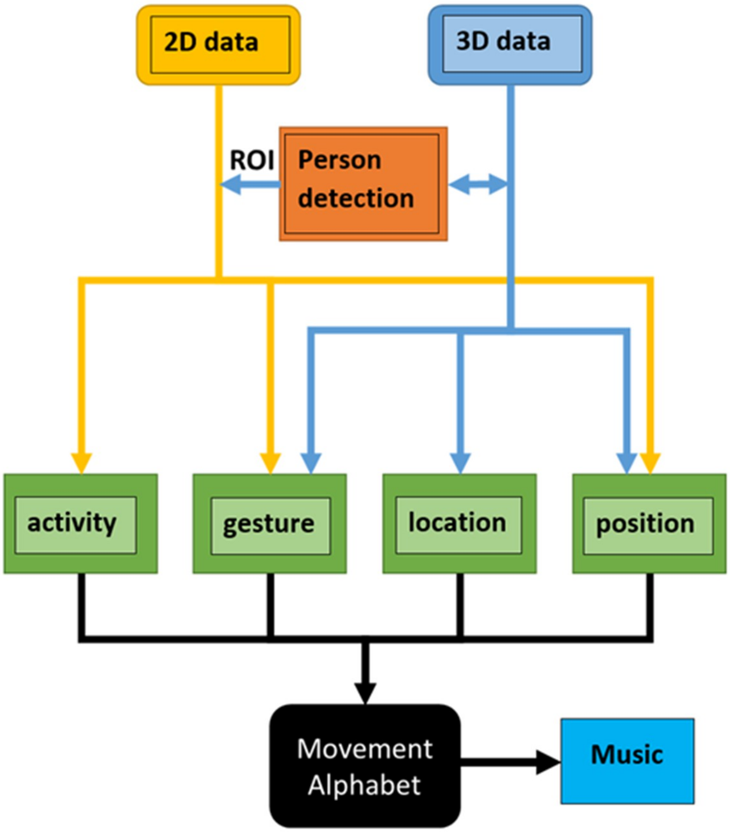


Fig. 3. Scheme of data flow at music generation

All collected data are the input for the "Movement Alphabet" (MA), which is a list of body-to-sound mappings which we deemed meaningful in the context of SoA, and in the sense of giving patients an intuitive sense of body-expression. MA consists of approximately 40 features, which for convenience, are divided into four types: location, position (form), activity and gesture. MA features have a heirarchy of parameters, thus, for example, the height of the patient's right hand is defined by the term:

playerID/shape/armHeight/right.

This term is delivered to the music software via the Open Sound Control (OSC) digital protocol, where, depending on the selected music environment (ME), certain sounds are produced. For example, the height of the hand might be used to determine the pitch of notes in a musical scale; the higher the hand, the higher the notes.

Other movements and gestures used include one or both arms moving to the side, head shaking, finger movements and even eye blinking. In addition to pitch, other sound parameters controlled include volume, rhythm, choice of musical instruments or soundbanks, density of discrete sounds, and so on.

For our system we use three different music environments: “Techno”, “Tonality”, and “Fields”.
The “Techno” environment is concerned with pop music. Starting the session, it begins with a more-or-less finished song. True to its name, the beat never stops, but the user is given ways to join in. They can pump up the base with strong torso movements, introduce filters, and play accents using quick sideward hand movements.
“Fields“ can be considered as a basic environment with the simple message: you move, you hear. Sound is created by the whole body. Small isolated sounds are possible, for example a single chirp of a bird comes from a small finger movement. Low activity may generate a singing bird whereas high activity may produce many birds, or larger birds (crows instead of finches).
“Tonality” environment provides the sounds of classical music instruments. Beginning with the piano, moving from left-to-right through the room implies playing low tones to high tones, sequentially. Discrete movements play single notes or small chords, while holding the arm in certain positions generates arpeggios.

4. System realization

4.1. Hardware setup

According to the described approach, we realized a setup of the system (see fig. 4) with the following features

- Captured area (= acting volume AV): 5 m x 3 m x 5 m
- Distance between device and AV: 2 m
- Spatial resolution in the AV: between 8 mm and 10 mm
- Size (width x length x height): 300 mm x 400 mm x 150 mm
- Weight: approx. 4 kg
- Working temperature range: 10 °C to 35 °C
- Power supply: 230 V
- Loudspeakers 2 x 20 W
- Power consumption: approx. 200 W
- Handling tool: tablet

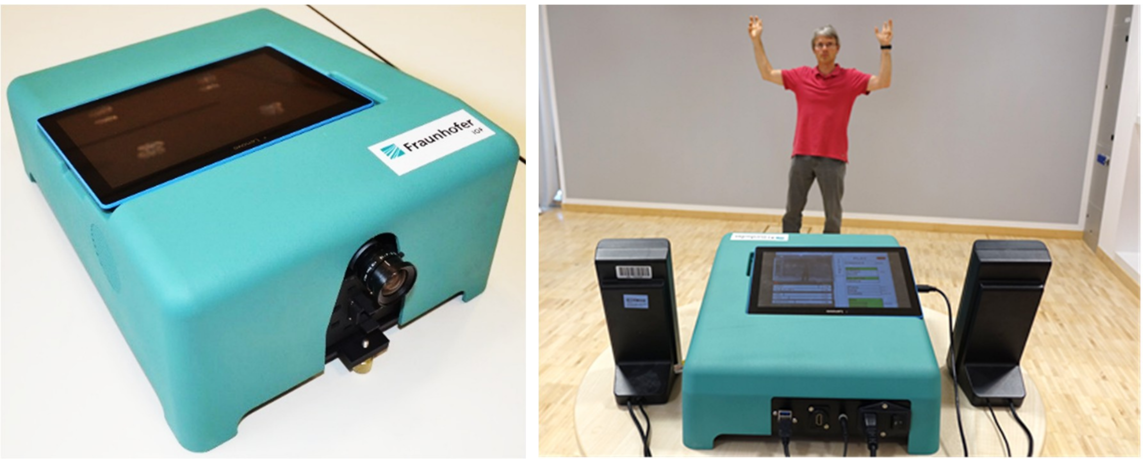
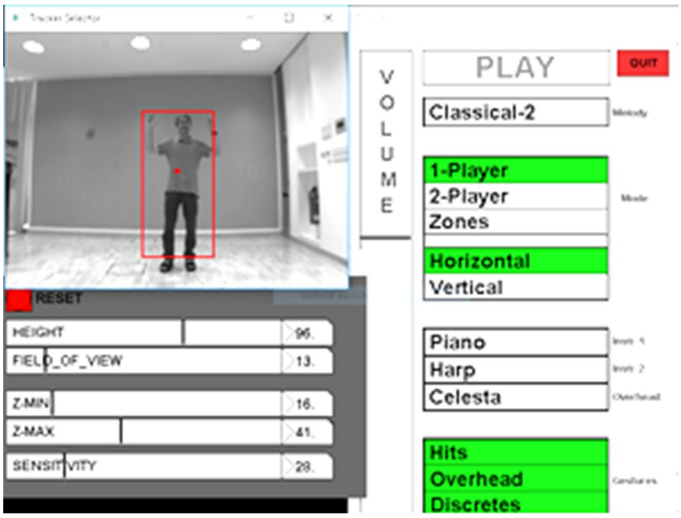


Fig. 4. The music generation system: main device (left), demonstrator in action (right)

4.2. Handling the system

In the following a brief description of the handling is given. In preparation a possibly large room should be selected without obstructions (like chairs and tables) in the tracking area. Room illumination should be set as needed. After switching-on the system, the music environment is chosen using the interface on the tablet. Initially, some basic calibration is carried out. All persons in the observed area should be detected by the system. The user may interactively select an active person by clicking. Now, all movements of that person generate defined results, for example, they may be given a certain virtual musical instrument which to play. Figure 5 shows the GUI of the different music environments.



HEIGHT 96.
FIELD_OF_VIEW 13.
Z-MIN 16.
Z-MAX 41.
SENSITIVITY 28.

Classical-2

1-Player
2-Player
Zones

Horizontal
Vertical

Piano
Harp
Celesta

Hits
Overhead
Discretes

Fig. 5. The GUI, current music environments: Tonality

Certain parameter such as volume and room size (width, depth, and height) can be adjusted. The latter is meaningful in order to avoid disturbances by movements or action outside the acting volume, which are nevertheless detected by the ToF camera. Additionally, active area can be adopted to the physical situation of the patient, i.e. “Room” is the complete area whereas “Chair” and “Bed” are more restricted regions for movement detection, designed for people in wheelchairs or in bed respectively.

5. Experiments and results

5.1. Preliminary experiments

First experiments were performed in order to evaluate the functional features of the system, i.e. check the working space, the influence of the illumination conditions, and the correct recognition of the acting persons and their movements.

A second series of experiments was performed in order to determine the influence of the environmental conditions. For this purpose, several acting scenarios were scheduled, applied, and the image data sequence was stored. Analysis of each experiment was done by subjective evaluation of the detected blobs in the image sequences. The following conditions were checked:

- Standard conditions
- Entering and exiting of the acting persons
- Actions too close and too far (outside the official acting area)
- Different clothing features (white clothes, skin, dark clothes)
- Different floor materials (carped, wood)
- Different illumination conditions (sun in room, sun on player, sun on floor, low light)
- Acting person in wheelchair (alone) and with someone pushing

Experiments were performed with one player and multiple players. Additionally, position, form factors, the level of activity, and gestures were determined and evaluated.

To begin with, a standard scenario was defined, called “basic scenario” (BS). Experiments were performed in a big well-lit room with light-colored walls. Sensors were placed at the viewing axis parallel to the walls. The floor was to be seen in the 3D data as well as the background wall but side walls were not visible. The room was bright, but without direct sunlight. Always one or two persons wearing “normal” clothes are in the Acting Volume (AV). Typically, the person is in the center of the AV and moves. Sometimes, additional movements by the arm or legs are performed.

First, correctness of person location and determination of the position and form factors were checked. Typically, these features were recognized satisfactorily. Sometimes blob lost feet, especially if the person was too far away from the sensor. Having more than one player, sometimes the blobs swapped IDs.

Next series of experiments included changes to the conditions of the BS. The following conditions were changed:

- Background not visible
- Side wall(s) visible
- Skewed sensor orientation
- Room with dark walls
- Varying environmental illumination
- Varying color and brightness of clothes
- Person sitting in wheelchair

Next series of experiments included different and varying movements of the acting persons, such as person lying on floor, jumping, entering AV, leaving AV, crossing AV from left to right, crossing AV from close to far and so on. Several typical gestured were made such as no movement, stretching the arms to the sides, to the front, or up, kicking with feet or hitting with arms.

The same was applied having a second person, whereby one person was doing one action and the second another.

All experiments were evaluated subjectively. In result of these experiments several changes were made in the algorithms, software tools, and parameter settings.

5.2. Application as therapeutic tool

One goal of our proposition was the development of therapeutic strategies for different groups of patients using the developed music device. For this reason, application scenarios were developed and preliminary experiments with groups of persons with different diseases were performed. According to the contacts and access possibilities of Grenzenlos e.V. society, the device was applied in sessions of person groups with a supporting person or under physician’s care belonging to different kinds of diseases, namely

- Blind people and persons with impairments of vision (B)
- Dementia patients (D)
- Psychologically disturbed people (P)
- People in wheelchairs (W)

Application of the device was performed concerning a predefined application scenario tailored to the special groups.

The goal of the application as therapeutic tool is an improvement of the mobility and pleasure to move on the one hand, and on the other hand a better self-confidence or other quality of life improvements especially for persons suffering from lethargy or depression.

5.2.1. General procedure

Every application starts with a warm-up program for the whole group, usually with support of external music. Sessions are performed either with single persons or groups with a small number of persons. The session typically started with selection of the music environment by the attending person(s). Then the subjects get to know the system by moving and hearing simultaneously. Each person moves individually according to the own feeling and physical possibilities.

5.2.2. Sessions with blind people

Blindness often implies that people feel insecure and partly without orientation in new unknown situations and environments. Sometimes they develop avoidance strategies in order to not meet the unknown and backtrack into known environments.

Blind people may get new experiences with the music demonstrator in unknown rooms. They are told they can move without obstacles in the room in which they produce acoustic stimuli through their movements and position in the room.

Typically, blind people test the location of the borders of the acting volume in the room. Here, it significantly helps to prevent music generation outside the acting volume. After the room experience, the blind probands move considerably safer and more confident. In repeated sessions the blind persons already start with more certainty in their movements.

In the long term, application helps to move more confidently (and not only within the music sessions themselves, but in general). The care-givers also observed a completely new self-awareness and an improved self-confidence of the blind people.

5.2.3. Sessions with dementia patients

Sessions with dementia patients were also performed with single persons or groups with approximately four persons. The mainly used music environment was “Fields”. By moving their body and recognition of the self-made music, dementia patients were sometimes more active, both by body movements and also by verbal communication. Especially animal sounds which can be used in the “Fields” environment helped dementia patients by bringing up their memories.

Typically, dementia patients are sad and skeptical in the beginning, but after a few minutes they are ready to try different settings of the device and to conduct their own new “compositions”. The general mood of the patient was observed to improve considerably.

5.2.4. Sessions with psychologically disturbed people

Sessions with psychologically disturbed people were mainly performed one-on-one, i.e. one care-giver and one patient. Introduction and application of the music demonstrator happened individually according to the degree of disorder and the wishes of the patient. As a result, an improved self-confidence and improved mobility was observed.

5.2.5. Sessions with persons in wheelchairs

Persons in wheelchairs typically suffer from different diseases: multiple sclerosis, paraplegia, or lost legs by accident or amputation. Hence, application of the music demonstrator is also individual. Some of the patients were very versatile with their arms and the head, while others only moved the head slightly or blink the eyes. However, in all cases application led to more mobility, a sense of joy, and sometimes to improved self-confidence.

6. Summary, discussion, and outlook

In this work, a new system for the fast generation of music through movement was introduced. Some significant improvements concerning the previously available system “MotionComposer MC2.0” were achieved, mainly a reduction of the latency time between movement and sound appearance to 180 ms. The new system could be successfully applied as therapeutic instrument with several groups of patients such as blind persons and persons suffering from dementia. Obviously, the system can also be used in the private range as a tool for making music by dancing and having fun and in this sense, can be seen as a tool for inclusion (participation by persons with and without disabilities together in one group).

Despite the improvements, the system is not yet perfect. Detection of more than one person and similarly the individual music generation by more than one person, does not yet work well.

Future work should start with further technical improvements of the system features. One goal is the further reduction of the latency between movement and corresponding sound from momentary 180 ms to a lower value. A reduction by approximately 30 ms can be achieved by immediate processing of the 2D data not synchronized with the relative low 3D frame rate of 20 Hz. Additionally, a shortening of the reaction time can be achieved using a more powerful soundcard in the PC. Final goal may be a latency time of about 100 ms.

Future work is also needed to replace the ToF-camera by a structured light projection based or passive stereo system in order to get 3D data faster, as well as lower the costs.

The current system will continue to be applied as therapeutic instrument, with testing planned into 2020 and beyond. Our partner Grenzenlos e.V. expects considerable therapeutic success with the different groups of patients. We hope that this kind of therapy finds a way into the catalogue of accepted therapies for the treatment of a variety of diseases.

Acknowledgements

This work was supported by the state Thuringia represented by the “Thüringer Aufbaubank” under grant label 2015 VF 0032. We also thank our partner Lucas instruments GmbH Jena [11] for building up the hardware of the system. Additionally, we thank Yves Schachtschabel for music software programming and implementation and Fusion systems GmbH Chemnitz for realization of a part of the data processing software.

References

[1] MotionComposer, http://www.motioncomposer.com/de., accessed 2019

[2] A. Bergsland and R. Wechsler, “Interaction design and use cases for MotionComposer, a device turning movement into music”, SoundEffects - An Interdisciplinary Journal of Sound and Sound Experience, special Edition on: Sound and Listening in Healthcare and Therapy. Vol 5, No 1, 2016, https://doi.org/10.24197/trp.31.2018.79-93.

[3] S. B. Gokturk et al., "A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2004, pp. 35–45, http://dx.doi.org/10.1109/CVPR.2004.291.

[4] H. Liu and L. Wang, “Gesture recognition for human-robot collaboration: A review”, International Journal of Industrial Ergonomicsm, Vol 68, 2018, pp 355-367, https://doi.org/10.1016/j.ergon.2017.02.004.

[5] Z. Zhang, “Microsoft kinect sensor and its effect”, IEEE MultiMedia, Vol 19 (2), 2012, pp 4-10, https://doi.org/10.1109/MMUL.2012.24.

[6] Orgue Sensoriel, http://orguesensoriel.com., accessed 2019

[7] Soundbeam, http://www.soundbeam.co.uk., accessed 2019

[8] C Bräuer-Burchardt et al., “Accurate Irritation-free 3D Scanning of Human Face and Body Sequences”, Proc. of 7th Int. Conf. on 3D Body Scanning Technologies, Lugano, 2016, pp. 231-238, http://dx.doi.org/10.15221/16.231.

[9] C. Bräuer-Burchardt et al., “High-Speed Accurate 3D Scanning of Human Motion Sequences”, Proc. of 6th Int. Conf. on 3D Body Scanning Technologies, 2015, pp. 194-201, http://dx.doi.org/10.15221/15.194.

[10] FusionSystems GmbH, https://www.fusionsystems.de/home.html., accessed 2019

[11] Lucas instruments GmbH, http://www.lucas-jena.de/., accessed 2019