TalkLIME: Mobile System Intervention to Improve Parent-Child Interaction for Children with Language Delay

Seokwoo Song, Seungho Kim, John Kim
School of Computing
KAIST
Daejeon, Korea
{sukwoo24, shkimj, jjk12}@kaist.ac.kr

Wonjeong Park, Dongsun Yim
Department of Communication Disorders
Ewha Womans University
Seoul, Korea
{thisisforwj, sunyim}@ewha.ac.kr

ABSTRACT
Parent-training is commonly used to support intervention of children with language delay. Unfortunately, parents find it difficult to apply the training to their child in everyday life and often give up on their parent-child interaction. In this work, we propose and evaluate TalkLIME – a mobile system that provides real-time feedback to improve the parent-child interaction and reinforce parent-training intervention. We first conduct a survey to understand parents' feedback preference for the mobile system and determine that a non-invasive feedback using the mobile phones screen is preferable. TalkLIME was developed to provide real-time feedback through the mobile phone screen while also providing motivation to the parents to consistently continue parent-child interaction through both short-term and long-term goals. A six-weeks user study was conducted with eight parents and their children with language delay. Our results show that the experimental group who used TalkLIME showed a significant improvement in the child's initiation ratio, an important metric in the language development of children.

Author Keywords
Language delay; language therapy; smartphone intervention; everyday treatment ; children ; parental education

ACM Classification Keywords
J.3 Life and Medical Sciences: Health; C.3 SpecialPurpose and Application-based Systems: Real-time and embedded systems; H.5.3 Group and Organization Interfaces: Computer-supported cooperative work

INTRODUCTION
Communication is an important aspect of our everyday lives and has a significant impact on our social relationships and academic achievements [21]. Speech-language development begins during early childhood, and development (or the lack of development) during childhood can have a lasting impact. As a result, properly identifying language development delay at a young age is important. Once a language delay is identified, it is necessary not only for the child to undergo therapy but parent-training is as important since child's environment has a significant impact on child's language development. In particular, the speaking habit of the parents (or caregivers) has a significant influence on the child. Over decades of research, Speech-language Pathologists (SLPs) have shown that effective parent participation or involvement is critical in the treatment of language delay in children [17, 5].

While parent-training can have the positive impact on the children, it is often difficult for parents to apply what they learned within their home environment. The SLPs often encourage the parents to spend 10-15 minutes [1] each day with their children to apply what the parents have learned [17, 5]. While 10-15 minutes is not a lot of time, unfortunately, it can feel like a long time for the parents for various reasons. It is also often a challenge for the parents to know whether they are correctly practicing what they have learned through parent-training. In this work, we propose and evaluate TalkLIME [2] – a mobile system that provides real-time feedback to the parent to improve the parent-child interactions with children that have the language delay. Based on the preliminary studies that we conducted, the objectives of our proposed system were the following:

1. Improve the parent-child interactions through a real-time feedback that is not disruptive.
2. Increase the parent's motivation to continuously engage in parent-child interactions.
3. Provide feedbacks on child's language development.

To provide a feedback that was not disruptive, we designed a mobile system where the phone screen was used to provide real-time feedbacks to the parents. We also simplified the system to make it easier for the parent to use – e.g., instead of requiring multiple devices [14], TalkLIME only required a single smartphone device. To motivate the parents to continuously use the system and engage in parent-child interaction sessions, we provided a timer to set a short-term goal for one session of parent-child interaction. In addition, TalkLIME provided historical statistics on how the child's language development changed over time. By providing a comparison to statistics of children without language delay, TalkLIME provided a long-term goal for the parents.

[1]We refer to this as *parent-child interaction session* in this work.
[2]TalkLIME stands for Talk **L**ess **I**s **M**or**E** as it is often better for parents to talk less to enable the children to talk more and have them initiate.

UbiComp '16, September 12-16, 2016, Heidelberg, Germany

DOI: http://dx.doi.org/10.1145/2971648.2971650

We conducted a 6-week user study with 8 parent-child participants and evaluated the impact of TalkLIME both quantitatively and qualitatively. The number of utterances is an important metric in child language development [17] but a more important metric is the *initiation* – i.e., the number of utterances that the child initiates. Our user study shows that TalkLIME resulted in significant improvement children's initiation ratio. We also provided qualitative analyses based on the post-study interview that showed TalkLIME helped the parents to understand parent-child conversation status through real-time feedbacks; TalkLIME also motivated the parents to continue parent-child interactions consistently by providing both short-term and long-term goals.

BACKGROUND AND RELATED WORK

Children with Language Delay and Intervention

Children with language delay are defined as children who do not acquire language as expected for their chronological age [15]. These children are slow in acquiring new words and also start to combine words into phrases later in their development compared to children with typical development [16]. Thus, children with language delay have difficulty in interacting with their parents and carrying out a conversation [4, 28, 25]. To identify children with language delays, standardized tests are commonly used by Speech-language Pathologists [24].

Early intervention in children with language delays can be performed in two ways [2, 15]. One approach is clinical-based intervention with SLPs and another method is a home-based intervention with the parents of children with language delay. Prior to home-based intervention, parent-training is done where strategies for promoting their child's language development and facilitating parent-child interaction is presented. It has been shown that intervention is more effective when home-based intervention is started at a young age for the child with a language delay [17]. While the guidelines provided by SLPs are often very clear, it is not necessarily easy to follow all of the guidelines for the parents and becomes a challenge. In this work, we propose a mobile system that provides intervention to help parents.

Device Support for Children with Language Disorders

Speech-language pathologists have begun to use "smart" devices for the early identification of language delay or to assist children with language delays have increased. Jeon et al. [18] explored an alternative AAC (Augmentative and Alternative Communication), with a robot to assist therapists in interventions with nonverbal children. Mohammed et al. [13] customized interactive speech-enabled games to help children who have language problems produce more intelligible speech with appropriate speech rate, typical pitch, and amplitude. LENA (Language ENvironment Analysis) [22] is a commercial device that can analyze the child's language environment influencing the child's language development. The device records the child's voice (and surrounding sound) for an entire day and then, analyzes the data off-line. While LENA can provide many benefits, it has some limitations. In particular, the cost of the system is relatively expensive (~$700) while our work leverages smartphones that are commonly available. In addition, LENA is a simple, passive device that only records the voice (and conversation) and do not provide any analysis itself or provide any feedback while TalkLIME that we propose enables real-time feedback during the parent-child interaction.

Children Care with Smartphone

There have been many recent works that leverage smartphones for infant/children care. Some examples include BiliCam to help monitor newborn jaundice at home [7], providing guidelines for caregivers of high-risk infants [23] and regulating smartphone usage among children [19]. Recently, researchers have explored the opportunities to leverage the technology to not only support parents but also provide intervention. Slovak [31] proposed a technology-based intervention to support parent-child interactions that could reinforce social-emotional skills learning at home. MOBERO [32] is a smartphone-based system to assist families of children with attention deficit hyperactivity disorder (ADHD) by encouraging the children to become more independent. TOBY [33] helps parents start early intervention for their children with autism to improve their child's abilities such as attention, memory, and recognition [33]. Pina et al. [27] presented a mobile application to help parents who have children with ADHD by detecting emotions of the parents and sending supportive messages and pictures to the parents. TalkLIME that we propose is similar as we also exploit the mobile system to provide intervention to the parents; however, the focus of this work is on parent-child interaction for children with language delay.

This work shares some similarity to TalkBetter [14] which also motivated a need for a mobile system to provide feedback to the parents of children with language delay. While TalkBetter provided a survey-driven study (of both SLPs and parents) to motivate the need for such system, no user study was done to test the effectiveness. The TalkBetter system also assumed that the feedback to parent was provided through a separate earphone. While this approach can provide a clear feedback, it can be disruptive to the parent; in addition, the intervention can be bothersome when the feedbacks are triggered because of errors in the system. Another drawback of the TalkBetter is the relatively complex system for a parent who is not proficient in technology since it requires 4 devices – two smartphones, a bluetooth mic, and a bluetooth headset.

OVERALL STUDY PROCEDURE

A high-level overview of our study procedure and methods is summarized in Figure 1 and consists of 3 phases.

Phase 1: The first phase consisted of preliminary studies to design our TalkLIME system – consisting of two separate interview-driven parent studies and a data-collection study. The first preliminary study was to understand parents' difficulties in continuously practicing what they learned from parent-training at home. The second preliminary study focused on how feedback should be provided to the parent through the mobile system. These two preliminary studies impacted the TalkLIME mobile system design used in our user-study. In addition, we collected data to understand the

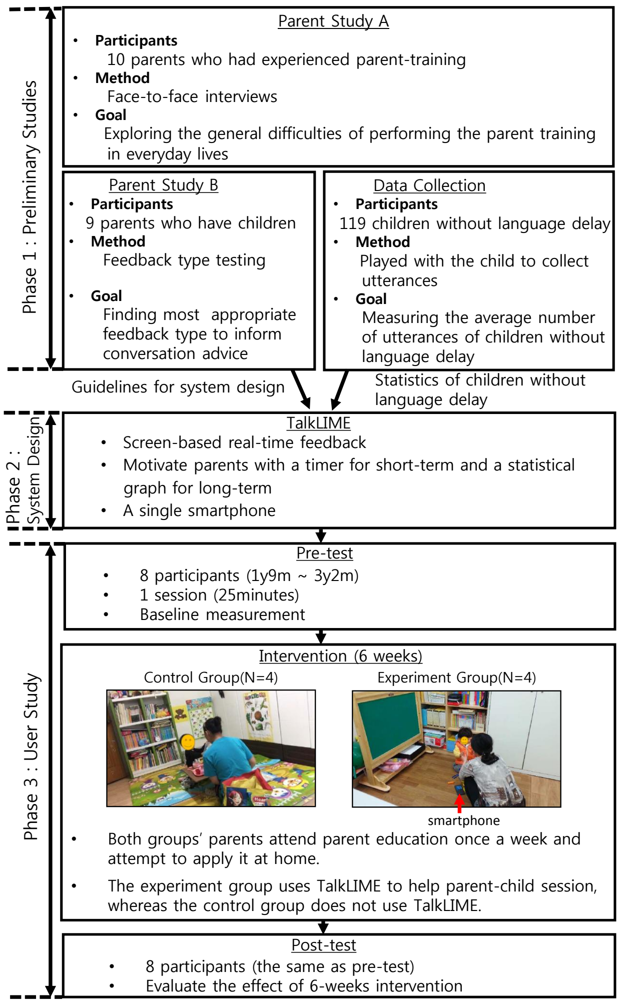


Figure 1. Overall study procedure.

average number of utterance for children (age: 2∼5) to provide some guidance for parents as part of the mobile system.

Phase 2: The TalkLIME system was implemented in this phase with the following key characteristics.
(1) Simplify system setup by building the system with a single smartphone.
(2) Provide less-disruptive screen-based real-time feedback for parents.
(3) Motivate parents by setting goals with a timer for as a short-term goal and a statistical graph for as a long-term goal.

Phase 3: A user study was conducted to evaluate the impact of TalkLIME system. The pre-test was first done to establish the baseline for both the children and the parents. During the intervention sessions of 6 weeks, the parents were divided into a control group and an experiment group. All parents received parent-training one session per week during the intervention, but only the parents in the experiment group were provided with the TalkLIME system. After finishing the intervention sessions, a post-test was done in the same manner with the pre-test to assess the changes in the children's language development from the pre-tests, we also conducted a post-test interview with the parents to qualitatively analyze the impact of TalkLIME.

Common problems from parents	TalkLIME
Difficult to know if parents are properly practicing their parent-training	provide real-time feedback [14]
Difficulty in continuously interacting with the children	provide a timer
Difficult to know if child is improving	provide data-driven interface and statistics
Less disruptive intervention	use-screen-based feedback
Increase accessibility	use a single phone-based system

Table 1. Summary of the different common problems faced by parents that we identified in our preliminary study and how TalkLIME addresses them

PRELIMINARY STUDY

Difficulties in Parent-child Interactions at Home

In our initial preliminary study, we interviewed 10 parents (10 females; ages: early 30 - early 40) who have children with language delay and have had experienced parent-training. Our goal was to better understand the difficulties in practicing parent-child interactions at home. The interviews were conducted face-to-face at their home or at language therapy center, with each interview lasting 30 minutes. The main questions asked were the following.

- Were parent-training helpful?
- Did you practice parent-child interactions continuously at home? If not, what were the difficulties?
- Did you ever stop or give up during a parent-child session at home? If so, why?

All of the interviews were recorded. All 10 parents unanimously answered that parent-training was very helpful in their parent-child interactions. However, by coding and iteratively clustering the interview data, we identified the two main difficulties in the parent-child interaction – the need for feedback to determine whether they are correctly practicing what they learned in parent-training and the motivation to continuously practice what they learned through the daily parent-child interaction. However, the first difficulty was knowing whether the parents were properly practicing what they learned during parent training – similar to the observations made in TalkBetter [14]. The parents often expressed that they made a lot of mistakes even though they were trained on how to interact with their children from the parent-training. The second difficulty was staying motivated during the parent-child interaction, both for the short-term (i.e., finishing a single parent-child interaction session) and long-term (i.e., continuing to carry out a parent-child interaction session). It has been shown that language improvements in a child often take a long period of time (e.g., several months) [35, 3]. Although most parents are initially eager to apply their parent-training, they gradually lose interest and motivation when they feel that their child's language development is not improving, even after putting in significant effort and time. "*I was initially really eager to improve my child's language development and tried many different parent-training. However, I soon got tired when I could not see any improvement in my child.*" In addition, while 10-15 minutes is not a lot of time, for parents tired from work or house chores, spending this 10-15 minutes can be tiresome – especially since children with language delay rarely provide any reaction to their parents during an interaction session – and results are some parents losing motivation and giving up: "*I tried to talk more and more to my child, but when my child continuously did not answer or react to me in any way, I was exhausted as I was already tired from other things.*".

Design Implications: Based on the preliminary study, while it is important to provide real-time feedback to the parents during a parent-child interaction, another equally, if not more, an important aspect was to prevent the parents from being discouraged during their parent-child interaction sessions at home. Even if the parent's interaction does not necessarily follow all of the guidelines from the parent-training, it is more important for the parents to continue to interact with the children [17] and thus, continuously motivating the parents was a significant priority for our proposed system.

Efficient Conversation Feedback

We performed a separate preliminary study to understand parent's preference in how feedback is provided to them during parent-child interaction session. 9 parents (9 females; ages: early 30 ~ early 40) who have a child (age: 2~5) participated in this study to understand the impact of different feedback. For this study, we asked the parents to interact with their children for approximately 30 minutes and the following three type of feedbacks were evaluated.

1. Post-session feedback: No real-time feedback is given but provided after the session by an SLP.
2. Sounds: With the assistance of an SLP, real-time feedback was provided using pre-recorded sound through an earpiece [14].
3. Screen: With the assistance of an SLP, real-time feedback was provided through a smartphone screen that was placed near the parent.

During the 30-minute session, each parent was exposed to all three type of feedback sequentially (i.e., each type of feedback for about 10 minutes) [3]. The SLP observed the parent-child interaction and gave the feedback when one of the following situations occurred: 1) the parent is speaking too fast, 2) the parent is not giving the child an opportunity to talk, 3) the parent is interrupting when the child speaks, 4) the parent is speaking in long sentences, and 5) the parents did not respond to the child quickly.

The feedback was provided manually by the SLP using a separate smartphone to "push" the feedback to the parent. For the sound feedback, when appropriate, SLP would push a button and a pre-recorded voice was played on the parent's earpiece. Similarly, for the screen feedback, the SLP would also provide the manual feedback and a text message would appear on the smartphone that was placed near the parent. For the post-session feedback, no real-time feedback was given but the feedbacks were summarized at the end by the SLP while the recorded conversation was replayed. In addition, the SLPs were asked to intentionally provide approximately 10-20% *incorrect* feedback – e.g., provide feedback "*wait for the child to speak*" when the child is actually speaking alone. Since any system is bound to have errors (e.g., difficulty in analyzing voice, noise, etc.) and can generate incorrect feedback, we wanted to understand the impact of such errors with the parents. After the tests, we asked the parents' comments on the strengths and the weaknesses of each feedback type. We summarize the comments on the three different type of feedbacks below.

[3]The order of feedback was different each participant in order to avoid any bias that might be created by the order of the feedbacks.

	Advantage	Disadvantage
Post-session feedback	Comprehensive feedback with voice recording	Not real-time feedback
Screen	Check feedbacks only when they want	Missed feedback
Sound	Immediate feedback	Very irritating on incorrect feedback

Table 2. Results for the different type of feedback mechanisms.

- **Post-session feedback**: Almost half of the parents (4 out of 9) answered that post-session feedback was most efficient because of its clarity since the SLP replayed the parent-child interaction and provided the feedback. The parents could realize their speaking habits: "*I realized how I am interacting with my child and how fast I spoke, and repetitive listening and efforts to alter bad habits would be helpful to interact with my child.*" Interestingly, the remaining parents found that post-session feedback would not be helpful since it is not a real-time feedback and felt it would not correct their habits: "*Feedbacks after the conversation was clear and not disruptive at all during the conversation, but I'm not sure that I can remember and apply the feedbacks directly in actual parent-child interaction.*"

- **Sound**: Many of the parents (5 of 9) this feedback was efficient since it provides direct (and immediate) feedback. However, all parents answered that the sound was most distracting, especially when incorrect feedback was provided. One parent said, "*I like the quick feedback provided through the earpiece, but when it sent obviously incorrect messages, I began to have doubts about the system and started to ignore the feedback.*"

- **Screen**: The screen feedback provides a trade-off between sound and post-session feedback. The screen allowed real-time feedback but it was less disruptive, compared with an earpiece feedback through sound. In addition, this feedback was most error-tolerant as well.

Based on the preliminary study (Table 2), we selected the screen interface to provide feedback since it was less disruptive than using sound and is more tolerant of errors. Since we were using a simple text message that appeared on the screen (and then, later disappeared), the feedback can actually be overlooked by the parent. Thus, in our TalkLIME implementation, we used a graphical interface that continuously changes to reflect the current parent-child interaction. This approach was also error-tolerant as well since explicit messages were not provided to the parents.

Language Developmental Comparison with Peer

In order to motivate parents, one of the important information is to provide an average statistics for the same age as their children for comparison. As a result, to gather the average number of utterances of children without language delay, we tested 119 (age 2 : N=29, age 3–5 : N=30) children without language delay. We visited children's homes or schools to gather spontaneous utterances from the children. One of the researchers and a child played together in a quiet place, and their conversation was recorded for approximately 30 minutes. To collect spontaneous utterances from the children, the researchers were trained to control their communication to focus on collecting spontaneous utterances from the children. The children's utterances were manually analyzed by two different researchers for verification. We used a one-way ANOVA to confirm the difference in the number of utterances by age. There was a significant difference between age groups (ρ <.05), and older children recorded a higher number of utterances. This average number of utterances was used in the data statistics provided to the parents in TalkLIME.

TALKLIME MOBILE-SYSTEM DESIGN

The goal of our TalkLIME system was to provide a less-disruptive feedback system that uses the smartphone screen. While the preliminary study used a simple text on the screen to provide the feedback, one limitation of this was that parents can miss the feedback if they are not aware of the screen. In addition, as discussed earlier, the mobile system cannot be 100% accurate in its analysis and feedback can be erroneously generated. To overcome these limitations while providing a simple interface, we design a screen that contains two balls – one ball representing the parent and the other ball representing the child. Based on the amount of utterance, turn-taking, and initiation, the screen interface will be changed to provide gradual feedback to the parent. The entire system was built on an Android smartphone.

Application Description

The mobile system application consists of two modes – the conversation mode and the result mode. As suggested by the parent-training, the parents are encouraged to spend 10-15 minutes of quality time with their children and to focus exclusively on applying what they learned during parent-training. After selecting a time for the parent-child interaction session, between 10-25 minutes, the system starts in the *conversation* mode (Figure 2(b)). When the selected minutes are complete, the *result* mode is shown (Figure 2(c)).

Conversation Mode

(a) The number of utterances – The size of the ball is proportional to the amount of utterance. The upper ball represents the child's utterances, and bottom ball represents the parent's utterance. The size of the ball changes based on the number of utterances over the past 3 minutes. This metric was empirically determined, given that larger values often result in a minimal change in the utterance values, whereas smaller values result in quick changes to the ball size. When the child's utterance is detected, the upper ball grows bigger and the bottom ball becomes smaller relatively, and vice versa. To count the number of utterances using the mobile system, an utterance is counted if there is silence for at least 0.3 seconds parent or child speaks. We experimentally chose 0.3 s via tests, but this metric is consistent with prior work [6] that analyzed the number of utterances. However, the actual number of utterances also depends on the sentence and the context of the word; thus, even if the speaker identification system is perfectly designed without any errors, using the mobile-system to determine the number of utterances will still result in some inaccuracies.

(b) Initiation Ratio – The brightness of the ball is used to represent the initiation utterance ratio of the total number of utterance that he/she spoke. Initiation utterances must be interpreted with meaning. For the same reason as utterance division, we predict one initiation utterance if someone speaks after there was silence for 3 seconds before. We experimentally chose 3 seconds or longer.

(c) Turn-taking – The glares in the ball determines who is speaking and shows the turn-taking that occurs between the parent and the child.

(d) Timer – When starting a parent-child interaction session, parents select the time length. The timer bar shows the amount of time remaining to motivates the parents by setting a short-term goal.

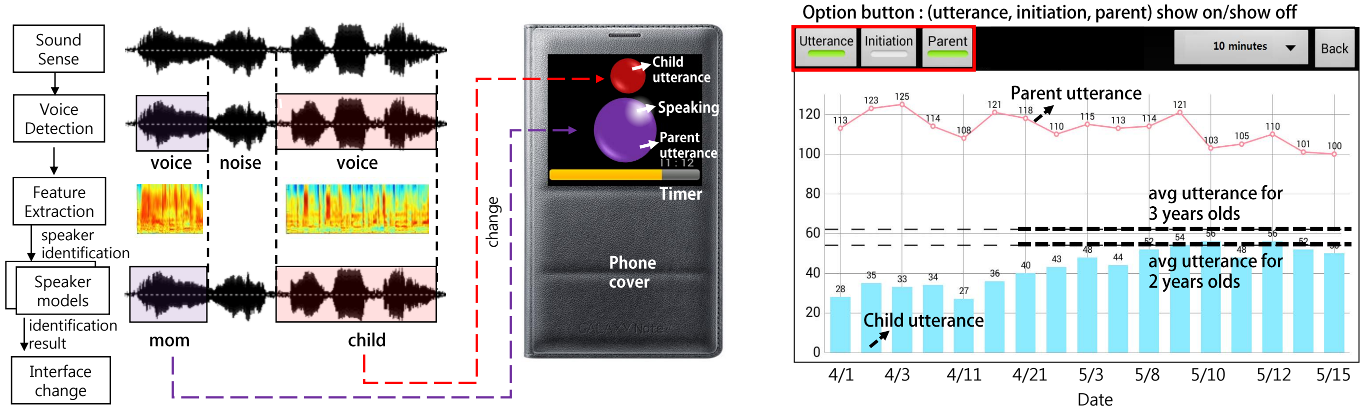
(d) Smartphone cover – Based on our preliminary study, children can be very interested in smartphones. To minimize distracting the children, we used a smartphone cover to minimize the amount of screen used for the feedback.

Result Mode

Following each conversation, the data analyzed for each conversation are saved; an example of a result is shown in Figure 2(c). By collecting and showing data from the previous sessions, the parents are able to observe the change in the number of utterances and the initiation ratio of the children as a function of time. The number of average utterances of the same age group is shown with a dotted line, motivating the parents by setting a long-term goal.

Speaker Identification

The application calculates the number of utterances by classifying the speaker of the utterance as either the parent or the child. Children's vocal characteristics have higher fundamental and formant frequencies, greater spectral variability, slower average speaking rate, and a higher variability in speaking rate compared with adults [37]. As a result, the voices of adults and children can be distinguished based on the different voice features. Common algorithms for identifying voices compares the pre-trained voice model and input sound by using machine learning algorithms [29, 8]. Figure2 (a) shows the high-level diagram of speaker identification that we implemented. The microphone within the smartphone senses the sound coming from the nearby environment, and Voice Activity Detection (VAD) algorithm is used to separate human voices from other noises. We used a recently implemented VAD algorithm [34] as it was robust even within a noisy environment. After detecting voice signal, feature extraction is performed for speaker recognition. The Mel-Frequency Cepstral Coefficient (MFCC) was used to extract



(a) Speaker identification system architecture (b) Interface design (c) Result graph

Figure 2. (a) TalkLIME System architecture, (b) the interface, and (c) result mode.

the features [10]. Speaker recognition algorithm is used for the identification of the voice [10, 30, 9]. For speaker recognition, speaker models must be pre-trained for feature comparison. We created the trained speaker models for both the parents and the children with the voice recorded from the pre-test prior to the start of the 6-week intervention. An open source speaker recognition algorithm [9] was used. The entire system TalkLIME was implemented on the Android operating system and executed on the smartphone itself.

METHODOLOGY

Participants

Institutional Review Board (IRB) approval was as obtained prior to the study. For our study, 8 children with language delay and their parents were recruited and the children were tested at a language therapy center affiliated with a local university in Korea. Compared with other HCI studies, the number of participants was relatively low; however, other studies involving special-needs participants often have small numbers of participants because of the difficulty in recruiting participants and the length of study.

The children selected for the study met the following criteria:
(a) they were 2–3 years of age;
(b) they could communicate through vocalization,
(c) they had a language delay, as they scored in the bottom 10 percentile [20] as indicated by the Korean MacArthur-Bate Communicative Development Inventories (K M-B CDI) [4] [26] and Korean Standard Receptive Vocabulary Test.
(d) they had no problems with perception and listening.

In addition, the parents of the children met the following selection criteria:
(a) they were the primary caregiver for the child,
(b) they had no problems with perception and listening; and
(c) they have never experienced parent-training to improve interaction with their children.

The participants were divided into a control group and an experimental group who used the TalkLIME mobile system

	Parents (avg)			Children (avg)		
	Age	# of utterance	Initiation Ratio	Age	# of utterance	Initiation Ratio
Control Group(n=4)	34	486	13.95%	2.41	76.75	14.85%
Experiment Group(n=4)	34	439.25	12.57%	2.52	160.5	8.70%

Table 3. Participants average functionality difference in the two groups.

(Table 3). Both groups went through the same parent-training with a speech-language therapist. While the utterance average between the two groups was different, the standard deviation was much higher for the control group. The T-test was used to confirm that there was no statistically significant difference between the two groups.

Procedure

The user study procedure consisted of a pre-test, 6 weeks of intervention through parent-training, and a post-test with the eight parents and their children with language delay, as shown in Figure 1. Prior to the actual user study, a pilot study was conducted with children who did not participate in this study to ensure that the system functioned properly during parent-child interactions. We also used the same toys that were used in the main user-study and wanted to determine what impact the smartphone has on the children's ability to focus on the toys. The pilot study demonstrated that the smartphone did not have any noticeable impact on the children.

a. Pre-test

The pre-test was conducted at the participant's house to measure the language development of the child. This measurement was used to determine the baseline for both the child and the parent. The participants (the child and the parent) played with 3 different types of toys (stickers, playing doctor, and playing house) for 30 minutes, with each toy being used for approximately 10 minutes. The conversation was video recorded and subsequently analyzed by an SLP to determine the baseline statistics for the child and the parent.

b. Intervention – Parent-training

The parent-training was performed for 6 weeks, one session per week, with all of the children's parents receiving the training simultaneously. Each session lasted approximately 1.5

[4]The K M-B CDI is based on the MacArthur-Bate Communicative Development Inventories (http://mb-cdi.stanford.edu/) which is commonly used for assessing language and communication skills.

Week	Topic	Contents
1	Let your child lead	Parent's and Child's interaction level, type
		Look at child's gesture, expression
		Wait for child to response
2	Follow your child leads	Participate child's play
		Mimic child's expression, sound, voice
		Comment to child's verbal and action
		Interpret child's response
		Appropriate interaction strategies
3	Video feedback about parents' interactions	
4	Take turns to keep the interaction going	Continue conversation with questions
		Give a sign for turn-taking
5	Add language to the interaction	Speak simple and short and slowly
		Extend child's message, highlight words
6	Sharing books and music	Interaction by reading books and singing

Table 4. Parent-training topics provided during 6 weeks of intervention.

hours and was based on *It Takes Two to Talk* [17], which provides strategies for interaction between parents and children with language delay. An SLP, who had completed the *It takes Two To Talk* workshop and had a license to teach, trained the parents. The topics of the parent-training for each session is summarized in Table 4. Following each training session, all of the participants were encouraged to apply the parent-child interaction strategies at home.

1) Control group: The parents in this group did not use TalkLIME for intervention. Instead, the parents in the control group were asked to maintain a daily diary of their interaction with their child. The purpose of the diary was to provide motivation for the parents to continuously interact with their child. The parents were instructed to record the amount of time they spent with their child and describe which strategies from parent-training that they applied in their parent-child interaction. Based on the preliminary study, parents often had difficulty in continuously interacting with their child and thus, the diary was used to help the parents in the control group. The diaries from the parents were collected by the SLPs weekly during the 6-week.

2) Experimental group: After the first session of the parent-training, the parents in the experimental group were provided with a smartphone (Samsung Galaxy Note 4) which had TalkLIME installed. The parents were instructed on how to use the system, including a description of the interface, the definition of the balls showed on the screens, the result graphs, for approximately one hour. The parents in this group used TalkLIME during the parent-child interactions and received real-time feedback through the smartphone. Similar to the control group, the parents were also asked to spend 10 to 15 minutes with their children every day with TalkLIME, but the parents were not asked to write a diary since the duration and conversation were recorded on the smartphone.

The location of the smartphone is important during parent-child interaction and should be placed where the parents can easily glance at it, without it becoming a distraction to the parent-child interaction. The parents were encouraged to place it on the table or on the floor (if they are sitting) and preferably away from the child.

c. Post-test

After finishing the intervention period of 6 weeks, a post-test was conducted similarly to the pre-test to assess the changes in the number of utterances and the initiation ratio for both the parents and the children.

d. Data analysis & Reliability

The results of the pre-test and post-test were manually analyzed by the researchers. The recorded conversations were shortened to 25 minutes to make the conversation time uniform across all of the participants. The number of utterances was manually counted for each test, and the initiation ratio was calculated based on the number of total utterances. The utterances were counted based on the criteria described in [12]. We defined the utterance as an *initiation* utterance if it starts a conversation with a new topic or was spoken after 3 seconds of silence [36]. Based on the number of total utterances and the number of initiation utterances, the initiation ratio was calculated based on the following ratio, i.e., (# of initiation utterances / # of total utterances) $\times$ 100.

To analyze the quantitative results statistics, Shapiro-Wilk normality test was done to ensure that ANOVA could be applied since the number of participants were low. Then two-way mixed ANOVA was used to observe the effect of using TalkLIME – both between the pre-test and the post-test within each group and also, between the control group and the experiment group.

The number of utterances and initiation ratio can be different depending on the researchers because some of the utterances can be interpreted subjectively. To understand the data analysis reliability, we randomly selected 20% of the analyzed data and re-analyzed the selected data with another researcher. The average consistency between the two researchers was 98.67% for the number of utterances of the parents, 95.23% for the initiation ratio of parents, 94.05% for the number of utterances of children and 92.66% for the initiation ratio of children.

EXPERIMENTAL RESULTS

The main components in language development are (1) the number of utterances and (2) the initiation ratio from the spoken utterances. One of the main goals of parental training is decreasing the number of parents' utterances to give the children a chance to speak, and a low initiation ratio of the parents shows that the conversation was led mainly by the children. Based on usage, on average, each parent used the mobile-system approximately 3 or 4 days per week.

Quantitative Results

Figure 3(a) shows the results of the number of parental utterances between the pre-test and the post-test for the both groups. There was a significant main effect between the pre-test and the post-test for both groups ($F_{(1,6)}$=25.630, ρ =.002), but there was no difference between the two groups ($F_{(1,6)}$=.156, ρ =.707). Thus, parents' decreasing utterances were the effect of the parent-training, not necessarily from using TalkLIME. Figure 3(b) shows the result of the initiation ratio of the parents. There was a non-significant main effect between the pre-test and the post-test for both groups ($F_{(1,6)}$=1.071, ρ =.341).

Figure 3(c) shows the results of the number of utterances from the children. Although the average number of utterance increased for both groups, there was a non-significant difference between the pre-test and the post-test for both groups ($F_{(1,6)}$=1.790, ρ =.229). Figure 3(d) shows the result

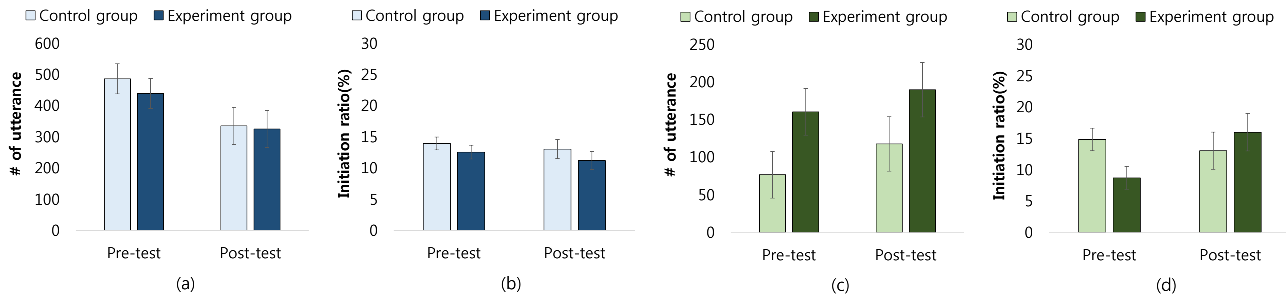


Figure 3. Results from the pre-test and post-test comparison for (a) parents' utterance, (b) parents' initiation ratio, (c) children's utterance, and (d) children's initiation ratio.

of the initiation ratio from the children. There was a non-significant main effect between the pre-test and the post-test for both groups ($F_{(1,6)}$=2.822, ρ =.144), whereas the experimental group showed improvement between the pre-test and the post-test ($F_{(1,6)}$=7.754, ρ =.032). The results showed that the experimental group using TalkLIME resulted in a statistically significant increase in the initiation ratio of the children.

To summarize the quantitative results, there was statically no difference between the control group and the experiment group for the parents. However, the initiation ratio of the children was improved in the experimental group. Although the amount children's utterances did not show a significant increase, the length of the study was not necessarily long enough to result in a significant improvement. It is not uncommon for children's improvement in language to occur over much longer periods of time (e.g., at least, 3 months [35, 3]). Because of the difficulty in recruiting subjects for such extended periods of time, we were limited to only 6 weeks for this study. However, the initiation is an important factor since it is a sign of improvement in language development [17]. *As a result, the significant improvement in the children's initiation shows the potential impact of using a mobile-system intervention to improve parent-child interaction.*

Qualitative Results

After post-tests, we conducted a survey with the parents in the experimental group using the questions listed in Table 5. The parents were asked to give a response using a Likert scale (1:strongly disagree, 5:strongly agree) and were also asked to comment on why they gave such score.

	Survey question	P1	P2	P3	P4
Q1	*Was the ball feedback helpful knowing the conversation status?*	4	4	4	5
Q2	*Was the timer helpful?*	2	4	5	5
Q3	*Was the result graph useful?*	3	4	3	5
Q4	*Do you want to use this application continuously?*	1	4	1	5

strongly disagree (1) ↔ strongly agree (5)

Table 5. Results from a post-study survey with the experimental group.

Screen-based conversation feedback (Q1)

Overall, the parents in the experiment group were satisfied with the screen-based real-time feedback with two balls. The parents were able to identify the relevant details of the conversation status – e.g., the number of utterance or initiation ratio of the children is low. Based on this feedback, the parent would try to increase the child's utterance using what they learned in parent-training. One parent mentioned: "*When I saw that my child's ball was too small, I tried to give my child a chance to speak.*" (P2). Interestingly, there was one parent who was stimulated because of the low number of utterance from the child, mentioning that "*After I knew the conversation was excessively led by me, I tried to decrease my utterances and induce the child's utterances by establishing the conversation strategies that I learned from the parent-training.*" (P4). However, there was one parent concerning the real-time feedbacks: "*When I knew the number of utterances of my child was very low, I tried to induce him to speak, but if the ball was still small despite my efforts, I felt exhausted.*" (P3).

Motivating parents to interact consistently (Q2, Q3)

Most parents in the experiment group agreed that the timer was helpful to continue the interactions, except one parent (P2). When they used the application at first, most of them felt that interacting with their children until 10-minutes timer is done is not easy: "*When I watched the timer after interacting with my child for quite a while, I surprised that just 3 minutes had gone, even though it seemed like it had been 10 minutes at least.*" (P3). The simple timer not only informed how long did the parents actually interact but also motivated the parents by proving some form of encouragement, setting a short-term goal. One parent said, "*I know that consistently interacting with my children is important, but I, sometimes, want to finish or skip the interaction with my child, because I am already too tired to talk due to house chores or other things to do. However, I cheered up by looking the remaining time in the timer when I felt it.*" (P4). Interestingly, one parent did not care about the timer. The therapist asked the parents to interact with their children for 10~15 minutes a day. Most parents felt that it is a long time, whereas the parent said it is no problem (P1).

Two out of four parents said the result graph was useful, and the others were neutral (Q3). Most parents are very enthusiastic when they initially begin parent-training, but if they do not feel that their children's language development improve for a long time, they might lose interest in practicing parent-training at home (children's language improvement occurs at least 3 months [35, 3]). The result graph was also useful in observing the changes in the language development of the children and to check the intervention time. When one parent who was passionate about children's education knew that there was a subtle change, she listened to the recorded voice

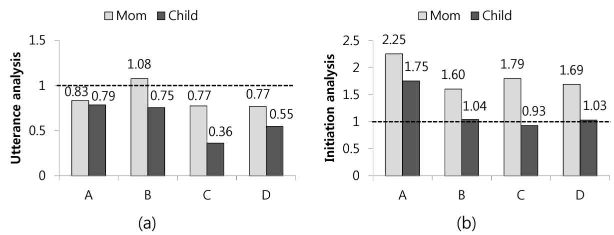


Figure 4. The accuracy of (a) utterance and (b) initiation utterance count.

to check their conversation: *"Did I make a mistake?", "Did I still have wrong speaking habits?" and "When did my child respond to me better?"* (P4). It also included the number of average utterances of the same age group with their children as a dotted line. The parents used the dotted line as a long-term goal of her child: "*I especially liked the dotted line showing the average number of utterances of the same age group. The difference between the dotted line and my child's line alerted me, motivating to practice more.*" (P2).

User experiences (Q4)

While two of the parents would use this system again, the other two parents disagreed. One parent express privacy concern since the parent-child conversation was recorded – however, this is a feature that can easily be disabled and for our purpose, the conversation was recorded to obtain ground truth for parent-child statistics. The other parent initially liked the system and served its purpose but felt it was no longer needed. "*This system was helpful in initially inducing my child's utterance. It was also helpful to know how I interacted with my child and learn who was leading the conversation. However, once I felt like I was interacting correctly, I did not feel like I needed the system any longer.*" (P1), and another parent said, "*It was also difficult to carry around the TalkLIME system and execute the application when interacting with my child.*" (P3). For our study, the parents were required to use a separate smartphone for the study but if our system can be ported to other smartphones, we believe that it would be more convenient to the parents if they can simply use their own personal smartphones.

Mobile-System utterance analysis accuracy

Figure 4 shows the results of the number of utterances and initiation utterances analyzed by the application and manually. The accuracy was calculated with the conversation that was recorded in the post-tests. The application accuracy was calculated with the following equation.

$$\text{Accuracy} = \frac{\text{\# of utterances analyzed by the computer}}{\text{\# of utterances analyzed manually}}$$

An accuracy of 1 means that the mobile system analysis and the manual analysis were identical. The value can be larger than 1 if the mobile system's analysis resulted in a larger number of utterances in the analysis. The number of utterances manually analyzed was larger than what the system analyzed for all children, and very low for child C and child D. Children with language delay often speak in an especially short and quiet manner; thus, the voice activity detection algorithm could not detect the sound as a human voice. For child C and child D, their utterances were especially short and quiet; even manual classification by an expert was difficult, but they were ultimately classified as utterances in the manual analysis. In addition, the numbers of utterances for child C and child D were relatively low – and thus, even a few errors resulted in a lower accuracy result. The number of initiation utterances tended to be higher than in the manual analysis. Since only silence was used to determine initiations, the system reported a larger number of initiation utterances. However, the error trend across the different parent-child was relatively consistent and for our system, the main purpose to observe the *change* in the parent and the child. Further discussion on the accuracy of our system is presented in the following section.

DISCUSSION

Difficulties in Recruiting Participants

We conducted our user-study with 8 parent children pairs, each group had 4 parent-child participants. The number of participants was relatively low, compared to other HCI studies, but studies involving special-needs or special-disorders often have a small number of participants because of the difficulties in recruiting participants. The 6-week length for the user study also made it difficult for some to participate in the study. We had initially recruited 10 parent-child participants but during the intervention period of 6-weeks, one parent within the control group gave up because of personal work and stopped attending parent-training. Another parent in the experiment group rarely used the mobile application despite our continuous request to use the system – thus, we excluded the data from these two parent-child participants. On several occasions, we also had to encourage the parents to not drop out and continue with the experiment during the 6-week intervention period.

Children's Awareness and TalkLIME

In this work, even though the child had a language delay, the focus was on the parents and providing effective intervention for the parents. The child was not made aware of TalkLIME because of their age as they were mostly pre-school age and it would be very difficult for them to understand the concept of "initiation", "utterances", etc. In addition, SLPs prefer "natural" and embedded techniques of intervention since children feel like they are playing and this is how language really develops [1], instead of explicitly increasing the awareness of their language delay. For older children, it remains an interesting future work to see how they can be made more active participants in a system like TalkLIME.

Objective feedback and Motivating Parents

As described earlier, the parents in the control group were asked to maintain a daily diary. Interestingly, the data collected from the diary of parents in the control group suggests that the parents did not skip a single day of parent-child interaction session during the 6 weeks. In comparison, the experimental group parents showed some variation during the 6 weeks. Figure 6 plots the amount of time that four of the eight parents spent with their children in parent-child interaction sessions. The results from the experimental group (Figure 6(a,b)) are based on the data provided by TalkLIME. It is clear that there is a variation in the amount of time spent with the child as the parents were not able to have a parent-child interaction sessions on some days. However, the results

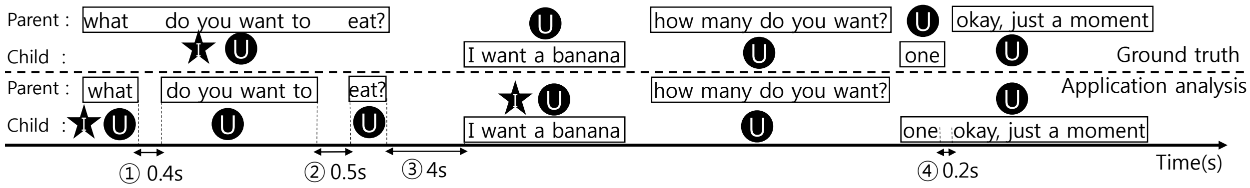


Figure 5. Analysis of the errors from the proposed TalkLIME system (Ⓘ : initiation , Ⓤ : utterance).

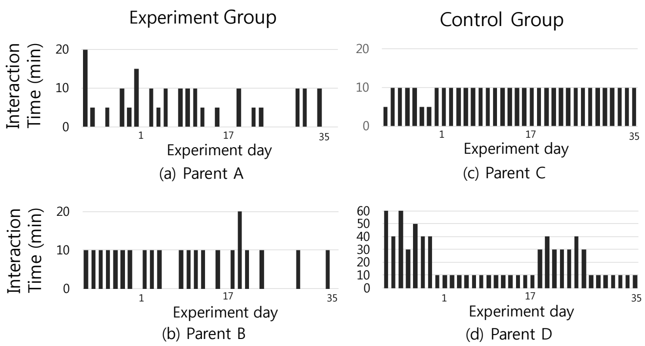


Figure 6. Amount of time spent in parent-child interaction session at home for parents in the experimental group (a,b) and control group (c,d).

from the control group (Figure 6(c,d)) suggests that these two parents did not skip a single parent-child interaction session.

According to the diary, the parents in the control group felt that general communication their daily lives were part of the strategies. For example, some of the comments in the diary were the following.

- "*When I tried to finish a bath, he wanted to stay in the tub. So, I gave him his favorite food so we could finish the bath.*"
- "*I read a book to my child and he paid attention.*"

Based on these comments, the parents in the control group tend to over-estimate the efforts they spend on the parent-child interaction and thus, it is difficult to understand how much effort the parents put in with their child. In comparison, more objective measurements are provided by TalkLIME and enables the parents to understand their effort. For example, parents can become frustrated when their child does not show any improvement in the language development – however, the parent might inaccurately assume that they are putting in the effort when in reality, they might not be.

Speaker Identification Accuracy

While the TalkLIME system had a positive effect on parent-child interaction sessions, the accuracy of the speaker identification can be improved and we describe some of the common sources of errors in our system.

(1) **Short utterances:** Some utterances tend to be very short (1-2 sec), especially since children have a language delay. It has been shown that at least several seconds of sound is needed to achieve high-level of accuracy [11].

(2) **Non-utterance sounds:** Sounds such as "hmm~" and "oh~" do not have meaning and are not necessarily utterances. However, since the sounds come from human voice, and voice activity detection algorithm identifies them as "voice" (or utterances).

(3) **Noise:** Some toys (e.g., blocks) used in parent-child interaction make significant noise. Voice overlapped with noise results in the sound having different vocal features and makes the speaker recognition more difficult.

(4) **Quickly speaking** – If a parent starts speaking (<0.3 s) quickly after a child, it is interpreted as a single long utterance.

Figure 5 shows an example of some inaccuracies from an actual parent-child interaction in our study. The top shows the ground truth analyzed by an SLP while the bottom shows the analysis from TalkLIME. Errors ① and ② occurred when the speaker talked slowly and resulted in a single utterance being broken up into multiple utterances. An incorrect initiation utterance was identified in Error ③ occurred when a speaker was late in responding. In this example, the child was responding to the parent and not initiating; however, since an initiation utterance was defined as an utterance after a silence of 3 s or longer [36], the child's utterance was incorrectly identified as an initiation utterance. Error ④ occurred when someone starts speaking quickly (<0.3 s) after another speaker's utterance and the two utterances are combined as one utterance. All of these factors contributed to the reduced accuracy of the speaker identification within TalkLIME. Improving the accuracy of the speaker identification needs to be addressed as part of future work.

CONCLUSION

In this work, we proposed and evaluated TalkLIME – a mobile intervention system to improve parent-child interaction for children with language delay. By providing less-disruptive feedback to the parent through the smartphone screen, TalkLIME provides real-time feedback to the parents on their interaction with their children. Based on the post-study interview, the interface of the TalkLIME system, including the timer and the data-drive interface, provided encouragement to the parents in their daily parent-child interaction. In addition, our user study showed that the experimental group using the mobile-system intervention showed a significant improvement in the initiation ratio from the children.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their insightful comments which improved this paper. This work was supported in part by the IT R&D program of MSIP/KEIT (10041313, UX-oriented Mobile SW Platform) and in part by the National Research Foundation of Korea Grant funded by the Korean government (Ministry of Science, ICT & Future Planning) (NRF-2013R1A2A2A03068010).

REFERENCES

1. Kaiser A.P., Hancock T.B., and J.P. Nietfeld. 2000. The effects of parent-implemented enhanced milieu teaching on the social communication of children who have autism. *Journal of Early Education and Development(Special Issue)* 4, 2 (2000), 423–446.
2. Ingersoll B. and Dvortcsak A. 2006. Including parent training in the early childhood special education curriculum for children with autism spectrum disorders. *Journal of Positive Behavior Interventions* 8, 2 (2006), 79–87.
3. Goldfield B.A. and Reznick J.S. 1990. Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of child language* 17, 1 (1990), 171–183.
4. Cunningham C.E. and Siegel L.S. et al. 1985. The behavioral and linguistic interactions of specifically language-delayed and normal boys with their mothers. *Child Development* 56 (1985), 1389–1403.
5. Alpert C.L. and Kaiser A.P. 1992. Training parents as milieu language teachers. *Journal of Early intervention* 16, 1 (1992), 31–52.
6. Xu D., Yapanel U., and Gray S. 2009. Reliability of the LENA Language Environment Analysis system in young childrens natural home environment. *LENA Foundation* (2009).
7. Lilian de Greef, Mayank Goel, Min Joon Seo, Eric C. Larson, James W. Stout, James A. Taylor, and Shwetak N. Patel. 2014. Bilicam: Using Mobile Phones to Monitor Newborn Jaundice. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14)*. 331–342.
8. Reynolds Douglas A., Quatieri F., Thomas F, and Robert B. Dunn. 2000. Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing* 10 (2000), 19–41.
9. Reynolds Douglas A., Quatieri F., Thomas F, and Robert B. Dunn. 2013. An Open-source State-of-the-art Toolbox for Broadcast News Diarization. *M., Rouvier and G., Dupuy and P., Gay and E., Khoury and T., Merlin and S. Meignier* (2013), 25–29.
10. Benot G. B. Fauve, Driss Matrouf, Nicolas Scheffer, Jean franois Bonastre, Senior Member, and John S. D. Mason. 2007. State-of-theArt Performance in Text-Independent Speaker Verification through Open-Source Software. *IEEE Transactions on Audio, Speech and Language Processing* (2007).
11. Gerald Friedland and Oriol Vinyals. 2008. Live Speaker Identification in Conversations. In *Proceedings of the 16th ACM International Conference on Multimedia (MM '08)*. 1017–1018.
12. Lee H.J. and Kim Y.T. 2001. Turn-Taking Characterist ics of Children with Specific Language Impairment and Normal Children. *Communication Sciences and Disorders* 6, 2 (2001), 293–312.
13. Mohammed E. Hoque, Joseph K. Lane, Rana El Kaliouby, Matthew S. Goodwin, and Rosalind W. Picard. 2009. Exploring speech therapy games with children on the autism spectrum. International Speech Communication Association, 1455–1458.
14. Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. 2014. TalkBetter: Family-driven Mobile Intervention Care for Children with Language Delay. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, 1283–1296.
15. Tsybina I. and Eriks-Brophy A. 2007. Issues in research on children with early language delay. *Contemporary Issues in Communication Science and Disorders* 34 (2007), 118–133.
16. Law J. and Garrett Z. 2003. Speech and language therapy interventions for children with primary speech and language delay or disorder. *Campbell Collaboration* (2003).
17. Pepper J. and McDade E., Weitzman. 2004. *It takes two to talk: A practical guide for parents of children with language delays*. The Hanen Centre.
18. Kyung Hea Jeon, Seok Jeong Yeon, Young Tae Kim, Seokwoo Song, and John Kim. 2014. Robot-based Augmentative and Alternative Communication for Nonverbal Children with Communication Disorders *(UbiComp '14)*. 853–859.
19. Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: Facilitating Participatory Parental Mediation of Adolescents' Smartphone Use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. 867–878.
20. Leonard L. 1998. *Children with specific language impairment*. Cambridge: MIT Press.
21. Rescorla L. 2005. Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research* 48, 2 (2005), 459–472.
22. LENA. *http://www.lenafoundation.org*.
23. Leslie S. Liu, Sen H. Hirano, Monica Tentori, Karen G. Cheng, Sheba George, Sun Young Park, and Gillian R. Hayes. 2011. Improving Communication and Social Support for Caregivers of High-risk Infants Through Mobile Technologies. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. 475–484.
24. Dunn L.M. 1997. PPVT-III: Peabody picture vocabulary test. Circle Pines, MN: American Guidance Service. (1997).
25. Moseley M.J. 1990. Mother-child interaction with preschool language-delayed children: Structuring conversations. *Journal of communication disorders* 23, 3 (1990).

26. Soyoung Pae and Kumjoo Kwak. 2011. *Korean MacArthur-Bates communicative development inventories (K M-B CDI)*. Seoul: Mind Press.

27. Laura Pina, Kael Rowan, Asta Roseway, Paul Johns, Gillian R. Hayes, and Mary Czerwinski. 2014. In Situ Cues for ADHD Parenting Strategies Using Mobile Technology. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '14)*. 17–24.

28. Paul R. and Shiffer M.E. 1991. Communicative initiations in normal and late-talking toddlers. *Applied Psycholinguistics* 12, 4 (1991), 419–431.

29. Douglas A. Reynolds. 1995. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication* 17 (1995), 91–108.

30. Seyed Omid Sadjadi, Malcolm Slaney, and Larry Heck. 2013. *MSR Identity Toolbox v1.0: A MATLAB Toolbox for Speaker Recognition Research*. Technical Report MSR-TR-2013-133.

31. Petr Slovák, Kael Rowan, Christopher Frauenberger, Ran Gilad-Bachrach, Mia Doces, Brian Smith, Rachel Kamb, and Geraldine Fitzpatrick. 2016. Scaffolding the Scaffolding: Supporting Children's Social-emotional Learning at Home. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. 1751–1765.

32. Tobias Sonne, Jörg Müller, Paul Marshall, Carsten Obel, and Kaj Grønbæk. 2016. Changing Family Practices with Assistive Technology: MOBERO Improves Morning and Bedtime Routines for Children with ADHD. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. 152–164.

33. Svetha Venkatesh, Dinh Phung, Thi Duong, Stewart Greenhill, and Brett Adams. 2013. TOBY: Early Intervention in Autism Through Technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. 3187–3196.

34. WEBRTC. *http://www.webrtc.org/*.

35. Kim Y.T. 2000. Content and Reliability Analyses of the Preschool Receptive-Expressive Language Scale (PRES). *Communication Sciences and Disorders* 5, 1 (2000), 1–25.

36. Kim Y.T. and Lee H.J. 2013. Conversational Turn-taking of Toddlers with Language Delay. *Special Education* 12, 3 (2013).

37. Yumin Zeng and Yi Zhang. 2007. Robust Children and Adults Speech Classification. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, Vol. 4. 721–725.