19th IEEE International Symposium on Robot and
Human Interactive Communication
Principe di Piemonte - Viareggio, Italy, Sept. 12-15, 2010

Using Child-Robot Interaction to Investigate the User Acceptance of Constrained and Artificial Languages

Omar Mubin[1], Suleman Shahid[2], Eva van de Sande[2], Emiel Krahmer[2], Marc Swerts[2],
Christoph Bartneck[1], Loe Feijs[1]
[1]*Department of Industrial Design, Eindhoven University of Technology, The Netherlands*
[2]*Tilburg Centre for Creative Computing, Tilburg University, Tilburg, The Netherlands*

***Abstract*— The possibility of improving speech recognition accuracy within human computer and robot interaction by having users interact in a constrained natural language or an artificial language has been explored. However, what has not been evaluated yet is the user acceptance of such forms of interaction. In this paper we discuss two separate but similar studies which were aimed at assessing the usability of constrained and artificial languages in contrast to natural languages. The interaction context was implemented in a game played between children and the iCat robot. We subjectively measured various variables related to gaming experience and interacting in the new languages. Our results reveal that there were no significant differences in the user experience across the two interaction mediums in comparison to natural languages.**

I. Introduction

Speech is one of the primary modalities utilized in Human Robot Interaction and is a vital and natural means of information exchange [1]. Therefore, improving the status of speech interaction in HRI could consequently lead to more efficient and more pleasant user-robot-interaction. Researchers in HRI have concentrated on designing interaction which can carry out some form of a social dialogue between humans and a robot. Reviewing various state of the art dialogue management systems unearthed several hindrances behind the adoption of natural language for robotic and general systems alike. The limitations prevailing in current speech recognition technology for natural language is a major obstacle behind the unanimous acceptance of Speech Interfaces for robots. Existing speech recognition is at times not good enough for it to be deployed in natural environments, where the ambience influences its performance. Recent attempts to improve the quality of automatic speech recognition of natural language for machines have not advanced sufficiently [2]. State-of-art automatic speech recognition has not advanced far enough for most applications, partly due to the inherent properties of natural languages that make them difficult for a machine to recognize. Examples are ambiguity in context and homophones (words that tend to sound the same but have different meanings) [3]. Dialogue Management and Mapping is one of the popular techniques used to model the interaction between a user and a machine or a robot [4]. However the inherent irregularity in natural dialogue is one of the main obstacles against deploying dialogue management systems accurately [5]. There has been attempt to solve such ambiguities by utilizing non verbal means of communication. As reported in [6], a robot tracks the gaze of the user in the case when the object or the verb of a sentence in a dialogue may be undefined or ambiguous. As a consequence of the prior discussed problems miscommunication occurs between the user and robot. The mismatch between humans' expectations and the abilities of interactive robots often results in frustration. Users are disappointed if the robot cannot understand them properly even though the robot can speak with its mechanical voice. To prevent disappointment, it is important to match the communication skills of a robot with its perception and cognitive abilities.

Recent research in speech recognition is moving in the direction of trying to alter the medium of communication in a bid to improve the quality of speech interaction. As stated in [7], constraining language is a plausible method of improving recognition accuracy. Constrained languages are defined as languages which are obtained by restricting the grammar and vocabulary of natural languages to reduce ambiguity. Artificial languages are another branch of languages which are deliberately designed for a specific purpose, e.g. to improve communication between humans. In linguistics, there are numerous constrained and artificial languages which address a user perspective by making communication between humans easier and/or universal (e.g. Basic English [8] and Esperanto [9] respectively). To the best of our knowledge, there has been no attempt to optimize a complete spoken language for automatic speech recognition for interaction with robots or otherwise.

Our goal lies in the design and evaluation of speech recognition friendly languages to be used while interacting with robots. The first criticism that might be drawn is that for any artificial language it would need to be learnt by

Manuscript received March 1, 2010.

Omar Mubin, Christoph Bartneck and Loe Feijs are with the Department of Industrial Design, Eindhoven University of Technology (TU/e), The Netherlands. (Phone: 0031402473842; e-mail: o.mubin@tue.nl, c.bartneck@tue.nl, l.m.g.feijs@tue.nl).

Suleman Shahid, Eva van de Sande, Emiel Krahmer and Marc Swerts are with the Tilburg Centre for Creative Computing, Tilburg University, The Netherlands. (e-mail: s.shahid@uvt.nl, evavdsande@gmail.com, e.j.krahmer@uvt.nl and m.g.j.swerts@uvt.nl).

978-1-4244-7990-0/10/$26.00 ©2010 IEEE

users. However, we wish to explore the benefits that an artificial language could provide if it's designed in such a way that it is speech recognition friendly. This benefit might end up outweighing the price a user has to pay in learning a new language. A second criticism that might be levied on our idea is that many artificial languages were created already but nobody ended up speaking them. Where our approach is different is that we aim to deploy and implement an artificial language for interaction with a robot and once several robots can speak a certain language it might lead and encourage humans to speak it as well. Moreover, we intend to deploy the interaction using artificial languages within specific scenarios so that the constraints can be taken care by the languages.

In this research we discuss the separate evaluation of two such artificially constructed languages: a constrained version of the natural language Dutch and an artificial language ROILA. We did not attempt to compare artificial and constrained languages but rather determine the user acceptance of both separately. Both case studies were carried out in an interaction context of a child playing a game in cooperation with a robot, as to provide the child with a scenario to interact in the new languages. Additionally, a gaming scenario is a realistic context as children use games in their everyday social and educational life [10, 11] and by having children play a game we can subjectively measure fun by virtue of their gaming experience, a strategy which has been used before [12, 13]. Therefore we could ascertain if the children had more fun playing the game in a specific language. By adopting a specified fixed scenario we aimed to have much better control over the study as we were working in the domain of new languages. Moreover, we selected children as users because we speculated that the children would be good learners of new languages as compared to adults [14].

We used the iCat robot [15] as the gaming partner of the children. The iCat robot is a cat like robot by Philips that has the ability to communicate to users via both verbal and non-verbal means such as by exhibiting facial expressions. We also speculated that due to its non-human like appearance the iCat robot would advocate lower expectations from the children [16] and consequently they would be willing to learn new languages to interact with it.

The iCat robot was controlled by the facilitators in a Wizard of Oz fashion and no speech recognition was taking place since our primary goal was to identify whether novel languages would be acceptable to users. In an actual setting; i.e. deployed with a speech recognizer, the accuracy of speech recognition would also affect the user experience. To summarize, the primary goal of this research was to evaluate whether children are comfortable with using constrained or artificial languages in comparison to natural languages and whether they are willing to invest some effort in learning such languages. If it would turn out that children are still relaxed, comfortable and have fun while communicating in constrained or artificial languages, then this finding would be a positive step forward in the implementation of speech recognition friendly languages, as the biggest objection with such new languages is that users would not be motivated or comfortable to interact in them.

II. Case Study I: Constrained Languages

The first study was carried out in Tilburg, the Netherlands, where children either collaborated in their native language Dutch or in a restricted set of Dutch utterances that was suitable for the communication purpose (i.e. collaborating in a game).

A. *Participants*

92 children took part in this study, that were between 9 and 12 years old. From them, 52 took part in the natural language condition (Dutch); the other 40 conversed with the iCat in constrained Dutch. We balanced gender in both conditions. All children had prior written consent by their parents and teachers to participate in this study and to use the results and audiovisual data for research purposes.

B. *Material*

The game employed had shown to induce emotions within children [17]. In the game, the child would see a row of six cards on the computer screen, where each card had a number written on it and only the first card was shown initially (see Figure 1). The other cards were placed upside down. The players' task was to guess whether the next card contained a number that was bigger or smaller than the previous one. The available card numbers were between one and ten and every number could only appear once within a solitary sequence. When the player guessed a number, the card would become visible. Then, the child would hear a characteristic non-speech sound (booing or clapping) to inform them about the correctness of their guess. To win a particular game, the child was required to guess every number correctly. The child played seven rounds of this game and was encouraged to discuss every guess with the iCat. We used a Wizard of Oz method to simulate both the verbal and non verbal behavior of the iCat. The wizard was located out of the child's vision, behind a screen. We received the input from the children through a camera and a microphone. The wizard could manipulate simple preprogrammed behaviors and animations that functioned as iCat's communicative response. A Dutch text to speech engine was also employed in order to elicit the responses of the iCat.

TABLE I
CONSTRAINED DUTCH SENTENCES (SOME EXAMPLES)

Constrained Dutch	English Translation
Wie is aan de beurt	Who's turn is it?
Mijn beurt	My Turn
Jouw beurt	Your Turn
Wat denk jij?	What do you think?
Ik denk hoger	I think higher
Ik denk lager	I think lower
We hebben gewonnen	We have won

C. Design of the constrained language

For the composition of the constrained language we considered the children's language level and the probability of using certain utterances when playing a game in a natural situation. With these aspects in mind we composed a constrained language consisting of fifteen permissible commands (see Table 1). When designing the constrained language, we also kept the difficulties in mind that speech recognizers would experience. For instance, the commands were designed such that they would have the least confusion amongst themselves. Moreover we used as few words as possible with which every sentence could be fully informative. Prior to interacting with the iCat, the children were given three minutes to study the constrained language so that they could recall and use it while playing the game. This was done out of the vision of the iCat, in a separate room. We checked whether the children could recall the commands, and whether they could easily restrict themselves to the defined vocabulary and recall appropriate commands. The duration of three minutes was finalized after conducting a few pilots which confirmed that it was long enough for the children to get acquainted with the language.

D. Procedure

The children were seated on a bench, which was placed in front of a table with a computer screen on it. As shown in figure 2, children sat beside the iCat. The iCat was positioned half diagonally, so that it could slightly turn its head for looking both at the screen and its game partner. When the children entered the room they first had an informal introduction with the iCat. After the game instructions the children played a practice trial together with the iCat. In this session, the experimenter was still present in the room in case there would be any questions. If the practice trial raised no further issues, the experimenter left the children's field of vision and started the game. After six game sessions, the experimenter guided the child back to another room where the child had to fill in his/her subjective evaluation. The experimenter (and iCat) was outside the children's view when they were filling in the questionnaire to avoid presence effects. Next, the experimenter asked the children some open questions, and rewarded the children with gifts. All sessions were video recorded. The video camera was placed on top of the monitor to record the child's face and upper body.

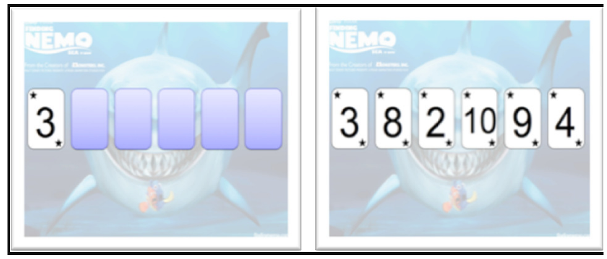

Fig. 1. Game I Example

E. Experiment Design and Measurements

The experiment was carried out between subjects with language type as the independent variable. To evaluate the children's social experience, we conducted several measurements by means of self-reports. We measured the fun that the children experienced during the game. For this, we adapted a Game Experience Questionnaire (GEQ [18]), which provides multiple questions to ask the children about the game's endurability, their engagement in the game, and whether their previous expectations about the game where met, all on a five-factor Likert scale. We also evaluated interaction in the constrained language, by means of the SASSI questionnaire (Subjective Assessment of Speech System Interfaces) [19], with which we measured factors such as cognitive demand and likeability. Given that all interaction was carried out in the form of a wizard of oz scenario we extracted only the relevant factors from the SASSI questionnaire and factors such as system response accuracy and speed were excluded.

F. Results

Independent samples t-tests were executed to ascertain if language type had an effect on the gaming experience. For the game experience, no significant differences between the language conditions were found for expectations ($t(88) = 1.119$, $p = .26$), endurability ($t(88) = 1.101$, $p = .27$) and engagement: $t(88) = 1.537$, $p = .12$). Results from the SASSI questionnaire revealed some interesting trends. The constrained language seemed to cause only slight cognitive demand, as ratings on this scale were below 3 on a scale from 1 (no cognitive demand) to 5 (complete cognitive demand) ($M = 2.06$, $SD = .96$). For the factor likeability, a significant difference was found between Dutch and constrained Dutch ($t(88) = 2.072$, $p < .05$) but the mean for both was rather high (Dutch = 4.65, Constrained Dutch = 4.40).

G. Discussion

Overall, the children were generally expressive to iCat during the game, both verbally and nonverbally (see Figure 3). The children had few difficulties with using the

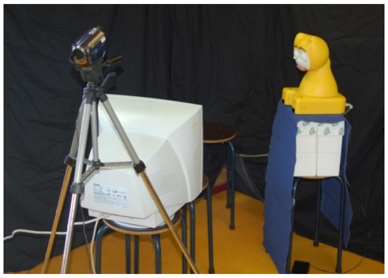

Fig. 2. Experimental setup

commands in the constrained language. They felt comfortable with the constrained language while playing, as they did not evaluate the constraints on verbal communication negatively. Furthermore, the children deviated only a few times from the constrained language while playing the game, and there were only three children that totally forgot one or more of the commands and could not continue with playing the game. However, the constrained language was not always used in its full potential, as at times children would only guess the card outcome and not argue the rationale of the choices made with iCat. For example, they would only use solitary words instead of complete sentences.

To evaluate whether effects of the constrained languages are indeed blurred by the minimalistic nature of the game, we could elicit a richer interaction by making a more extensive game where more choices could be made therefore leading to elaborate discussion. Therefore we decided to conduct a second study by adopting a different game where the interaction context was defined by an artificial language. It is worth pointing out that we could not have conducted a study with the three languages in question (natural, constrained and artificial) operating together as the three independent variables, primarily because of practical reasons. Consequently had this been the case, for a within subject analysis children would have to learn three languages and for a between subject analysis we would require a lot more participants for each condition.

III. Case Study II: Artificial Languages

In this study children played a game with the iCat in either their native language or in an artificial language: ROILA. Primarily, based on our results from case study I, we used a different game. The main goal of this study was to determine if a certain artificial language hampered the user experience of the children. We realized that the game used in case study I would not be entirely appropriate to evaluate artificial languages mainly due to its minimalistic nature. Despite losing the consistency between the two case studies we had no other choice but to design a new game where we could potentially explore the effect of artificial languages by having players engage in enhanced discussion.

Fig. 3. Children involved in the game

A. Participants

24 children took part in this study, that were between 10 and 13 years old. From them, 13 took part in the natural language condition; the other 11 conversed with the iCat in the artificial language ROILA. All children had prior written consent by their parents and teachers to participate in this study and to use the results and audiovisual data for research purposes.

B. Material

The game was a simple matching game in which the children had to match a given word with another word from a set of words based on some logical reasoning (see Figure 4). We anticipated that such a game would encourage the children to be much more verbally involved with the iCat as they would have to discuss the rationale of their choices and hence provide us with an opportunity to evaluate the artificial language with fairness. Therefore unlike the game employed in case study I they did not have one option but several and they had to discuss and deliberate their choices with the iCat. The children were only allowed one final guess and they would then be informed about the correctness of their guess. Yet again the iCat was controlled in a Wizard of Oz fashion with input from the children conveyed via a microphone and camera.

C. Design of the artificial language ROILA

The artificial language ROILA was designed on the basis of a genetic algorithm which used a confusion matrix of phonemes as its fitness function, design details can be found here [20]. The end product was a vocabulary with ideally the least confusion amongst its words and would hence as a whole be speech recognition friendly. Semantics were assigned based on a relation between frequency of use of English words and word length of the ROILA words. Shorter words would be assigned meanings of frequently used words so that the children would find them easier to learn. These words were then used to construct a command set via a simplified grammar. The children were asked to learn a set of roughly 25 commands in ROILA (see Table 2) and were given one day for training with regards to vocabulary and pronunciation. Before commencing with the experiment, necessary pronunciation checks were carried out for fluency. These commands were not fixed and the children could make changes to them based on the game situation.

TABLE II
ROILA SENTENCES (SOME EXAMPLES)

ROILA	English Translation
babo etat ujuk	Whos' turn is it?
ujuk ajne	My Turn
babo ajne	Your Turn
babo wimo	What do you think?
unso jabut Fish Sea	I think the correct answer is Fish and Sea
obat kutak	Which one should we pick

D. Procedure

The procedure and setup were identical to case study I (see Figure 5). The text to speech engine of the iCat was replaced with audio recordings to ensure correct pronunciations. The recordings were slightly transformed to sound robot like. We are also aware that for this purpose we could have used an open source Text to Speech engine such as Festival [21]. We aim to accomplish this in future experiments.

E. Experiment Design and Measurements

Language type was the main independent variable, as children played in either the natural or artificial language. Yet again we adopted the SASSI questionnaire [19] and used the following 3 factors from it: likeability, cognitive demand and habitability, where habitability is defined as the extent to which the user's conceptual model of the system agrees with what the system is actually doing.

F. Results and Discussion

For the three factors we achieved Cronbach alphas of $0.7<alpha<0.8$, which gives us sufficient reliability in the SASSI questionnaire. Language Type did not have an effect on any of the factors: likeability $t(22) = 1.43$, $p = 0.17$, cognitive demand $t(22) = 1.23$, $p = 0.22$ and habitability $t(22) = 0.22$, $p = 0.83$. On average the natural language was ranked as the more preferred for the three factors but both languages were ranked on the positive end of the 5 level likert ranking scales (Likeability for native language = 3.31 and for ROILA = 3.00). Generally the children were quite positive in interacting with ROILA as was exemplified by their subjective rankings. However we did note that one day was not entirely sufficient for the children to achieve complete fluency in ROILA. Therefore they had to undergo a supplementary training session with the experimenters prior to playing the game.

IV. DISCUSSION AND CONCLUSION

In this paper we have presented an evaluation of how children interact with a robot using constrained and artificial languages. We compared these two languages with natural languages (the native language of participants) and showed that the children as users were comfortable in communicating with a robot using both languages. Similar results have been reported in [22] where users preferred a constrained language over a natural language however no studies have been carried out to ascertain the user acceptance of artificial languages. Moreover, the results that we obtained by evaluating constrained and artificial languages are quite positive as we concluded that there was no significant cognitive overload exerted by such languages in comparison to natural languages. Some significant differences were found, e.g. in terms of likeability, but both non natural languages were ranked positively. In hindsight we believe that it was a good decision to change the game for case study II as it resulted in almost full use of the artificial language ROILA. In comparison to the constrained language condition where usually the children did not initiate a dialogue with the iCat or found it difficult to constraint themselves, in the case of ROILA the children did take an initiative and started deliberating with the iCat.

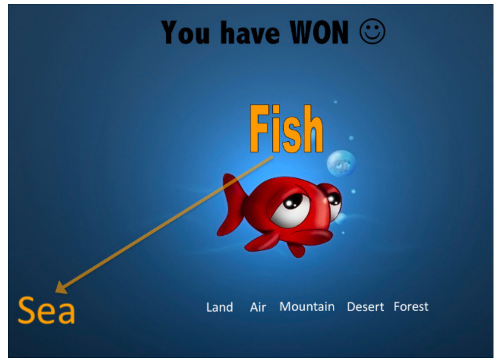

Fig. 4 Game II Example

We can conclude that at least in a wizard of oz situation children were not reluctant to learn new languages and there was no cognitive overload felt in doing so. Ultimately when a complete application is ready with speech recognition in place we can expect that children would be willing to invest some effort in learning a language that is optimized for speech recognition.

V. FUTURE WORK

Certain limitations do apply to our results, which we aim to eradicate by future research. Firstly, we aim to evaluate subjective satisfaction of artificial and constrained languages with speech recognition in place, as that would also affect the subjective experience e.g. an inaccurate speech recognizer would cause frustration to the user [2]. Other related factors are the tendency of users to hyper articulate when speech recognition does not work eventually leading to even worse recognition accuracy [23]. Therefore such issues must be kept in mind while extrapolating the behavior of participants between the two situations of with and without speech recognition in place.

The evaluation of the artificial language ROILA with regards to only speech recognition has been dealt with

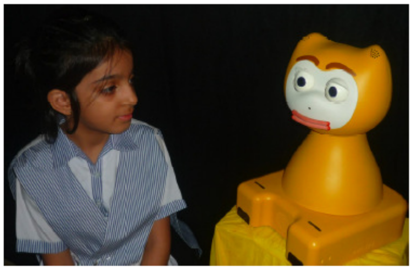

Fig. 5. Child interacting with the iCat: Case Study II

elsewhere [20]. Now it is a matter of combining both aspects and evaluating them together.

Secondly, we must acknowledge that children are rather easy to please and adults would be much more critical of any extra effort in learning new languages. We are aware that we conducted subjective evaluations only and given that children are rather easy to please it would be worthwhile to explore objective measures to determine the user acceptance of novel languages. Such objective measures are also mentioned here [24].

Thirdly, we must keep in mind the wow-effect of interacting with the iCat. It could have been that the children were mesmerized by the iCat and this clouded their judgments. This could potentially be solved by carrying out longitudinal evaluations.

Lastly, we adopted a gaming scenario which provided extra motivation to the children to learn the new languages. Therefore for future experiments similar scenarios could be utilized which provide an extra motivation and a gain factor for participants to learn new languages.

Acknowledgment

We would like to thank all the children, parents and teachers for their cooperation and assistance.

References

[1] M. A. Goodrich and A. Schultz, "Human Robot Interaction: A Survery," *Foundations and Trends in Human–Computer Interaction,* vol. 1, no. 3, pp. 203-275, 2007.

[2] B. Shneiderman, "The limits of speech recognition," *Commun. ACM,* vol. 43, no. 9, pp. 63-65, 2000.

[3] M. Forsberg, "Why is Speech Recognition Difficult?," 2003.

[4] J. Fry, H. Asoh, and T. Matsui, "Natural dialogue with the Jijo-2 office robot," in *In Proceedings of the 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems* Victoria, B.C., Canada, 1998, pp. 1278-1283 vol.2.

[5] G. E. Churcher, E. S. Atwell, and C. Souter, "Dialogue management systems: a survey and overview," University of Leeds, Leeds, UK, 1997.

[6] Z. M. Hanafiah, C. Yamazaki, A. Nakamura, and Y. Kuno, "Human-robot speech interface understanding inexplicit utterances using vision," in *CHI '04 extended abstracts on Human factors in computing systems*, Vienna, Austria, 2004, pp. 1321-1324.

[7] R. Rosenfeld, D. Olsen, and A. Rudnicky, "Universal speech interfaces," *Interactions,* vol. 8, no. 6, pp. 34-44, 2001.

[8] C. David, "The Cambridge encyclopedia of language," in *The Cambridge encyclopedia of language*: Cambridge University Press, 1997.

[9] P. Janton, *Esperanto: language, literature, and community*: State University of New York Press, 1993.

[10] T. Johnstone, "Emotional speech elicited using computer games," 1996.

[11] N. Lazzaro, "Why we play games: Four keys to more emotion without story," *Hämtad den,* vol. 18, 2006.

[12] A. Al Mahmud, O. Mubin, J. Octavia, S. Shahid, L. Yeo, P. Markopoulos, and J. Martens, "aMAZEd: designing an affective social game for children," in *Interaction Design for Children*, 2007, pp. 53-56.

[13] A. Al Mahmud, O. Mubin, J. Octavia, S. Shahid, L. Yeo, P. Markopoulos, J. Martens, and D. Aliakseyeu, "Affective Tabletop Game: A New Gaming Experience for Children," in *Tabletop*, 2007, pp. 44-51.

[14] J. Rubin, "What the" good language learner" can teach us," *Tesol Quarterly,* pp. 41-51, 1975.

[15] A. Breemen, X. Yan, and B. Meerbeek, "iCat: an animated user-interface robot with personality," in *Fourth International Conference on Autonomous Agents & Multi Agent Systems*, Utrecht, 2005.

[16] C. Bartneck, T. Kanda, O. Mubin, and A. Al Mahmud, "Does the design of a robot influence its animacy and perceived intelligence?," *International Journal of Social Robotics,* vol. 1, no. 2, pp. 195-204, 2009.

[17] S. Shahid, E. Krahmer, and M. Swerts, "Alone or Together: Exploring the Effect of Physical Co-presence on the Emotional Expressions of Game Playing Children Across Cultures," in *Fun and Games*, Eindhoven, the Netherlands, 2008, pp. 94-105.

[18] W. IJsselsteijn, Y. de Kort, and K. Poels, "The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games," *Manuscript in preparation,* 2008.

[19] K. Hone and R. Graham, "Towards a tool for the subjective assessment of speech system interfaces (SASSI)," *Natural Language Engineering,* vol. 6, no. 3&4, pp. 287-303, 2001.

[20] O. Mubin, C. Bartneck, and L. Feijs, "Using Word Spotting to Evaluate ROILA: A Speech Recognition Friendly Artificial Language," in *CHI 2010 Conference on Human Factors in Computing Systems*, 2010, pp. 3289-3294.

[21] A. Black, "Festvox". [Online]. Available: http://www.festvox.org/

[22] S. Tomko and R. Rosenfeld, "Speech Graffiti vs. Natural Language: Assessing the User Experience," in *Proceedings of HLT/NAACL*, 2004.

[23] E. Shriberg, E. Wade, and P. Price, "Human-machine problem solving using spoken language systems (SLS): Factors affecting performance and user satisfaction," in *Human Language Technology Conference*, 1992, pp. 49-54.

[24] O. Mubin, S. Shahid, C. Bartneck, E. Krahmer, M. Swerts, and L. Feijs, "Using Language Tests and Emotional Expressions to Determine the Learnability of Artificial Languages," in *CHI 2009 Conference on Human Factors in Computing Systems*, 2009, pp. 4075-4080.