MOGAT: Mobile Games with Auditory Training for Children with Cochlear Implants

Yinsheng Zhou[1], Khe Chai Sim[1], Patsy Tan[2], Ye Wang[1]
[1]School of Computing, National University of Singapore, 117417, Singapore
[2]Singapore General Hospital, 169608, Singapore
{yzhou86, simkc, wangye}@comp.nus.edu.sg, patsy.tan.l.p@sgh.com.sg

ABSTRACT

Cochlear implants have improved the lives of tens of thousands of the hearing impaired by providing sufficient auditory perception for speech, but these devices are far from satisfactory for music perception. Many cochlear implant recipients, especially pre-lingually deafened children, have difficulty recognizing and producing specific pitches. To improve musical auditory habilitation for children post cochlear implantation, we developed MOGAT: MObile Games with Auditory Training. The system includes three musical games built with off-the-shelf mobile devices to train their pitch perception and intonation skills respectively, and a cloud-based web service which allows music therapists to monitor and design individual training for children. The design of the games and web service was informed by a pilot survey (N=60 children). To ensure widespread use with low-cost mobile devices, we minimized the computation load while retaining highly accurate audio analysis. A 6-week user study (N=15 children) showed that the music habilitation with MOGAT was intuitive, enjoyable and motivating. It has improved most children's pitch discrimination and production, and several children's improvement was statistically significant ($p < 0.05$).

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: User-centered design; H.5.5 [**Sound and Music Computing**]: Signal analysis, synthesis, and processing; K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities

General Terms

Design, Experimentation, Human Factors.

Keywords

Music, mobile, game, auditory habilitation, cochlear implant, children.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MM'12, October 29–November 2, 2012, Nara, Japan.
Copyright 2012 ACM 978-1-4503-1089-5/12/10 ...$15.00.

1. INTRODUCTION

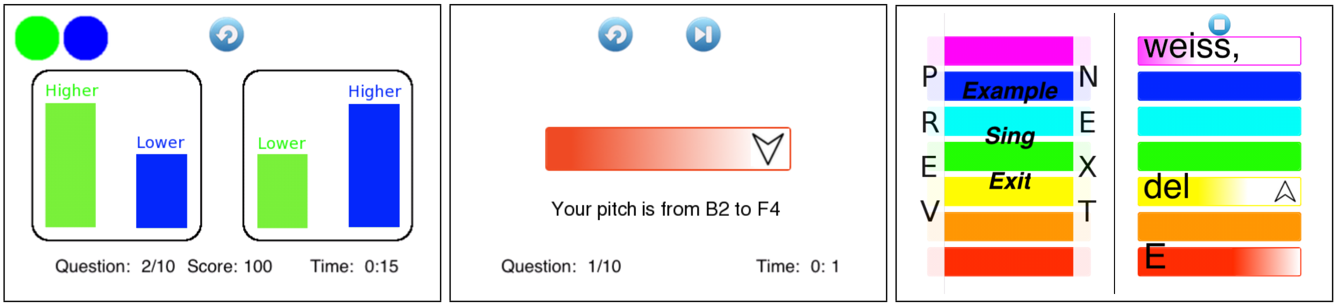
Music plays an important role in people's lives. The vast majority of people perceive music with their unaided ears, but millions of people have partially or profoundly impaired hearing. How can they experience music? One approach is through amplifying vibrations with mechanical devices; Beethoven was an early adopter of such technology [19]. Another approach is to sense the tactile vibrations; the highly acclaimed percussionist Evelyn Glennie "feels" music with different parts of her body [26]. It must be noted that these musicians experienced sounds before the onset of deafness. As a result, they could imagine music in their minds [28] without hearing it. However, some children are deaf before forming memories of music or even language. For those pre-lingually deafened children, the only avenue to develop their hearing sense is via cochlear implants (CI). By surgically implanting electronic devices in their cochleas, children[1] with CI are able to hear sound and music [21].

CIs are quite successful in supporting spoken communication, but they are still less than ideal for encoding and transmitting musical sounds [30]. The rich spectrum of musical sounds are not well preserved by feature extraction devices; intelligible speech only requires a very narrow frequency bandwidth. CI recipients generally have poorer perception and identification of melodic patterns [22] and musical timbre [18] than normal hearing people.

Auditory habilitation is an important part of the implant process to boost recipients' adaption for the devices post cochlear implantation. It must encompass many features including not only speech but also music perception and production [13]. As such, musical auditory habilitation is usually set up as an adjuvant habilitation process to their standard habilitation programs [35]. Musical habilitation has shown improvement in recipients' melody recognition [24], timbre identification [23], and can enhance recipients' self-esteem and increase their motivation for practicing [13]. The main obstruction is a lack of appropriate teaching resources, professional training and administrative support [27].

In order to focus on their musical habilitation, it was imperative to involve children with CI in the design loop from the beginning. Therefore, a pilot study is conducted to understand their deficiency in terms of pitch and rhythm perception/pitch production in contrast to normal hearing peers. Based on those findings, we designed and developed

[1] Estimates show that as of December 2010, approximately 219,000 people worldwide have received implants [12], of which at least 60,000 are children recipients [11] whose number is increasing dramatically every year.


Higher
Lower
Lower
Higher
Question: 2/10 Score: 100 Time: 0:15
Your pitch is from B2 to F4
Question: 1/10 Time: 0: 1
P R E V
Example
Sing
Exit
N E X T
weiss,
del
E

(a) *Higher Lower* (b) *Vocal Matcher* (c) *Ladder Singer*

Figure 1: The game interfaces in MOGAT

MObile Games with Auditory Training (MOGAT) using off-the-shelf mobile devices. MOGAT aims to provide a fun, intuitive, and cost-effective way to enhance musical habilitation for children with CI.

Since MOGAT was designed for pre-lingually deafened children with cochlear implants, their unique characteristics must be considered when selecting objectives, content, and particular stimuli. Pre-lingually deafened children with CI are not exposed to sound until their implantation, while post-lingually deafened adults with CI have in their mind the experience of sound established prior to their deafness. As such, the training methods for children with CI should be specifically designed, which will be different from the existing music training program for adults [25]. Delay in their language development may affect their cognitive and behavioral development [33, 32], which will further inform the choice of vocabulary and material for those students.

MOGAT contains three structured music games focusing on pitch-based habilitation. *Higher Lower* (Figure 1a) trains pitch interval perception ability; *Vocal Matcher* (Figure 1b) focuses on single pitch production with appropriate voice control; *Ladder Singer* (Figure 1c) combines pitch, breath, and lyrics in an intuitive user interface to guide users in singing songs. Using highly accurate singing analysis algorithm, it transcribes users' pitch and provides real-time feedback to help them to sing correctly. With optimized computation load, MOGAT can be built into low-cost mobile devices to provide a cost-effective way for children's habilitation. Audio is recorded on mobile devices and uploaded to a server. With all the data synchronized, stored, and managed in our server, MOGAT realizes a cloud computing service to enable music teachers and therapists to support a large number of children's habilitation. The MOGAT web application allows teachers to visualize the progress of individual students over days, weeks, and months. Furthermore, teachers are able to pinpoint students' singing problems and send comments or encouragements to their students.

This paper's main contributions can be summarized as:

- MOGAT is the first integrated solution to support musical (rather than general audio) habilitation for children with cochlear implants.
- An analysis of users' pitch/rhythm perception and intonation accuracy that guides the design, which is catered specifically to their musical needs and cognitive abilities.
- We conducted systematic and in-depth user evaluation to test the effectiveness of MOGAT in enhancing musical habilitation for children with CI.

2. RELATED WORK

2.1 Auditory Habilitation and Its Applications

For children with CI, auditory habilitation is critical to their hearing and speech development [38]. Due to the spectrally degraded signal pattern provided by the implant and the large interpersonal variability [37], passive adaptation via long-term use of the devices may not be adequate. Nevertheless, active learning via auditory habilitation has been shown to be effective in the speech recognition and production of the hearing impaired [15, 16, 36]. Meanwhile, auditory training with music stimuli can help to improve music recognition and production for cochlear implant users [24, 13, 22]. However, because of time and cost considerations, it is almost impossible for hearing healthcare professionals and therapists to provide extensive and intensive auditory therapy to CI recipients [21].

Recently, computer-assisted speech training (CAST) system has been developed to augment auditory habilitation approaches by providing greater flexibility with minimal costs and supervision. Research shows that moderate amounts of auditory training performed at home with the computer-based speech training (CAST) software resulted in significant improvements in the speech recognition for both adult CI recipients [21, 20] and children with CI [38]. A typical example is Sound Express Auditory Training (SEAT) system [8], a self-directed auditory training program on personal computers. Although it has some useful features (e.g., interactive interface and feedback) to help CI users to practice their perception of spoken sounds, it is not optimized for musical habilitation and lacks teacher guidance.

Unlike speech perception, music perception relies more strongly on pitch perception and thus contains information very differently from verbal communication. Due to the device limitation, implant listeners are reported to have great difficulty with complex pitch perception in comparison with speech perception [30]. Unfortunately, relatively few studies have explored the effects of auditory training on CI users' music perception or production. The only system for this purpose was designed for post-lingually deafened adults [25, 24]. However, music perception by pre-lingually deafened children with CI is very different from post-lingually deafened adult CI users [22] who already have the experience of acoustic sound before their deafness. Pre-lingually deafened children begin to form their concept of sound until implantation, and all their central speech and music patterns are developed in the context of electric hearing. Therefore, it is imperative and important to develop a musical auditory system specifically for children with CI.

2.2 Music Education Applications

The objectives of the existing music training applications are not compatible with our purpose. Most vocal training applications (e.g., [10, 9, 29]) were designed to develop specific professional listening and performing techniques for users who already have decent hearing acuity (e.g., recognition of chords, harmonics, and development of unique vocal style or instrument skills). Therefore, the components of these applications are not able to assist children with CI for their habilitation which has a different focus (e.g., pitch and rhythm perception and fundamental singing ability). Although Karaoke games [5, 4, 3] seem to have some values in learning singing songs using real-time visual feedback and machine scoring, our study show that it is harder for them to understand and use this kind of Karaoke games than our designed singing game. Family Ensemble [31] and MySong [34] use automatic accompaniment generation technique to enhance users' motivation in piano playing and vocal singing. MOGCLASS [40, 39] provides a collaborative music system to enhance students' music experience during music classes. However, all these projects are designed for normal people, so they will not be so useful for our special user group.

3. AUDIO ANALYSIS

3.1 Automatic Note Annotation

3.1.1 Note Segmentation

Since users will be singing the hard consonant "La" (see Section 4.1.1), notes onsets are easily identified in a spectrogram (Figure 2). Our input audio is a monophonic signal at 24 kHz. We take the short-term fourier transform (STFT) with a hamming window, using a window size of 512 and a FFT length of 512. The detection function is constructed using the half-wave rectified spectral flux,

$$SF(k) = \sum_{i=0}^{n-1} H(s(k,i) - s(k-1,i)) \tag{1}$$

where $s(k,i)$ is the magnitude of the i^{th} frequency bin in the k^{th} frame, and $H(x) = (x + |x|)/2$ is a half-wave rectifier assigning zeros for its negative arguments. The rectification emphasizes onsets rather than offsets. The spectral flux was first normalized to $[0, 1]$ by subtracting the minimum and dividing the maximum absolute difference. Then, a low-pass filter was applied to remove jitter and noise. Finally, a high-pass FIR filter adaptive threshold was subtracted from the normalized spectral flux to create a "pruned" flux.

$$SF_{pruned}(k) = \alpha + \frac{\beta}{H}\sum_{i=k-H/2}^{k+H/2} SF(i) \tag{2}$$

We empirically determine the moving window size $H = 10$, and $\alpha = 0.03, \beta = 1.2$. After post-processing and thresholding the detection function, peak-picking is used to identify the local maxima in the adjusted spectral flux above the defined threshold.

3.1.2 Pitch Estimation

We used the YIN pitch estimation algorithm [17]. In order to find the periodicity (indicated by $\widetilde{\tau}$, i.e., the number of samples in the period) of a discrete time-domain signal s, we begin by calculating the squared difference function $d(\tau)$ for a desired range of lag values:

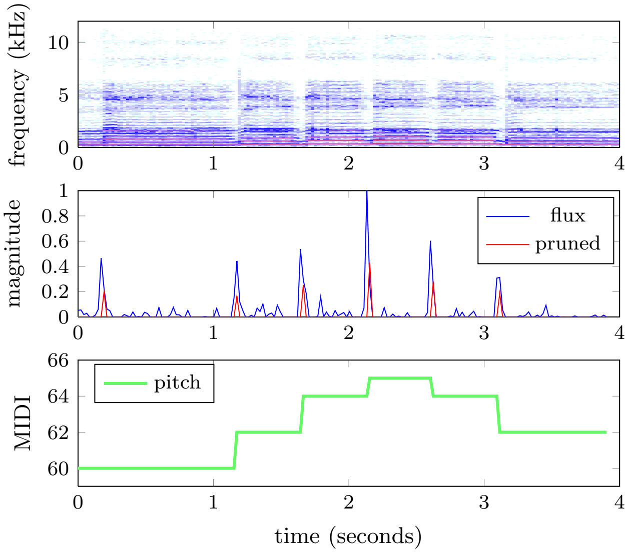


Figure 2: Note segmentation result on a singer's recording. The top plot is a spectrogram; the lower plot is the normalized and adjusted spectral flux.

$$d(\tau) = \sum_{n=0}^{N-1} (s(n) - s(n+\tau))^2 \tag{3}$$

We then use a cumulative mean normalized difference function to determine the aperiodicity of the audio frame:

$$d'(\tau) = \begin{cases} 1, & \tau = 0 \\ d(\tau)/[(1/\tau)\sum_{j=1}^{\tau} d(j)], & \text{otherwise} \end{cases} \tag{4}$$

Next, we search for the smallest value of τ that gives a minimum of $d'(\tau)$ smaller than a given absolute threshold $\kappa = 0.10$. If no such value is found, we instead search for the global minimum of $d'(\tau)$. Once we find the lag value $\widehat{\tau}$ from last step, we interpolate $d'(\tau)$ at $\widehat{\tau}$ and its immediate neighbors with a second order polynomial. The length of the period $\widetilde{\tau}$ corresponds to the minimum of the polynomial in the range of $(\widehat{\tau} - 1, \widehat{\tau} + 1)$, and the pitch is estimated as the sampling rate divided by $\widetilde{\tau}$. Since consonant and silence frames have relatively high aperiodicity, we omit values of $d'(\tau) > 0.15$ (value set experimentally). We then convert the pitch (in Hz) to a MIDI pitch value.

Within each note segment, we adopt the median as the pitch value for all frames. After note segmentation and pitch estimation, the output of automatic note annotation is the note sequence $\mathbf{O} = o_1, o_2, ..., o_t$, which will be the input for the singing evaluator in the following section.

3.2 Singing Evaluator

In this study, "intonation accuracy" refers to the similarity between subject's pitch contour and the reference one. In order to find the optimal matching path between pitch contours $\mathbf{A} = \{a_1, a_2, ..., a_N\}$ (the reference from sheet music) and $\mathbf{B} = \{b_1, b_2, ..., b_M\}$ (the detected values from Section 3.1), we adopt the classic note-level Dynamic Time Warping (DTW) method. A singer may shift the pitch up or down ("transposition" in musical terms) by a constant in-

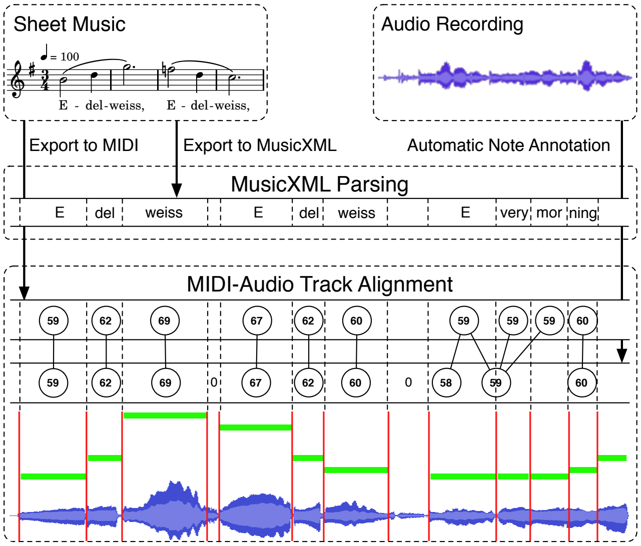


Figure 3: Alignment of recorded audio with the reference MIDI and MusicXML files. There are three rows of information for alignment from top to bottom: lyrics, MIDI pitch sequence, and audio track annotation. A "pitch" of 0 indicates breath noise or silence.

terval to fit his/her vocal range. To detect transposition, we enumerate 12 semitones in a octave and shift the subject's pitch contour from one octave down to one octave up to find the minimum matching cost. The absolute differences between two pitch contours then are averaged across all the notes in reference sequence **A** (5).

$$C_{DTW}(\mathbf{A},\mathbf{B}) \;\; = \frac{1}{N}\min_{i\in[-12,12]\bigcap\mathbb{Z}}\{Dist(\{a_1,a_2,...,a_N\}, \{b_1+i,b_2+i,...,b_M+i\})\} \tag{5}$$

where

$$Dist(\{a_1,a_2,...,a_N\},\{b_1,b_2,...,b_M\}) = D_{N,M} \tag{6}$$

$$D_{i,j} = d(a_i,b_j) + \min(D_{i-1,j-1}, D_{i-1,j-2}, D_{i-2,j-1}) \tag{7}$$

$$d(a_i,b_j) = |a_i - b_j| \tag{8}$$

where $d(a_i,b_j)$ is the absolute difference between note a_i and b_j. $D_{i,j}$ is the minimum cumulative absolute difference up to a_i and b_j. $Dist(\mathbf{A},\mathbf{B})$ is the absolute difference between two pitch contours **A** and **B**, note by note.

3.3 Audio Alignment to MIDI and Lyrics

The meta-data in the game of *Ladder Singer* consists of the pitches, onsets, lyrics, and sample audio. The sample audio was recorded from one of our female teachers' singing, while other data was extracted directly from the sheet music. For score editing, we used Noteflight [7] to associate each note with its corresponding syllable in a word in lyrics. After editing, the notes and lyrics were exported into MIDI and MusicXML files respectively. To control the animation synchronously with audio playback, we need to align the MIDI with the audio track.

Alignment is performed with the algorithm in Section 3.1 to detect breath and silence events before aligning notes with the score (Figure 3). The alignment is done by finding the minimum cumulative cost in DTW (see Section 3.2). There are sometimes ambiguous note boundaries among some consecutive notes with the same pitch which are occasionally detected as one long note. In order to separate them for further manual adjustment, we automatically separate the long note into a number of matched notes in proportion to their lengths in the MIDI file. After alignment, we modify the "Note On" and the "Note Off" events for each note in MIDI files to its matched note onset and offset in the audio track. Experiments show that over 90% notes are aligned to the accurate positions and the rest are positioned at the approximate positions which are then adjusted manually. As a result, our alignment method significantly reduces the time and effort required for annotation.

Table 1: Subjects in pilot survey

Grade	Age	CI students	NH students
Primary 2	7 - 9	9	9
Primary 1	6 - 8	8	8
Kindergarten	5 - 7	13	13

4. MOGAT DESIGN

In order to understand the disadvantages of children with CI and to further analyze their musical needs, we performed a pilot study to compare the music abilities of children with CI and normal hearing (NH) children. This led to several design principles which informed the design of three games and our web service for teachers.

4.1 Pilot study

All children in the study (see Table 1) were from Canossian School and its affiliated school for the hearing impaired. The study was approved by the school principal and carried out during their normal school time.

4.1.1 Procedure

We adopted their regular music assessment exercises as our assessment protocol, which were built into our iPad application beforehand to easily and quickly collect their answers and recordings. The app contains three modules for testing users' abilities in pitch perception (10 questions), rhythm perception (10 questions), and intonation accuracy (11 questions). In three modules, piano sound is used for audio playback. This is because the music educators use piano to teach music for children with CI, and thus they are more familiar with this instrument than other instruments.

In the pitch perception module, subjects first hear two notes played by a piano sound and then they are asked to identify if they are the same or different by choosing one of two buttons displayed on the touch screen. The maximum interval between two notes is a fifth.

In the rhythm perception module, subjects first hear a two-bar rhythmic phrase synthesized by a piano sound and then they need to tap on the touch screen to reproduce all the note onsets. The app records the time stamps of user tappings and saves them into a log file.

The singing module is relatively more complicated than the other two. During the example demonstration, the app plays a synthesized two-bar melodic phrase using a piano sound. Immediately after the melody finishes playing, the app displays countdown ("3, 2, 1, Go") label on the screen, and subjects are then to sing "La" for each note without hearing the synthesized piano sound. Recording stops automatically once the animation finishes. Subjects are always

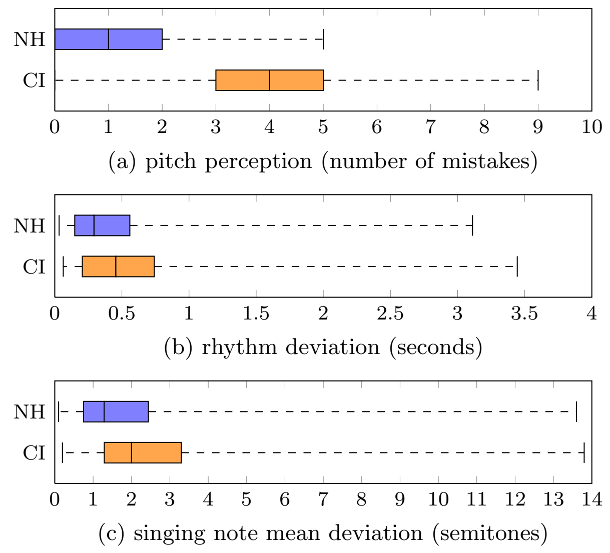


Figure 4: Three metrics used for evaluating music perception and singing ability in the two subject groups. Each box plot shows the lower limit, lower quartile, median, upper quartile, and the upper limit of the data. Lower numbers indicate fewer mistakes.

shown the visual animation of the note sequence in a piano roll format during demonstration and recording, where the visual note sequence moves from right to the left piano keyboard, indicating the active note currently being played.

4.1.2 Analysis

Our hypotheses were that the results of pitch perception, rhythm perception and intonation accuracy would be worse for children with CI than NH children. We used unpaired t-test to test the difference of means between the control and the test groups. We took the statistical significance p-value equal to 0.01. Results are presented in Figure 4.

Pitch perception: This test presented children with a choice of two pitch intervals; children with CI chose the incorrect option a significantly higher number of times than NH children with $t(56) = 5.935, p = 7 \times 10^{-8}$. Since this test is asking for a choice of between two options, random guessing should result in a score of 5/10.

Rhythm perception: We define our rhythm perception metric as the mean absolute deviation between user taps and reference sequence after aligning the first detected onset to the first reference onset. CI and NH children did not differ significantly in this measure with $t(161) = 2.392, p = 0.018$. On the basis of this experiment we reject the null hypothesis and conclude that the rhythm perception of children with CI has not shown to be worse than children with normal hearing, which agrees with previous research [30, 18].

Intonation accuracy: We used the mean note deviation calculated by the singing evaluator in Section 3.2 to represent intonation accuracy. Children with CI demonstrated significantly larger mean note deviation than NH children with $t(661) = 4.039, p = 6.1 \times 10^{-5}$. Our data also revealed large individual variability in both NH children and children with CI.

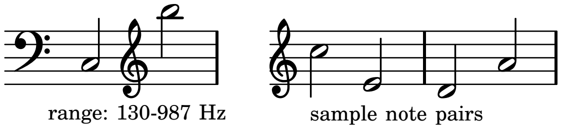


Figure 5: Range of *Higher Lower*, and minimum difference between the pairs of notes used for CI children.

4.2 Design Goals

Based on our pilot study, we established four goals:

- Improve students' pitch perception skills by determining the relative pitch difference.
- Improve students' pitch production skills with appropriate use of voice and breath support in singing.
- The interfaces should be easy and intuitive to use, and the games should be fun and interesting to play.
- Supply a remote centralized administration allowing teachers to easily monitor and personalize habilitation.

The emphasis on pitch perception and production arises from the deficiencies found in the pilot study. Our music therapist required that the system support breath control in singing by testing children's ability to sustain the correct pitch for a certain duration.

On the basis of our pilot study and design goals, we created three games for students and a web service for teachers.

4.3 Game Design

Children with CI are a special user group, in terms of not only their hearing disabilities but also their cognitive ability. In order to achieve an intuitive design, we organized a multi-discipline research team and adopted the relevant design methodology in HCI (e.g., participatory design, iterative design, and user-centered design). We actively involved all stakeholders in the design process including music teachers, music therapists, CI children, and the school principal.

4.3.1 Higher Lower (pitch perception)

The game begins by playing two notes with a piano sound because of their familiarity with this instrument. The student then indicates whether the first note is higher than the second, or vice versa (Figure 1a). The total range of the pitches is shown in Figure 5, while the minimum difference between two pitches can be altered according to the players' ability. Players can replay the sound by pressing a button on the interface.

In our user study, we followed the advice of the children's music educator and chose 7 semitones as the minimum difference between two notes. To a skilled musician this may appear to be a rather easy game, but test results show that some children with CI find this quite challenging.

4.3.2 Vocal Matcher (singing individual pitches)

In this game, players listen to a note, and then they sing the pitch and sustain it for 1 second until the note bar is filled up (Figure 1b). In order for a player to practice the pitches matching his/her vocal pitch range, the game will search for the player's pitch range in the beginning of the game. When the pitch range is found, the program will log the data into the device, and will randomly select notes from this range for playing in the future. We provide automatic note checking for players' pitch in comparison with the reference.

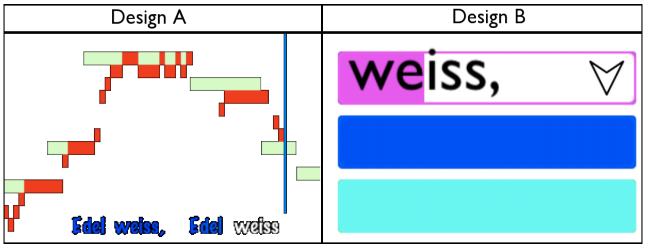


Figure 6: The comparison of two game designs. In Design A, the reference MIDI is in green; the users' pitch contour is in red. In Design B, the downward/upward arrow on the right means that users' pitch is higher/lower than the reference and they should lower/increase their pitch.

When players are singing the correct pitch, its note bar will gradually fill up until they sustain that pitch for the note duration; when players are singing the wrong pitch, on the right of the note bar appears an arrow to indicate whether players should increase or decrease their pitch. Players can replay or skip the sound if they deem it too difficult to sing.

Following the advice of the children's music teacher, any pitch within 3 semitones of the correct one will be accepted.

4.3.3 *Ladder Singer (singing a melody)*

To design this game, we began by studying the common features of existing Karaoke games and found that most of them display one row of pitch contour and another row of lyrics in parallel, with animation highlighting the relevant portion in time with the playback of audio track. In order to investigate the usability of this kind of games in our scenario, we implemented a Karaoke game incorporating the basic mechanism of Karaoke games on mobile devices shown in Design A of Figure 6. However, feedback from special educators and users suggested the following 2 problems:

1. Pitch correction: Design A uses a vertical bar to indicate the current singing progress, under which players' pitch is displayed. However, it does not check the correctness of players' pitch, and thus players have to rely on the relative positions of their past pitch contour compared with the reference to do pitch adjustment. This adds additional cognitive burden on players during the game play.

2. Lyrics reading: Due to problem 1, lyrics are difficult to read as players' vision is already overloaded with information from reading pitch feedback, especially when lyrics are displayed as merely one sentence at the bottom of the screen while the pitch contour occupies the most screen space.

These are not serious problems for adults and children with normal hearing, singing well-known songs for entertainment, as they likely have the lyrics memorized and have good pitch detection and singing ability. However, our target group is younger children with CI, so a different design is necessary.

Design B in Figure 6 shows the interface design of *Ladder Singer*. We used a "color ladder", a simple metaphor used in their math textbooks, where each note is one "stair" in the ladder and notes are sorted ascendingly from bottom to top on the screen, which is a consistent high-low concept adopted in *Higher Lower*. In order to guide players to sing each note, we first empty its corresponding note bar. To solve problem 1, automatic note checking as in *Vocal Matcher* is provided to help players to adjust their pitch. To solve problem 2, we display each corresponding word inside the note bar. In this way, all the necessary information including note, duration, lyrics, and hints for correction are seamlessly integrated into the narrow space of a note bar. Furthermore, we break down the whole song into phrases so that students can learn the song phrase by phrase. As a result, it is easier for players to concentrate on both lyrics and note information simultaneously. Figure 7 shows the internal game state of *Ladder Singer*. There are two stages in the game: listening and singing. Players begin the game by selecting a musical phrase to listen to with note animation. During singing, the game checks the correctness of players singing pitch and provides the relevant feedback. Meanwhile, it will automatically proceed to the next note when players finish the current one, and return to the listening stage when they finish the phrase. We will validate the usability of Design B compared with Design A in Section 6.3.

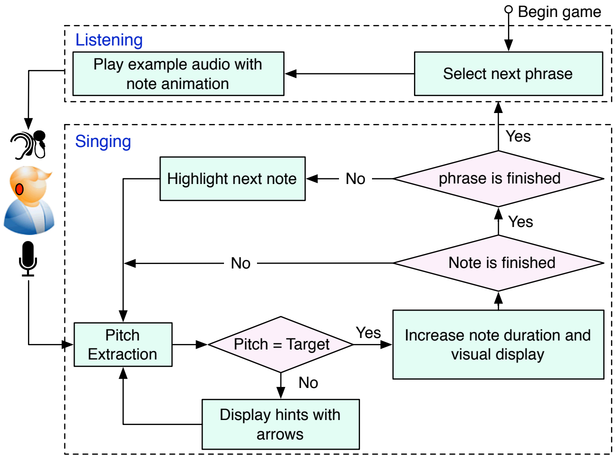


Figure 7: Internal game-state of *Ladder Singer*

4.4 Cloud Computing Service

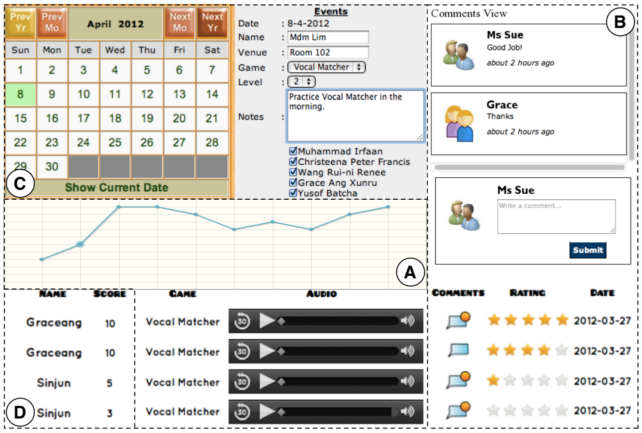
In order to facilitate special music teachers' communication with children with CI, we built a cloud computing service for this community with the following main features:

A. Individual progress tracking: Teachers can view a graphical visualization of a student's scores over a daily, weekly, or monthly period. Furthermore, they can listen to students' singing recorded in games to pinpoint students' problems.

B. Enabling reciprocal interaction: Teachers can examine students' singing, give overall rating, and post comments. The rating displayed is the collective average of all teachers' ratings. The elements of social media interaction in the website grant more flexibility and accessibility to students and teachers for communicating with each other.

C. Events planning: Teachers can plan students' habilitation in an event calendar by which they can decide time, location, game, and difficulty for a student to play according to his/her performance.

D. Leader board: Students and teachers can check the score leader board within one day, a week, and a month, which can introduce a moderate amount of competition to increase motivation to practice.



Events

Date : 8-4-2012
Name : Mdm Lim
Venue : Room 102
Game : Vocal Matcher
Level : 2
Notes : Practice Vocal Matcher in the morning.

☑ Muhammad Irfaan
☑ Christeena Peter Francis
☑ Wang Rui-ni Renee
☑ Grace Ang Xunru
☑ Yusof Batcha

Comments View

Ms Sue
Good Job!
about 2 hours ago

Grace
Thanks
about 2 hours ago

Ms Sue
Write a comment...

Name	Score	Game	Audio	Comments	Rating	Date
Graceang	10	Vocal Matcher			★★★★★	2012-03-27
Graceang	10	Vocal Matcher			★★★★☆	2012-03-27
Sinjun	5	Vocal Matcher			★☆☆☆☆	2012-03-27
Sinjun	3	Vocal Matcher			☆☆☆☆☆	2012-03-27

Figure 8: A montage of teacher view

5. IMPLEMENTATION

5.1 Games

The games were developed using Objective-C in iOS SDK. We adopted the cocos2d [2] framework for the graphics and animation rendering. The score layout layer, which parses midi and lyrics files for scheduling the note animation and displaying lyrics respectively, was implemented using the C++ library libjdkmidi [6]. In order to permit widespread deployment, we targeted older 2^{nd}-generation iPod Touches to reduce the devices' cost for children's parents. As a result, we must seriously consider the less advanced computational power of these devices during the development. We compared the computational cost and latency of two audio frameworks, Audio Unit (AU) and Audio Queue Services (AQS) in the CoreAudio framework.

To investigate the performance and latency of AQS and AU in our application, we implemented both in our application and performed the same tasks (i.e., pitch estimation and audio recording) in their callback functions. The sample rate was 24000 Hz and we used three buffers of 1024 samples ($\approx$ 43ms). We measured the average CPU usage of the application and average latency of callback functions using Activity Monitor of instruments in Xcode 4, and present the results in Table 2. We found that the CPU usage and latency were much better with AQS. It should be noted that AU allows highest level of control and simultaneous audio I/O with low latency, and other high-level audio frameworks including AQS are built upon it. However, the render callback in AU lives on a real-time priority thread on which subsequent render calls arrive asynchronously [1], which has a very strict performance requirement.

If the callback performs a time-consuming task such as autocorrelation-based YIN algorithm ($O(n^2)$), it will lose its advantage of low latency and high performance and even get gaps in the sound. Rather, AQS use the buffer queue whose callback function gets called whenever its audio buffers come in. It allows computing with less strict time constraints so that we can synchronously perform an accurate but time-consuming algorithm as well as writing audio buffers to the file system.

With iOS vDSP framework, the spectral domain YIN algorithm [14] ($O(n\log(n))$) can fully optimize the computational cost for both AU and AQS (see Table 2). In order to handle both audio recording and real-time audio processing simultaneously, AQS is used finally in conjunction with Audio File Services which is for writing audio buffers into recording files.

Table 2: Performance comparison of *Vocal Matcher* between AQS and AU in our app on an iPod Touch (2^{nd}-generation).

	YIN		YIN FFT	
	AQS	AU	AQS	AU
CPU Usage (%)	28.49	53.47	27.57	37.06
Latency (ms)	0.021	0.624	N/A	0.022

In the callback function of AQS, we defined a fixed threshold to exclude frames of silence or irrelevant background noise and to preserve those representing potential singing voice. The volume was calculated by getting the decibel value from the root mean square of the audio signal within the frame. We empirically chose 30 dB as the threshold. We implemented spectral domain YIN algorithm with vDSP framework to do pitch detection for any frames which were not deemed to be silence.

5.2 Cloud Service

The server backend of the web service used PHP server to handle the HTTP requests and MySQL for the database. For the web front-end, we use HTML5, JavaScript, CSS3, and jQuery, a commonly used JavaScript framework. Furthermore, the web service adopts the RESTful architecture. The metadata (including scores, user ids, and recordings in games) are automatically sent to the web server via JSON, which are then parsed and stored into our database. This allows minimal end user actions, which makes the data transmission process seamless and offers a better user experience.

6. USER STUDY

15 students with CI were selected at random from a Canossian School (ages 6 to 10) for the hearing impaired. Their average hearing age after implantation is 4 years and 10 months. The evaluation approved by the school principal was carried out in their daily school time.

MOGAT was installed on 15 2^{nd}-generation iPod Touches which do not have built-in microphones. We plugged an audio adapter containing a microphone into each iPod Touch to enable its audio input. We connected students' cochlear devices directly with audio adapters via their own personal audio cables, which provided better sound quality than the speakers on iPod Touches.

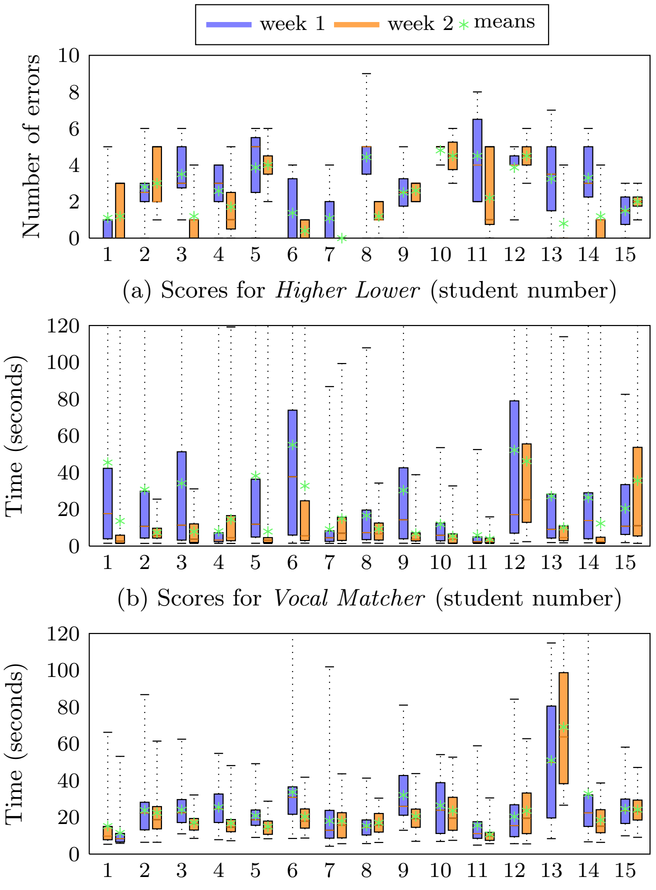
The first three games were evaluated in the order of *Higher Lower*, *Vocal Matcher* and *Ladder Singer*. Students were asked to play each game once everyday for two weeks under their teachers' supervision.

6.1 User Performance Evaluation

We tracked the students' skill in the games.

Higher Lower (*HL*): To match the initial pilot study, we tracked the number of incorrect answers to measure their pitch perception skills.

Vocal Matcher (*VM*): Since students should sing a specific pitch for 1 second before the game progresses to the next exercise, we consider the total time a student spends on each exercise to indicate how quickly they can reach the correct pitch. A student's intonation skill is therefore in-



(a) Scores for *Higher Lower* (student number)

(b) Scores for *Vocal Matcher* (student number)

(c) Scores for *Ladder Singer* (student number)

Figure 9: Evolution of students' scores during the user study: Students' scores in the first week are compared to their scores in the second week in three games. Lower numbers indicate fewer mistakes, i.e., better proficiency.

versely proportional to the time. We excluded those recordings with extremely long duration ($\geq$ 120 seconds) when students were not familiar with the interface.

Ladder Singer (*LS*): A music phrase is a combination of multiple individual notes. We consider the total time a student spends on completing all the notes within the phrase to measure their pitch production skills in singing the song of Edelweiss.

Figure 9 shows individual students' score in the first and the second weeks, while Table 3 shows the one-way ANOVA test comparing scores from two weeks for those students whose scores improved between the two weeks. The mean error in *HL* decreased significantly from 2.82 to 1.98 ($F = 9.892, p = 0.002$). The mean time for completing single pitch in *VM* decreased significantly from 25.99s to 13.86s ($F = 36.49, p < 0.001$). The mean time for completing single phrase in *LS* decreased significantly from 24.33s to 18.97s ($F = 21.23, p < 0.001$). Furthermore, none of these students, who did worse in the 2nd week (N/A in Table 3), showed statistically significant degradation in score. The improvement among subjects was also highly variable. In *HL*, there were totally 9 students who have improved their score within two weeks, and 3 out of these 9 students (S3, S7, and S8) had statistically significant performance improvement ($p < 0.05$). In *VM*, there were totally 12 students who improved their skills in two weeks, and 7 out of 12 children (S2, S3, S5, S8, S9, S10, and S13) children had statistically significant improvement ($p < 0.05$). In *LS*, there were totally 12 students who had improvement, and 6 out of 12 children (S3, S5, S6, S9, S11, and S14) had statistically significant improvement ($p < 0.05$). In sum, 5 out of 15 children (S3, S6, S10, S11, and S14) had improvement in all three categories and 1 child (S3) had significant improvement among all categories.

Table 3: The one-way ANOVA test results (F-statistic and p-value) on comparing each student's scores in the first week and his/hers in the second week ($^*p < 0.05$,$^{**}p < 0.01$). N/A means no improvement in their second-week scores compared to their first-week scores.

	Higher Lower		*Vocal Matcher*		*Ladder Singer*	
	The factor of weeks					
	F	p	F	p	F	p
S1	N/A		2.479	0.126	1.185	0.287
S2	N/A		**12.561**	**0.001****	0.056	0.814
S3	**18.778**	**0.012***	**15.161**	$\mathbf{10^{-4}}$******	**10.862**	**0.002****
S4	0.641	0.454	N/A		0.416	0.53
S5	N/A		**10.757**	**0.002***	**4.651**	**0.038***
S6	3.447	0.137	2.378	0.134	**8.402**	**0.007****
S7	**10**	**0.025***	N/A		0.578	0.451
S8	**46**	**0.002***	**5.497**	**0.023***	N/A	
S9	N/A		**23.506**	$\mathbf{10^{-5}}$******	**13.037**	**0.001****
S10	0.273	0.638	**7.368**	**0.01***	0.385	0.55
S11	0.388	0.578	1.942	0.171	**5.664**	**0.024***
S12	N/A		0.198	0.659	N/A	
S13	3.103	0.153	**6.566**	**0.014***	N/A	
S14	4.966	0.09	1.789	0.189	**4.785**	**0.037***
S15	N/A		N/A		0.004	0.947

6.2 User Experience

Our user experience evaluation is based on the following three criteria: *naturalness*, *enjoyment*, and *motivation*. At the end of the first week, the students were asked to rate questions using a five-point likert scale.

- *I feel the game is easy to play.*
- *I enjoyed playing this game.*
- *I would play this game for fun if I had it.*

Figure 10 shows the results averaged over all the students. In terms of naturalness, *HL* is the most intuitive one to play with. *VM* is a simplified version of *LS*, and it has the practice and carryover effect on *LS*. Therefore, although *LS* is relatively hard to play with, the naturalness score of

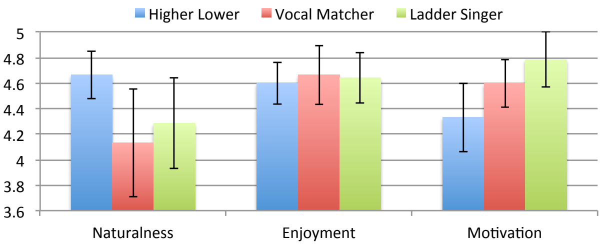


Figure 10: Results of user experience

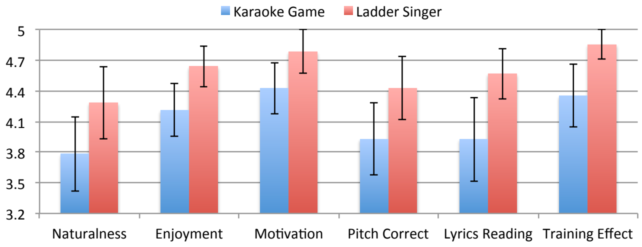


Figure 11: Karaoke Game v.s. *Ladder Singer*

LS was slightly higher than *VM*. Furthermore, the students enjoyed playing three games to a similar extent. Last but not least, the students expressed strong motivation to play all three games for fun in the future (ratings are all over 4.3), especially *LS* (rating is 4.8).

6.3 Ladder Singer v.s. Karaoke Game

We organized a comparison study between Karaoke Game (Design A) and *LS* (Design B) mentioned in Section 6 before the evaluation of *VM*. During the process, we randomly chose the order of two games to counterbalance the practice and carryover effects. After each game, users were asked to rate the games using three additional factors (*pitch correction, lyrics reading, training effect*) in addition to the criteria of user experience.

- *I can correct my pitch based on the game feedback.*
- *I can follow the lyrics during singing.*
- *The game can help me with learning this song.*

Figure 11 shows the results averaged across 15 users. We can see that *LS* was ranked higher than Karaoke Game in all aspects. During the experiment, we observed that most of their pitch contours displayed in Karaoke Game were quite flat, which illustrates that they could not easily correct their pitch according to the reference. Nevertheless, it is impossible in *LS*, where they can proceed to the next note if and only if users perform correctly.

6.4 Web Service Evaluation

Our purpose is to evaluate whether our cloud-based web service can enhance special teachers in supporting students' musical habilitation. We recruited two special musical educators in this study. First, they received a sheet of instructions demonstrating the interface and features of the website. Teachers were then asked to use it by themselves without facilitators' intervention. Finally, teachers answered a questionnaire related to the usability of the web service.

Overall, the participants responded that MOGAT web service was fairly easy to use, all giving it a 4 on a scale of 1 (extremely difficult to use) to 5 (extremely easy to use). Participants also expressed satisfaction and willingness to use the website to support children's habilitation. However, they requested that we improve the documentation for more advanced features such as setting up an event. They asked us to describe the ratings detailedly in the website so that different teachers can keep a consistent rating criteria.

7. DISCUSSION

While most children benefited from MOGAT-enhanced auditory training, there was large individual variability in the amount of improvement among subjects. Many factors may affect the outcomes. For example, training materials (i.e., sound stimuli), training difficulties, and training duration. Although these independent variables were controlled in our user study, it is most desirable for us to deeply understand how they affect individual's performance. It could help us to design the most suitable individualized training protocols. We will investigate this problem in the future.

We would like to mention that teaching hearing-impaired children singing allows them to not only just sing better but also do what normal children can do, which can enhance their confidence and self-esteem. Moreover, singing can help them to learn language and improve their speech intelligibility. Thus we focus on this special user group and it is almost impossible for hearing-impaired children to achieve the same skills by merely playing any other video games.

It is important to emphasize that throughout the project we aim to train users' relative pitch production ability rather than their absolute pitch ability because different singer has his/her own vocal pitch range. The singing evaluator was thus developed to transpose the pitch to fit their vocal range.

The limitation of the work is that the audio-MIDI alignment algorithm is not ideal and requires manual adjustment. Nevertheless, the statistical model can be used to solve this problem. First, the DTW-based algorithm can help to build the training dataset, saving the work in human annotation. We can then train a Hidden Markov Model (HMM) on the dataset to do the alignment. It will become our future work.

8. CONCLUSION

We have presented the design, development, and deployment of MOGAT, the first integrated training system to assist pre-lingually deafened children with music habilitation. After investigating their problems in pitch-based metrics in comparison with normal hearing children from our pilot study, we developed three music games with off-the-shelf mobile devices which are designed specifically for their musical needs and cognitive abilities. In order to maximize the limited teaching resources, we developed a cloud-based web application to connect special music teachers with children with CI to provide them with administrative and teaching support. A comprehensive user study has demonstrated the effectiveness and efficiency of MOGAT in enhancing children' musical habilitation as well as teachers' teaching and management tasks, and its large potential in enhancing pitch perception and production of children with CI.

ACKNOWLEDGEMENTS

We are grateful to Ms. Terry Theseira and her colleagues and students from Canossian School for assistance and participation in the user studies. We would like to thank Graham Percival for his kindly support in paper proofreading, and thank Kenny Tan and Yong Shen Wong for their system implementation and design. We appreciate all our reviewers for their excellent comments and feedback.

9. REFERENCES

[1] Audio unit hosting guide for ios. `http://developer.apple.com/library/ios/`.

[2] Cocos2d. `http://www.cocos2d-iphone.org/` .

[3] Glee karaoke. `http://itunes.apple.com/sg/app/glee-karaoke/id360736774?mt=8`.

[4] Karaoke revolution. `http://www.konami.com/games/karaoke-revolution/`.

[5] Karaokeparty. http://www.karaokeparty.com/.

[6] libjdkmidi. http://opensource.jdkoftinoff.com/jdks/docs/libjdkmidi/.

[7] Noteflight. http://www.noteflight.com.

[8] Sound express auditory training, tigerspeech technology. http://www.tigerspeech.com/tst_soundex.html.

[9] Theta music trainer. http://trainer.thetamusic.com/.

[10] Voice tutor. http://voicetutorapp.com/.

[11] Cochlear implants for children, July 2008. http://www.cochlear.com/au/hearing-loss-teatments/cochlear-implants-children.

[12] National institute on deafness and other communication disorders, cochlear implants, March 2011. http://www.nidcd.nih.gov/health/hearing/pages/coch.aspx.

[13] S. Abdi, M. Khalessi, M. Khorsandi, and B. Gholami. Introducing music as a means of habilitation for children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, 59:105–113, 2001.

[14] P. Brossier. *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Queen Mary University of London, UK, August 2006.

[15] P. A. Busby, S. A. Roberts, Y. C. Tong, and G. M. Clark. Results of speech perception and speech production training for three prelingually deaf patients using a multiple-electrode cochlear implant. *Br J Audiol*, 25(5):291–302, 1991.

[16] P. W. Dawson and G. M. Clark. Changes in synthetic and natural vowel perception after specific training for congenitally deafened patients using a multichannel cochlear implant. *Ear and Hearing*, 18:488–501, 1997.

[17] A. De Cheveigne and H. Kawahara. Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 2002.

[18] W. R. Drennan and J. T. Rubinstein. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *J. of Rehabilitation Research & Development*, 45(5):779–790, 2008.

[19] G. T. Ealy. Of ear trumpets and a resonance plate: Early hearing aids and beethoven's hearing perception. *19th-Century Music*, 17(3):pp. 262–273, 1994.

[20] Q.-J. Fu, J. Galvin, X. Wang, and G. Nogaki. Moderate auditory training can improve speech performance of adult cochlear implant patients. *Acoustical Society of America*, 6(3):106–111, 2005.

[21] Q.-J. Fu and J. J. Galvin. Computer-assisted speech training for cochlear implant patients: Feasibility, outcomes, and future directions. *Seminar in Hearing*, 28:142–150, 2007.

[22] J. J. Galvin, Q.-J. Fu, and R. V. Shannon. Melodic contour identification and music perception by cochlear implant users. *The Neuroscience and Music III - Disorders and Plasticity*, 1169:518–533, 2009.

[23] K. Gfeller, S. Witt, M. Adamek, M. Mehr, J. Rogers, J. Stordahl, and S. Ringgenberg. Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *Journal of the American Academy of Audiology*, 13(3):132–145, 2002.

[24] K. Gfeller, S. Witt, J. Stordahl, M. Mehr, and G. Woodworth. The effects of training on melody recognition and appraisal by adult cochlear implant recipients. *Journal of the Academy of Rehabilitative Audiology*, 33:115–138, 2000.

[25] K. Gfeller, S. A. Witt, K.-H. Kim, M. Adamek, and D. Coffman. Preliminary report of a computerized music training program for adult cochlear implant recipients. *Journal of the Academy of Rehabilitative Audiology*, 32:11–27, 1999.

[26] E. Glennie. Hearing essay, 1993. http://www.evelyn.co.uk/Resources/Essays/Hearing%20Essay.pdf.

[27] V. S. Hagedorn. Musical learning for hearing impaired children. *Research perspective in music education*, (3):13–17, 1992.

[28] P. Harrison. The effects of deafness on musical composition. *Journal of the Royal Society of Medicine*, 81:598–601, 1988.

[29] O. Mayor, J. Bonada, and A. Loscos. The singing tutor: Expression categorization and segmentation of the singing voice. In *Proceedings of the AES 121st Convention*, 2006.

[30] H. J. McDermott. Music perception with cochlear implants: A review. *Trends In Amplification*, 8:49–82, 2004.

[31] C. Oshima, K. Nishimoto, and M. Suzuki. Family ensemble: a collaborative musical edutainment system for children and parents. In *ACM Multimedia*, 2004.

[32] C. C. Peterson. Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *J Child Psychol Psychiatry*, 45(6):1096–1106, 2004.

[33] A. L. Quittner, P. Leibach, and K. Marciel. The impact of cochlear implants on young deaf children: new methods to assess cognitive and behavioral development. *Arch Otolaryngol Head Neck Surg*, 2004.

[34] I. Simon, D. Morris, and S. Basu. Mysong: automatic accompaniment generation for vocal melodies. In *ACM CHI'08*.

[35] D. L. Sorkin and N. Caleffe-Schenck. Cochlear implant rehabilitation. Online, June 2008. http://www.cochlear.com/files/assets/ci_rehab_not_just_for_kids.pdf.

[36] R. Sweetow and C. V. Palmer. Efficacy of individual auditory training in adults: A systematic review of the evidence. *Journal of the American Academy of Audiology*, 16(7):494–504, June 2005.

[37] E. A. Tobey, A. E. Geers, C. Brenner, D. Altuna, and G. Gabbert. Factors associated with development of speech production skills in children implanted by age five. *Ear and Hearing*, 24:36S–45S, February 2003.

[38] J.-L. Wu, H.-M. Yang, Y.-H. Lin, and Q.-J. Fu. Effects of computer-assisted speech training on mandarin-speaking hearing-impaired children. *Audiology and Neurotology*, 12:307–312, 2007.

[39] Y. Zhou, G. Percival, X. Wang, Y. Wang, and S. Zhao. Mogclass: a collaborative system of mobile devices for classroom music education. In *ACM Multimedia '10*.

[40] Y. Zhou, G. Percival, X. Wang, Y. Wang, and S. Zhao. Mogclass: evaluation of a collaborative system of mobile devices for classroom music education of young children. In *ACM CHI'11*.