Singing proficiency research for new music interface design

My general research interest is to design new music interfaces to enable non-musicians to create music. During my research, I find voice to be a particularly interesting musical instrument. Compared to other instrument, voice has very limited frequency range, which spans typically 2 octaves in comfort zone. And our voice cannot produce polyphonic sound. However, voice is the only instrument that everyone, including non-musicians, can control pitch naturally and accurately given a natural tempo within voice range, with incredible variations of different timbres. Given such an intuitive ease of using voice for human, I'm interested in designing a music interface that can strengthen people's ability for musical expression and compensate the disadvantages of voice, making voice a more expressive and intuitive musical instrument for non-musicians.

Nowadays, smart mobile phones are becoming more and more powerful in computational ability with more diversified ways of interaction, such as multi-touch, accelerometer, microphone. As well, mobile phones are very convenient to carry around. What's more interesting, it has a great social interaction platform for people all over the world to communicate remotely with each other. Therefore, a lot of new music interfaces have been invented in smart phones, like magic guitar, magic piano, ocarina, glee, etc. Therefore, I started to research on the design of mobile phone music interface related to singing.

What techniques do I plan to use for my project

I plan to use Mobile Music (MoMu) toolkit[1] for iOS development in iphone. It is an open source software API for mobile phone music application that provides a convenient framework for not only audio process features such as DSP or reverberation filters, but also for various sorts of interactive features such as multi-touch, graphics, networking-OSC, etc. In addition, iphone has very good computational power to support audio synthesis.

Now I have a working prototype in my iphone to reproduce my voice in real time with very simple sound effects such as delays and amplification. The next step would be to use various filters in Momu toolkit audio API. The first filter I'm going to explore would be a reverberator, such as JCRev.cpp, NRev.cpp, PRCRev.cpp, because reverberation is popular audio applications. However, it is necessary at the same time to find out how the computational demands of reverberation processes can be handled in my prototype.

How singing proficiency research gives insight to the design of this project

For this project, I would like to explore how people would react to digitally processed feedback of singing sound. In a natural singing, hearing auditory feedback of one's own voice has a very short feedback loop consisting of two different pathways. On the one hand, it has bone conduction of the original voice; On the other hand, it also has air-conducted sound feedback from the environment. Using the iphone, the latter can be modified by applying different digital filters. With the different feedback, what kind of voice music interface design would be more intuitive, not confusing for people? For example, if pitch shift effect is added in phone sound process, can the user learn to adjust to the new feedback, or get confused? Or maybe some harmonic pitch shift works well, but inharmonic pitch shift would make people confused. These are just simple examples, but there would be many more possible applications. But how to choose the effective design is a big problem. This leads to the question as to what is the nature of human singing abilities. In music perception literature, one research paper [2] on singing proficiency among non-musicians I recently have read gives me an interesting opportunity to think about these questions.

This paper describes the study about singing proficiency in normal non-musician population. Previous research has found that 10% to 15% population cannot sing a familiar melody correctly at a slow tempo, and that poor singing might be caused by various mechanisms, thus leading to different patterns of impairment in singing. The study by Dalla Bella and colleagues investigated whether they can find various types of singing impairment. The authors designed 3 singing tasks, and they tested 39 musically untrained university students. The first task was Sung Performance Battery task, including repetition of isolated pitches, intervals and short novel melodies; The second task was the production of familiar melody at a natural, slow tempo; The third task was repetition of familiar melody at a fixed tempo. The second task required the participants to sing from memory, while the third task started with listening to the song that they had to reproduce afterwards. Their performance was recorded acoustically and analyzed with variables including: numbers of errors in pitch interval (relative pitch), pitch contour, time (relative tempo), as well as deviation of pitch interval, initial pitch (absolute pitch, repetition task only), tempo (absolute tempo, repetition task only), and temporal variability.

The result shows that poor singing happens much more in the pitch domain than in the time domain. They further compared pairs of different types of errors and found that within the pitch domain, there are people who perform good at pitch interval but poor at pitch contour, good at pitch interval deviation but poor at initial pitch, and within tempo domain, there were people who were good at temporal variability but poor at tempo deviation, or vice versa. Based on both absolute and relative accuracy in performance, the authors classified poor singing in 4 “phenotypes”, including poor pitch transposer, poor tempo transposer, poor pitch internal singers, poor duration singers, indicating that different mechanisms of pitch and time processing might be independent in some degree.

Discussion of the paper

I think this paper is interesting for the following two reasons. First, the authors redefined the abstract definition of “poor singing” into four categories: for absolute measurement, it’s divided into poor pitch transposers and tempo transposers; for relative measurement, it’s divided into poor pitch-interval singers and poor duration singers. By decomposing the singing task, more accurate study in different aspect of singing is done and new music interfaces could be developed to help improve people with certain kind of singing impairment. Second, it suggests that the relative/absolute processing mechanisms of pitch and time processing may have some degree of functional independence, which deserves further research in music perception.

However, I also see some problems in this paper. First, only 39 samples in a certain university participated in this research. The distribution of their performance measures does not seem to be able to represent the general singing proficiency distribution of human. People from different culture, age, might have dramatically different distribution compared to this 39 people group. Second, in the classification of poor singing, for each of the classification of four different singing "phenotypes", only 1-3 samples are thought to represent the type. The number of the samples seems too small to justify the categorization.

Future research for singing music interface design

I would like to conduct experiments by modifying the following aspects of singing feedback, and see how music perception research would benefit music interface design.

First, I would like to use a broader pitch range than the individual's voice range in the feedback. Since people's voice pitch range is very limited in music expression, it'll be interesting to see how people use a broader range to control to enhance their musical expression. I plan to implement algorithms that produce different pitch feedback with input voice pitch. Finger multi-touch control will be used to offer broader pitch control. It would be interesting to see how people respond to the feedback sound when they hear a different processed pitch output and a certain fraction of bone-conducted original pitch combined together.

Second, I would like to use pattern control. Usually when people are singing or are playing most instrument, the output sound has a one-to-one correspondence to people's input. As a result, people learn instrument from one note to a phrase, to a section and finally to a whole piece, which usually takes a long time to learn and many people give up learning in the process because only at the end of training can people actually hear the whole song. This kind of bottom-up learning makes instrument practice often boring to a lot of children. Thus, I would like to explore how people would respond when, instead of learning from a single musical unit to build up, people can control a block of output notes with one input source. For example, with a single input sound, the output could be 3 related notes coming out together, or coming out one by one, or certain style of improvisation. The pattern can be selected and controlled by finger multi-touch or single-touch on screen. Also the output could be a certain sound effect filter real time process. It would be interesting to see how people react to this kind of one-to-more music interaction.

Further music perception and interface design will offer me much more insight on how the two areas would benefit each other, opening the door to more intuitive music interface design that targeted exactly what non-musicians really need to create music, as well as deep understanding of human music ability with new tools being experimented.

References

1. Bryan, N. J., Herrera, J., Oh, J., and Wang, G. 2010. “MoMu: A Mobile Music Toolkit.” In Proceedings of the International Conference on New Interfaces for Musical Expression. Sydney 2010.

2. Simone Dalla Bellaa,b and Magdalena Berkowskaa. 2009.“Singing Proficiency in the Majority Normality and “Phenotypes” of Poor Singing”. The neurosciences and music III—disorders and plasticity.