Dialog Toys - A New Sector In Toy Market

Karina Matthes, Rico Petrick, Horst-Udo Hain

Linguwerk GmbH, Schnorrstr. 70, 01069 Dresden, Germany
{karina.matthes, rico.petrick, udo.hain}@linguwerk.de

Abstract

Automatic speech recognition (ASR) is currently used in some specific domains like interactive voice response systems (support hotlines), voice activated dialing (calling contacts from address books on mobile phones) or voice destination entry in car navigations systems. Several constraints like recognition rate (especially in noisy environments), computational effort, memory consumption or privacy aspects prevent ASR from a widespread application.

On the other hand, under well defined circumstances ASR can be quite handy, for instance if there is a resticted set of words to be recognised and the user can be guided by a dialogue system. Then also system requirements can be limited to a level where even embedded platforms may come into consideration.

In this paper we present a toy for pre-school children that uses speech input and output for communication. It takes the child to a journey into another world: Lingunia. Based on a story that is shown in a picture book the toy explains different topics like animals, colours, numbers, seasons etc. and integrates the child into the fictional situation. By asking for words and facts which previously have been mentioned, the child becomes part of the adventure and – on the fly – improves it's active vocabulary. It has even the opportunity to control the dialogue by choosing which game it wants to play or which part of the story should be described. The world's smallest ASR system Picard allows us to create a system that does not require internet connection for the recognition task because everything is performed on the embedded platform.

1. Introduction to Dialog Toys

Dialog Toys are educative toys with an integrated high-tech speech recognition. They enter a controlled dialog by talking to the children and requesting answers from them. Dialog Toys understand the children's speech and reply appropriately. Therefore the children not only passivley listen to the Dialog Toy, but also actively improve their vocabulary and knowledge. The first toy of this brand is called *Lingufino* and is dedicated to children of age 4 to 7 years.

In the toy store you will find two packages: the basic set and the expansion set. The basic set consists of the toy Lingufino, a hidden electronic unit inside Lingufino, a picture book that illustrates an adventure, and finally a pluggable memory module that belongs to the illustrated story. Additionally we provide expansion sets with several new stories which include a picture book and a memory module.

2. Embedded Speech Dialogue System

Lingufino (project is based on previous studies [1, 2, 3] and funded by [4, 5]) implements a speech dialogue system (SDS) running on an embedded microcontroller platform mounted hidden insight of the soft toy. The human machine interface (HMI) communicates via speech only, no additional modes such as keys, touch functions or displays are applied. The system consists of three parts (Fig. 1):

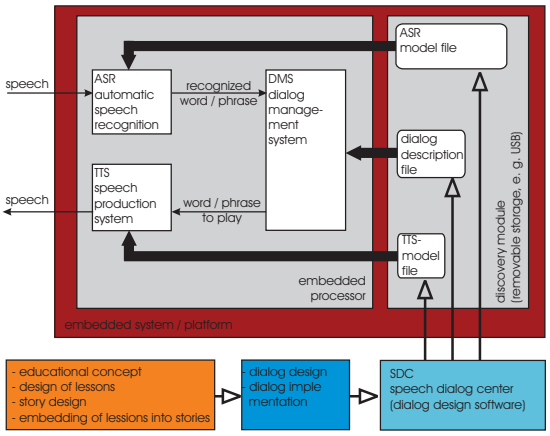

Figure 1: *Speech dialogue system and associated design process.*

- TTS – speech production, output is performed by the limited word/phrase based text-to-speech system using prompts that are prerecorded from a professional to achieve the best quality regarding intelligibility and prosody which is very important for the educational effect. One of the goals is to teach the child how to pronounce words, therefore the voice output must be as clear as possible using a pleasant voice with "human" prosody. This is currently not available using synthetic speech.
- ASR – speech input, an automatic speech recognition (ASR) system named *Picard ASR* is able to recognise words, phrases and sentences (64 in parallel). Its very low memory consumption (RAM: 15 KB, FLASH: 90 KB, CPU: 40 MHz) enables Picard ASR to run on very low price microcontrollers, which is necessary for the toy market but also for other consumer markets. It is a phoneme based Hidden Markov Model (HMM) recogniser using 64 shared GMMs, configurable tristate mono- or triphones. Feature extraction runs with 13 primary MFCCs and additional Δ and $\Delta\Delta$ features.
- DMS – the dialogue management system, runs dialogues that are given by dialogue description files. These include fixed dialogues but also dynamic structures driven by random processes to make the toy dialogue more alive. One task for the dialogue manager is to handle speech responses by the child according to the quality or correctness. If, for instance, the toy asks for a "red table", then the answer "table" is quite good (the child understood what has been asked for), but the answer "red table" is much better (the child can assign properties to objects). The DMS can reply with different levels of awards, and is on the other hand able to determine future questions or games based on the educational level of the child.

3. Didactical Approach

The Lingunia World of Learning is a fun interactive game for children aged four or older. It includes the soft toy Lingufino, which features modern speech technology and lovingly illustrated adventure books. Each book comes with a discovery module (a removable memory) which covers the speech dialogue configuration files (ASR model files, TTS model files, DMS description files) for the current adventure dialogue.

Lingufino speaks to players, recognises what has been said (up to 2,500 different words and phrases) and guides them through fantastic adventures. Over 1,500 voice responses encourage the player to interact through speech. The integrated language games are based on scientific findings on language development in early childhood. Intelligent dialogue and a variety of games make the Lingunia World of Learning a unique interactive voice gameplay experience.

The words and phrases used in the dialog for speech input and output are carefully selected from the vocabulary of preschool children. The dialog system is able to adapt to the educational background of the child and therefore can choose appropriate prompts and questions. This avoids overwhelming the child with the story or the dialog. Main focus of the toy is that the children have fun and like to play with it. The educational approach is a hidden for the child, the teaching takes place "by the way". Depending on the answers given by the child the DMS can decide which level of questions is best for this particular child. This allows for a story line that is both vivid and challenging.

4. Dialog between Child and Toy

Lingufino tells the story that is depicted in the book. An alive and interactive dialog guides through different story lines. Due to the different opportunities to respond to the questions the child experiences the adventure in different ways which keeps the toy fascinating.

Lingufino lives in an imaginary world called Lingunia together whith its friends Elfi and Edison. In the first story (basic set) they go on a journey in Lingunia and find some magic powder and a spell. In every new story (expansion set) they travel using this magic powder and the magic spell to the childish human world. There they have exciting adventures related to topics children are interested in like kindergarten, zoo, farm holidays and so on.

All stories contain quests (finding correct answers to several questions) and speech games (like rhymes or counting games) and have been developed in cooperation with pedagogues, educators, educationalists, speech therapists, media educators and so on. The teaching and learning content is represented survey-like for the parents on a separate page.

5. Design Workflow

The speech dialogue design process includes three competencies (Fig. 1): (i) educational siences, (ii) media siences, (iii) speech technology sciences.

(i) First the educational concept is developed. The question here is what should be taught or explained (colors, numbers, animals, etc.). Then the vocabulary has to be determined which can be used for children of the appropriate age. The words, phrases and sentences must not be too complicated, but should on the other hand explain the facts as accurate as possible.

(ii) Based on the vocuabulary and the content to be communicated, a story line is developed and then compiled into an adventure. Designers and illustrators create figures and pictures to display the story in a way that children can understand.

(iii) These stories are to be implemented as a dialogue structure using a dedicated dialogue design software tool SDC (Speech Dialogue Center). Dialog designer, educational expert and illustrator work hand in hand to decide what should be asked for, how it is asked for, and how it is depicted in the picture book. This is an iterative and time consuming process.

Later on speech recognition and speech production are to be designed for the given dialogue. The required speech prompts are recorded by a professional speaker supervised by the dialog designer. The ASR engine is configured to be able to recognise all phrases that are necessary for the dialog. At the end, SDC automatically generates all necessary configuration files which are stored on the memory module for the regarding adventure.

6. Conclusion

In this paper we introduced "Dialog Toys" as a group of toys that are able to enter a dialog with children. The first toy of that group is called Lingufino and is dedicated to pre-school children at the age of 4 to 7 years. It has a build-in hidden electronic unit that provides speech recognition, speech output and a dialog management. Every adventure consists of a story book together with a pluggable memory module. The toy guides the child through an interactive journey and motivates the child to answer questions related to the story that is told. This is on the one hand an exciting and funny taks for the child, and on the other hand it improves just by the way its pronunciation and active vocabulary.

The contents of the story, the dialog, the pictures and the appropriate speech prompts and phrases are designed by experts like educationalists, dialog designers, and speech therapists. The used vocabulary is adjusted to the knowledge of preschool children to avoid that the children are dissastisfied by the dialog.

7. References

[1] Jokisch, O., Hain, H.-U., Petrick, R. and Hoffmann, R., "Robustness Optimization of a Speech Interface for Child-Directed Embedded Language Tutoring," In Proc. of Workshop on Computer Child Interaction (WOCCI) 2009, Boston, USA, 2009, CD-ROM.

[2] Matthes, K., Claus, F., Hain, H.-U. and Petrick, R., "Herausforderungen an Sprachinterfaces für Kinder," (in German, engl. translation: Challenges on Speech Interfaces for Children), In Mixdorff, H. (Ed.): Electronic Speech Signal Processing 2010, Berlin, TUDpress, 2010, ISBN 978-3-941298-85-9, pp. 180 – 187.

[3] Claus, F., Gamboa Rosales, H., Petrick, R., Hain, H.-U. and Hoffmann, R., "A Survey about ASR for Children," In Proc. of Workshop on Speech and Language Technology in Education (SLaTE) 2013, Grenoble, Frankreich, 2013, pp. 26 – 30.

[4] Bundesministerium für Bildung und Forschung, 2015.

[5] European Social Fund and the Free State of Saxony, 2015.