# Music Interfaces for Novice Users: Composing Music on a Public Display with Hand Gestures

Gilbert Beyer
University of Munich
Oettingenstr. 67
80538 Munich, Germany
gilbert.beyer@ifi.lmu.de

Max Meier
University of Munich
Oettingenstr. 67
80538 Munich, Germany
max.meier@ifi.lmu.de

## ABSTRACT

In this paper we report on a public display where the audience is able to interact not only with visuals, but also with music. The interaction with music in a public setting involves some challenges, such as that passers-by as ‘novice users’ engage only momentarily with public displays and often don’t have any musical knowledge. We present a system that allows users to create harmonic melodies without being in need of a previous training period. Our software solution enables users to control melodies by the interaction, utilizing a novel technique of algorithmic composition based on soft constraints. The proposed algorithm does not generate music randomly, but makes sure that the interactive music is perceived as harmonic at any time. Since a certain amount of control over the music is assigned to the user and to ensure the music can be controlled in an intuitive way, the algorithm further includes preferences derived from user interaction that can be competing with generating a harmonic melody. To test our concept of controlling music, we developed a prototype of a large public display and conducted a user study, exploring how people would control melodies on such a display with hand gestures.

## Keywords

Interactive music, public displays, user experience, out-of-home media, algorithmic composition, soft constraints

## 1. INTRODUCTION

An important goal of interactive public displays reacting to e.g. body movements or hand gestures of passers-by is that interaction has to be such intuitive that novice users can start interacting immediately: Passers-by should be able to walk-up and use the content, or ideally control it in the intended way already by their initial, unconscious interaction. Interactive displays often allow manipulating visual objects that can for example be constituent parts of a brand identity, like a brand logo that can be moved along the display surface by hands or feet. For some reason however acoustic events do not appear at all or play only a secondary role within the interactive experience: often they are delimited to immutable sound objects just supplementing the visual interaction, or statically playing background music. Nevertheless, the enrichment by sound can enhance the interactive experience, and last but not least the identity of many brands that are advertised for on public displays is defined by both a visual and acoustic appearance.

On the other hand, beyond the context of interactive installations in public spaces, interactive music systems have become increasingly popular: with social music games like Guitar Hero, well-known songs can be re-played together, and easy-to-use musical applications for mobile devices such as the iPhone give everyone the possibility for musical expression, even without having any musical knowledge (see Figure 1).

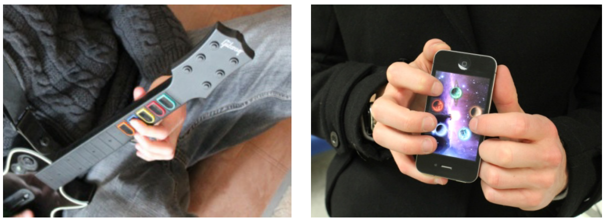

**Figure 1. Interactive music making with Guitar Hero and the iPhone**

In spite of enjoying great popularity and commercial success, such interactive musical applications have barely been employed in public spaces so far. We propose that the trends of interactive out-of-home media and interactive music making will successfully combine in the future, producing new media that will enable passers-by not only to play with, but also manipulate and shape melodies by means of interactive control mechanisms.

In this work, we present an approach which brings together interactive displays in urban spaces and interactive music systems. When combining public displays and music systems, the question arises how harmonic melodies can be created by ‘unskilled’ passers-by in a suitable way. With our approach for engaging with music in public spaces, it is possible to create music in many different styles. For example, music can be generated in such a way that it resembles the melody of a well-known song. This way, it is possible to develop interactive musical applications that give musical laypersons the feeling of successfully playing an instrument or composing music.

## 2. REQUIREMENTS FOR COMPOSING MUSIC ON A PUBLIC DISPLAY

How and if users interact with public displays depends amongst others on the external surroundings and usage context, the type of the display, as well as the number of individuals approaching. Usually passers-by can be assumed to be novice users and laypersons in regard to any application provided on such displays. Especially when it comes to interaction with

*NIME'11*, 30 May–1 June 2011, Oslo, Norway.

music, the question arises how an engaging user experience and a feeling of success can be achieved. Ideally, the demands of the input technique should be simple (yet allow an expressive performance), while the produced music should be as appealing as possible. To give novice users the feeling of successfully composing music in a public space and having fun during their short-term engagement with the application, we follow an approach where they are only capable of manipulating some musical parameters, and a software in the background makes sure that the generated melodies are perceived as harmonic and are reminiscent of some well-known musical themes. The input for the music generation also has to comply with the chosen interaction paradigm, which in the case of public displays is often multi-touch or vision-based interaction with hands. As people stopping in front of public displays are often pairs and small groups of individuals, a public display capable of multi-user interaction should also provide means to play music together in a successful way.

To comply with such requirements, we make use of a novel technique that allows generating music in real-time with respect to so-called preferences that express 'how the music should sound'. With this approach, it is possible to automatically derive preferences from given melodies in such a way that their characteristic properties can be preserved up to a certain extent (e.g. distinctiveness of a melody), while at the same time it is possible to flexibly alter them based on user interaction. Not only can the musical context of a melody be varied (e.g. instrumentation or style), also the melodic material itself can be subject to dynamic changes.

We use three types of preferences: First we use preferences for a single instrument which are derived from user interaction, e.g. a touch display or a motion tracking system. These preferences reflect how the user wants the music to sound, for example 'I want to play fast notes with a high pitch'. In our approach we generate music with only two parameters – 'pitch' and 'energy' – which are usually simple to extract from user interaction with both hands but are also expressive. Intuitively, these parameters continually control the note pitch (high/low) and the speed (fast/slow) at which the instrument should play. Interfaces based on these parameters are easy to play because they require only few musical skills (e.g. making exact rhythmic movements) – nevertheless, they provide much control over the music in a very direct way with immediate musical feedback.

The second type of preferences expresses general melodic rules: With this kind of preferences, it is possible to make the music consistent with a certain musical style (e.g. Hip-hop or Jazz). Furthermore, it is also possible to make the resulting melodies comply with a songs distinct acoustic identity. In most cases, the preferences derived from user interaction will be competing with a songs prominent characteristics, i.e. the user interaction does not fit the tune with respect to both tonality and rhythmics. Since a certain amount of control over the music is assigned to the user, it is inherently not possible to exactly play a given melody note by note. Nevertheless, it is possible to generate melodies which are similar to it by using note pitches as well as tonal and rhythmic patterns appearing in the tune's distinct melody. This way, melodies can be generated considering both interactivity and the recognition of a tune.

At last, we use preferences that coordinate several instruments playing simultaneously, for example a single player with static background music or multiple players among each other. This coordination is made by preferring harmonic intervals between different instruments. Furthermore, it is also possible to coordinate multiple instruments such that they play similar rhythmic patterns.

# 3. RELATED WORK

Of interest to our work are generally works on user-controllable music within digital media in public spaces. Yet, we currently know of no work that focuses on how to control a distinct melody within the interactive experience. A good overview on algorithmic composition is provided by [3] and [10]. Examples for interactive music composition and generation systems are Electroplankton [8] or Cyber Composer [7].

Related to our work are approaches for imitating musical styles: typical techniques for dealing with this problem are based on musical grammars or statistical models [4]. The Continuator [13] combines style imitation and interactivity. Based on a statistical model, the system is able to learn and generate musical styles e.g. as continuations of a musician's input. Our approach for generating music is based on constraint satisfaction problems. Automatic musical harmonization deals with the problem of creating arrangements from given melodies with respect to certain rules. Pachet and Roy made a detailed survey on musical harmonization with constraints [12].

To our knowledge, there is currently no work describing the combination of music generation and interactive applications in public spaces. In [11] a system for musical performance is described that acquires a user's physical actions and physiological state to alter stored data representing a music piece. In [9] pressure-sensitive controls allow people with disabilities to control the generation of music. The system introduced in [2] uses a performance device to interactively control several aspects of a composition algorithm. A general-purpose position-based controller, where the position signal may also be used for generating music, is described in [14].

Our approach for generating music is based on a reasoning-technique called soft constraints which allows dealing with soft and concurrent problems in an easy way. Bistarelli et al. [1] introduced a very general and abstract theory of soft constraints based on semirings. Building on this work, in [6] monoidal soft constraints were introduced, a soft-constraint formalism particularly well-suited to multi-criteria optimization problems with dynamically changing user preferences. Soft constraints have successfully been applied to problems such as optimizing software-defined radios [15] or orchestrating services [16]. We introduced a soft-constraint based system for music therapy in [5], giving us basic proof of concept with this technique.

# 4. COMPOSING MUSIC WITH SOFT CONSTRAINTS

To realize interactive, user-controllable music systems in public spaces we developed a technical solution for real-time music generation that helps to coordinate the different characteristics of user interaction, the acoustic identity of a tune and the general harmonic and rhythmic concordance of instruments.

We make use of a framework for algorithmic composition of music which is based on soft constraints [5]. With this framework, music can be interactively generated in real-time by defining preferences as described in the previous section. All preferences can also be generated dynamically, which allows to compose music in real-time, e.g. based on user interaction by continually defining preferences which reflect 'how well the music matches the interaction'. In general, a soft constraint expresses how well an assignment of values to variables (a valuation) matches a desired result. A valuation is a function from variables to values:

$$Valuation = (Variable \to Value).$$

The extent to which this valuation is desirable can be expressed in various ways. The cited theory introduces a very elegant way of rating valuations with a set of grades and several operations for combining or comparing grades. Many concrete kinds of grades can be used, for example based on numbers or Boolean values. A soft constraint assigns a grade to each valuation:

$$SoftConstraint = (Valuation \rightarrow Grade).$$

Typically, one is interested in the best possible valuation which can be computed with a general solver for soft constraints. In our application of soft constraints for generating music we want to assign actions to voices: each voice corresponds to a certain sound (e.g. a piano, guitar or synthesizer sound); actions are for example 'play a note' or 'pause'. When an instrument should be polyphonic, it has to have an according number of voices. We use soft constraints to rate action assignments:

$$(Voice \rightarrow Action) \rightarrow Grade.$$

At certain time intervals, each instrument is being asked to state preferences for its own notes. These preferences from all instruments are then extended with global coordination preferences and combined to a single constraint problem. This problem is being solved, yielding an action for each voice which satisfies the preferences best. In the next section, we will introduce a prototype where hand gestures are used to control music. Based on an optical tracking system, we derive two parameters from a user's movements: the total amount of movement (corresponding to the rate of played notes) and the average vertical position of all movements (corresponding to pitch). Based on these two parameters, preferences are generated reflecting the desired speed and pitch. For example, when the user makes fast movements and lifts his hands up, the music should also be fast and have a rather high pitch. Vice-versa, when the user is moving slowly and his hands are down, the music should be slow with a low pitch.

The music should fit the user interaction on the one hand, but we also want it to fit to a given tune on the other hand. This is realized with an additional preference reflecting 'how well the music matches a tune's distinctive melody'. This preference is generated based on a timed transition model representing the tune's note pitches and rhythmic patterns as well as transitions between notes (e.g. 'C is often followed by E or another C'). Our approach is based on a custom transition model which represents sequences of events aligned upon a structured metric grid. Intuitively, the model represents (1) how often an event occurs at a certain metric position and (2) how often other events follow this event at this position. Following typical terms from the closely related area of probability models, the 'events' are called states. The discrete metric positions (representing 'time') are called steps:

$$State$$
$$Step = \{0, \dots, n\}$$

In each step, each state has a certain weight for a given voice. This weight represents how often the state occurs at the given step:

$$stateWeight_{Voice} : Step \times State \rightarrow \mathbb{R}$$

The transitions between states at a given step are represented with the following function. The first two arguments define the original step and state – the third argument defines the next state. Transition weights are always defined for subsequent steps; the state in the third argument is implicitly assumed to be on the next state:

$$transitionWeight_{Voice} : Step \times State \times State \rightarrow \mathbb{R}$$

Figure 2 visualizes a timed transition model with three steps and states. State weights are visualized with black circles: the bigger the circle, the higher the weight. The transition weights are visualized with arrows (a thicker arrow indicates a higher weight). When the model is untrained, all weights are the same. Training the model modifies the weights; the right picture visualizes a trained model with shifted weights.

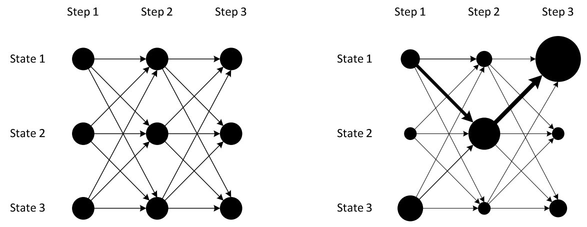

**Figure 2. Transition model visualization (left: empty model, right: trained model)**

The actual states can be modeled in several ways: the simplest way is to directly use the existing set of actions as states. However, there would be a little disadvantage: if note pitches are directly used within the states, it is not possible to play a model in another tonal scale. If this is desired, it is better to use abstract stages in a tonal scale rather than concrete note pitches. Now, we define a constraint which expresses 'how well an action matches the data represented in the model'. Given the last step and the last actually executed state (the state corresponding to the last action chosen by the constraint solver), we can compute a total weight for each state on the subsequent step. This is done by just summing up the transition weight and the step weight itself:

$$\begin{aligned}
&totalWeight_{Voice} : Step \times State \ \times Step \times State \rightarrow \mathbb{R} \\
&\ totalWeight_{v}(lastStep, lastState, step, state) \\
&= transitionWeight_{v}(lastStep, lastState, state) \\
&+ stateWeight_{v}(step, state)
\end{aligned}$$

The constraint itself for a certain voice is constructed based on the last step, the last executed action and the current step. When the sets of actions and states are identical, the constraint can be defined like this:

$$\begin{aligned}
&\ modelConstraint_{Voice} : (Step \times Action \ \times Step) \\
&\rightarrow ((Voice \rightarrow Action) \rightarrow \mathbb{R}) \\
&\ modelConstraint_{v}(lastStep, lastAction, step)(val) \\
&= totalWeight_{v}(lastStep, lastAction, step, val(v))
\end{aligned}$$

To sum it up, we have preferences based on user interaction as well as preferences reflecting the similarity to a tune, and – in most cases – these preferences will be competing among each other. Furthermore, it is also possible to coordinate several instruments with additional global preferences. In our public display scenario, we define a global constraint which maximizes the amount of musical harmony between the interactive instrument and background music. Soft constraints are very appropriate for dealing with such problems and allow accommodating several concurrent preferences in an easy yet expressive way. When the preferences have been stated, a soft constraint solver can be employed for computing the best possible notes with respect to all preferences. We use a soft constraint solver which was originally prototyped in Maude and that we later implemented in a more efficient version in C#, making it possible to use it in a soft real-time environment.

## 5. PROTOTYPE AND EVALUATION

To explore how novice users can compose music on a public display with our soft-constraint framework, we developed a prototype with which users can interactively play music with hand movements. The sensing of hands is realized using marker-based techniques. To examine which gestures users would use to manipulate music, we developed several sample applications where the note pitch of the music can be controlled by up-and-down movements of the hands and the rate of played notes by the velocity of hand movements (see Figure 3).

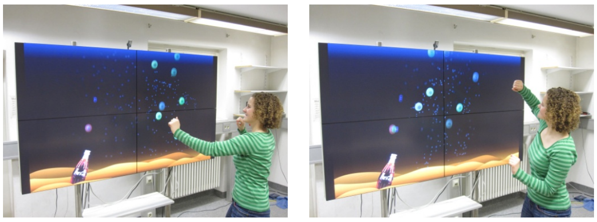

**Figure 3. Public display prototype that allows novice users to control music by hand movements**

When someone starts to interact with the system, he can realize the connection between his movements and the notes he hears: when the movements become faster, the notes will also play faster – not moving at all leads to silence. The resulting melodies do not only fit to the person's movements, they are furthermore being generated in a way that they comply with a well-known melody. We developed different gesture-based techniques for controlling the music with hands:

The first interaction technique allows the user to control visual elements and acoustic events only with one hand at a time. The note pitch of the music is controlled by up-down movements, and the rate of played notes is controlled by the velocity of movements. The second technique allows the user to control visuals and music with both hands, and for computing note pitch and rate of played notes the mean values of both the hand vertical positions and velocity are taken. The third interaction technique extends the second technique by allowing the user to control the acoustic events (note pitch and rate of played notes) with separate hands, i.e. one hand controls the note pitch and the other hand controls the rate.

Even without any previous instructions, most users were aware that they have control over the music. Only 2 out of 21 people stated they did not recognize the connection between their hand movements and the music. No user stopped interacting while standing in front of the system for a longer period, and the average user made hand gestures for over 90% of the time which gives us further confidence that people understood the basic interaction paradigm. Based on the videos, we analyzed how long it took until people interacted in the way we intended, i.e. when they started to primarily make hand gestures which are relevant for the music generation. The variant based on only one hand took 132 seconds on average, the variant based on the mean values of both hands took 118 seconds and the third variant with separate hands for both parameters took 92 seconds. Even if most users seemed to interact in an effectual way interviews revealed that not everybody did consciously identify the parameters 'pitch' and 'rate of played notes' and how they can be controlled: 12 out of 21 people stated that they used up-and-down movements to control the music and 10 out of 21 people could tell how note pitches can be controlled; only 2 users understood how they can vary the rate of notes. Nevertheless, the results from the user observations make us confident that hand gestures are well-suited for interacting with music without any previous training.

## 6. CONCLUSION

We introduced an approach for musical composition in public spaces, combining the trends of interactive public displays and interactive music systems in the future. Systems that allow controlling sounds by the interaction can open up new opportunities in advertising, entertainment, or installation art. Yet, as passers-by are usually novice users of any deployed interactive installation and often musical laypersons, we believe that systems where users can play note by note offer fewer opportunities for experiencing music. Instead means should be offered that give users the feeling of success when interacting, while still having a certain amount of control over the music. First user tests with our prototype of a large public display showed that music generation with soft constraints serves this purpose quite well. The next step is to investigate how users interact with the proposed system in the wild.

## 7. REFERENCES

[1] Bistarelli, S., Montanari, U., Rossi, F. Semiring-based constraint satisfaction and optimization. *Journal of the ACM*, vol. 44(2), 1997, 201–236.

[2] Chadabe, J. *Interactive music composition and performamce system.* US Patent 4526078, 1985.

[3] Essl, K. Algorithmic composition. In: Collins, N., d'Escrivan, J. (eds.) *Cambridge Companion to Electronic Music.* Cambridge University Press, Cambridge, 2007.

[4] Farbood, M., Schoner, B. Analysis and Synthesis of Palestrina-Style Counterpoint Using Markov Chains. In *Proc. of the Intl. Computer Music Conf.* Havana, 2001.

[5] Hölzl, M., Denker, G., Meier, M., Wirsing, M. Constraint-Muse: A Soft-Constraint Based System for Music Therapy. In *Proc. of Third International Conference on Algebra and Coalgebra in Computer Science (CALCO'09).* Springer, Udine, 2009, 423–432.

[6] Hölzl, M., Meier, M., Wirsing, M. Which soft constraints do you prefer? In *Proc. of Workshop on Rewriting Logic and its Applications (WRLA 2008).* Budapest, 2008.

[7] Ip, H., Law, K. Kwong, B. Cyber Composer: Hand Gesture-Driven Intelligent Music Composition and Generation. In *Proc. of 11th International Multimedia Modelling Conf. (MMM'05)*, Melbourne, 2005, 46–52.

[8] Iwai, T., Indies Zero and Nintendo: *Electroplankton.* Game for Nintendo DS, 2005.

[9] Jubran, F. *Sound generating device for use by people with disabilities.* United States Patent 2007/0241918 A1, 2007.

[10] Nierhaus, G. *Algorithmic Composition.* Springer, Heidelberg, 2008.

[11] Nishitani, Y., Ishida, K., Kobayashi, E., Yamaha Corporation. *System of processing music performance for personalized management of and evaluation of sampled data.* United States Patent 7297857 B2, 2007.

[12] Pachet, F., Roy, P. Musical Harmonization with Constraints: A Survey. *Constraints* 6 (1), 2001, 7–19.

[13] Pachet, F. The Continuator: Musical Interaction With Style. In *Proc. of the International Computer Music Conference*, ICMA, Gotheborg, 2002, 211–218.

[14] Wheaton J. A., Wold E., Sutter A. J., Yamaha Corporation. *Position-based controller for electronic musical instrument*. United States Patent 5541358, 1996.

[15] Wirsing, M., Denker, G., Talcott, C., Poggio, A., Briesemeister, L. A Rewriting Logic Framework for Soft Constraints. In *Proc. of Workshop on Rewriting Logic and its Application (WRLA 2006).*Vienna, 2006.

[16] Wirsing, M., Clark, A., Gilmore, S., Hölzl, M., Knapp, A., Koch, N., Schroeder, A. Semantic-Based Development of Service-Oriented Systems. In *Proc. FORTE 2006*. Springer, Heidelberg, 2006, 24–45.