COMUNICA - Tools for Speech and Language Therapy

William Rodríguez, Oscar Saz, Eduardo Lleida, Carlos Vaquero and Antonio Escartín
Communications Technology Group (GTC)
Aragon Institute for Engineering Research (I3A)
University of Zaragoza, Zaragoza, Spain
{wricardo,oskarsaz,lleida,cvaquero}@unizar.es, aescartinv@gmail.com

ABSTRACT

This paper introduces the systems and technologies used for the development of computer-aided speech and language therapy tools. These tools ("Pre-Lingua", "Vocaliza" and "Cuéntame") aim to help speech and language disabled people to improve their communication abilities, covering all the processes in the acquisition of the spoken language (from phonation and articulation to descriptive and comprehensive language). The applications are conceived with the idea of supplying an easy interface for speech therapy in any language, although focusing on the needs of speech therapists in Spain and Latin America. One of the key points in the applications is the possibility of automating the process of speech therapy thus the child who is under therapy can run it in an unsupervised way after a short time configuration done by the speech therapist. These tools require some improvements in Speech Technologies such as Automatic Speech Recognition (ASR) and pronunciation verification in order to help users to improve their communication skills.

1. INTRODUCTION

Language acquisition is one of the most relevant aspects in young children education. Language is the most powerful tool that human beings have to interact with their environment everyday, to achieve their goals and also for problem solving. In this way, delays in the language acquisition can involve a general delay in the development of the rest of the educative abilities of a patient with speech disorders. The situation becomes especially critical for children with physical handicaps or development disorders that have to deal with an increase in their barriers with the rest of the society.

Language therapy is a subject that requires professionals with a wide academic formation and an extensive experience, so there is a great difficulty in making that this small number of professionals fulfill the demand of their work. Thus, a single speech therapist can have under tuition several children with speech difficulties, every one of them requiring long, continuous and time-consuming sessions with the therapist. And all of this becomes more complicated in the world of special education, when the speech therapist has to deal with several speech impaired children who also have further educational needs.

2. TOOLS FOR SPEECH THERAPY

All the needs exposed could be directly checked by the authors in contact with the staff and educators of the Public School for Special Education (CPEE) "Alborada", located in Zaragoza (Spain). The professionals in the CPEE "Alborada" have always shown a special interest in the field of technical aids for the improvement of the quality of life of their pupils and all the handicapped community.

The development of tools for speech and language therapy that is currently being carried out under the framework of "Comunica", whose main goal is to provide the community of speech therapist with applications that can reduce the time they need to see each one of their patients by the automation of many of the activities they carry out every day. "Comunica" intends to be a long-term way of distribution of all the tools for speech and language therapy developed under its framework.

All the tools are distributed under a freeware license for the use of all the community of speech therapists and speech therapy users who could be interested in them[1]. All the applications in "Comunica" have been developed aimed to the community of speech therapists in Spain and Latin America by choosing the Spanish language as the initial language of operation in these tools.

"Comunica" works in three parts, first "PreLingua" which works the pre-language stage and include voice activity detection, intensity, breathing and intonation control, and vocalization; the second part "Vocaliza", which works on the the phonological, semantic and syntactic levels of language, and finally "Cuéntame" which works the pragmatic level of language.

2.1 "PreLingua"

"PreLingua" gathers a set of small games that use speech processing to train children with speech development disorders in order to assist the work in speech therapy oriented to phonation. A speech processing diagram like the one shown in Figure 1 is used for the training of five speech features in the games (voice activity, intensity, breathing, tone and vocalization).

All the games within the "PreLingua" framework do not require of any previous configuration and their educative value relies on the robustness of the speech processing like shown in Figure 1. Hence, "PreLingua" can dramatically reduce the time that a therapist spends with children affected with severe speech impairments.

[1]URL http://www.vocaliza.es

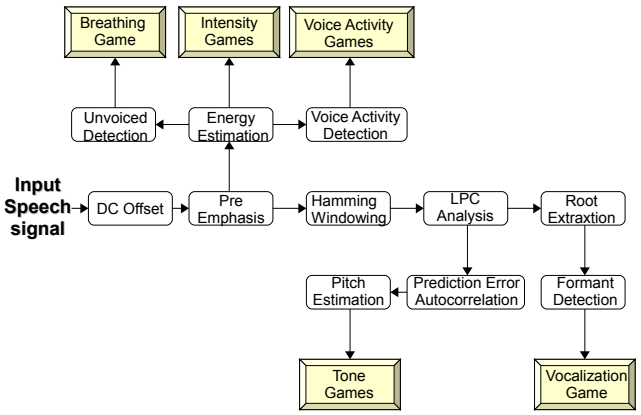

Figure 1: Speech processing blocks in "PreLingua".

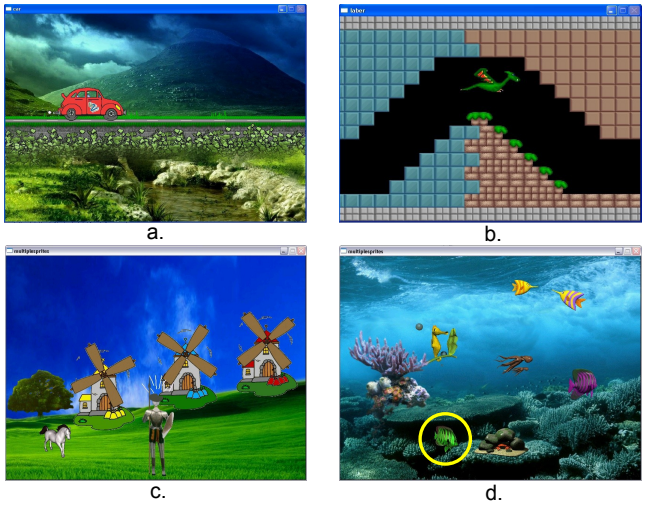

Figure 2: Games in "PreLingua"; a.Voice Activity, b.Intensity, c.Breathing and d.Tone.

2.1.1 *Voice Activity*

Voice activity games [3] are oriented to children with delays in their speech who still do not associate their production of sounds to changes in their environment. In this case, a game with a very simple interface is required because is oriented to children in very early stages of development. A speech technology called VAD (Voice Activity Detector) gives a binary signal (0 : *unvoice/silence*, 1 : *voice*), if voice is detected the 1 obtained will subsequently produce a reaction on the screen of the computer. For instance in Figure 2 part *a* shows a car which move when a value of 1 is obtained by the VAD system, at the beginning the car appears the left side, when voice is detected it moves to right side. The child will become aware of the possibility of change their environment by use of voice.

Extra functionality is obtained by these games for the early stimulation of children suffering the most severe development disorders. As with the rest of the "PreLingua" games, the voice activity detection games do not require any further configuration besides the normal configuration of the audio signal via a microphone.

2.1.2 *Intensity*

Intensity games are the next step in the development of speech therapy games. Once the child has acquired the ability to distinguish his own speech production, he has to learn to control the volume of his vocal production. In this situation, the energy estimation of speech signal is required, this value becomes in the vertical position of a cartoon. The feedback given by the game is the completion of a goal; as, for instance, reaching the end of a labyrinth in which the dragon is moved up and down accordingly to the intensity of the speech production (Figure 2 part *b*). A set of other games are also provided in "PreLingua" in which the only objective of the game is moving an image up and down accordingly to the sound intensity; these games are addressed to children with a higher development difficulties.

2.1.3 *Breathing*

Breathing is important for keep fluency communication. In order to induce a child to control breathing, the system analyze the energy of speech signal and only consider unvoiced segments. The Figure 2 part *c* shows that in presence of blowing the *Quijote* animate a windmill and the intensity of the signal will increase or decrease the rotation speed.

2.1.4 *Tone*

Tone games follow the same philosophy that intensity games in the fact that they require the user to control a character on screen raising and lowering his fundamental frequency or pitch. Control of tone is required in a correct speech production and it is extremely needed in some speech features like prosody. In this case, a mathematical technique called Linear Prediction Coefficients (LPC) [4] is required to separate the influence of the glottal pulse from the articulation features and the vocal tract. Once the LPC analysis is done, the system obtains an estimation of the fundamental frequency $F0$ and becomes in the vertical position of diferent cartoons. An application like the one presented does not require of a fine pitch tracking algorithm because the user is told to utter a long voiced sound like a vowel to obtain the pitch estimation. The games provided in "PreLingua" include the movements of a butterfly or fish controled by pitch ($F0$), in Figure 2 part *d* the green fish (yellow circle), is controlled by voice's child and has to interact with the others animals modifying their tone.

2.1.5 *Vocalization*

The transition between phonation and articulation in small children initially occurs with vowels. The set of vowels for every language is unique, so the strategy in the vocalization games has been to make the development only for the vowels in Spanish, although expansions to any other language could be done by defining the representations of the vowels of that language to the first and second formants ($F1$ and $F2$) space [4]. Spanish contains five vowels (/*a*/, /*e*/, /*i*/, /*o*/ and /*u*/ in their SAMPA notation) whose representation in the space of the formant frequencies $F1$ and $F2$ is a triangle.

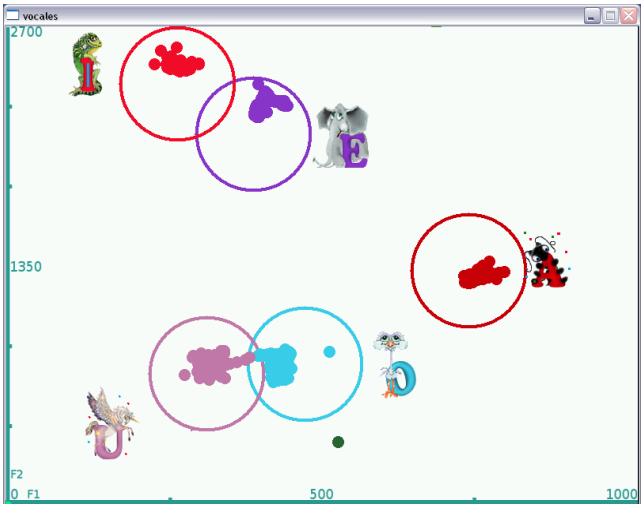

Figure 3: Vocalization Game.

One key point in the correct detection of vowel production refers to vocal tract length normalization. The canonic values of $F1$ and $F2$ for the five Spanish vowels are given by several authors but these values are purely theoretical as every user of the game will have a different value of the formants for his production of vowels. Hence, it is necessary to make an estimation of the vocal tract length to normalize the calculated formants by this value and make them fit to the canonic values of $F1$ and $F2$ for every vowel. Once this little step is taken, the game is ready to be used by the child in an unsupervised way, unless it is noticed that it is needed to re-adjust the vocal tract length.

In the vocalization game, the idea is encouraging the user to utter each vowel and the system draws a dot according to $F1$ and $F2$ estimation, if the pronunciation is close to the correct form the system uses a similar color to the vowel in order to draw a dot and will animate a cartoon, otherwise, the dot will be draw with different colors and the cartoon won't animate. The Figure 3 shows the results of pronounce the five vowels of correct form.

2.2 "Vocaliza"

This application is oriented to the speech training of the articulation abilities of the patient in isolated words and short sentences. While focusing mostly on the articulatory side of the language, it also introduces the user to the semantics and syntax levels of language with different activities [5] .

There are three main parts in the application:

- The configuration or set-up interface is the way in which the therapist creates the profiles for the different users of the application. These profiles contains the different words and sentences that every used will be prompted to utter accordingly to the specific needs of the child. It also contains information about the use of images, sound and text during the games for that user. As it will explained in Section 2.2.3, different Augmentative and Alternative Comunication (AAC) systems are used to adapt the user interface to the educational needs of every user. Acoustic information is also stored for the creation of speaker dependent acoustic models to be used in the Automatic Speech Recognition (ASR) system.
- Once a user profile is created, the core of the application is the set of four games developed for speech and language therapy and the Speech Technologies embedded in the application to make the user improve the language. The games are described in Section 2.2.1 and the Speech Technologies in Section 2.2.2.
- Finally, the user interface just require a speech input from the patient. Different games as well as the feedback given by the core of the application will keep appearing as the user keeps advancing through them. With this design, no supervision from the speech therapist is required for the child to practice language.

The user interface, the games for speech and language therapy and the Speech Technologies in "Vocaliza" are explained in the following sections:

2.2.1 Speech training and games

To make speech training attractive for children (main targets of the applications) "Vocaliza" exercises three levels of the language (phonological, semantic and syntactic) presenting different games.

Phonological level is exercised encouraging the user to utter a set of words previously selected by the speech therapist or educator during the configuration procedure to focus on the special needs of every user. The application runs an ASR decoding on the utterance to accept or reject it and evaluates the accepted utterances displaying a grade as the final outcome of the game. Two games are designed for training the phonological level: In the first one, the user is prompted to utter a word like in Figure 4 (candy), and in the second one, the user can utter freely any of the words set up by the therapist in the user profile and the word appears on screen.

Semantic level is exercised presenting a riddle game previously defined by a speech therapist or educator. The application asks a question to the user and gives three possible answers. The user must utter the correct answer and the ASR system must accept it to go on with the next riddle. The application will show again a grade depending on the ability of the user to solve the riddle.

Syntactic level is exercised encouraging the user to utter a set of sentences, previously selected by a speech therapist or educator. Again, the application will use ASR decoding to accept the input utterance of the user and in that case will evaluate user utterances to display a grade, marking the improvement of the user.

2.2.2 Speech Technologies for speech and language therapy

The Speech Technologies provided by the core of "Vocaliza" are ASR, speech synthesis, acoustic user adaptation and pronunciation verification.

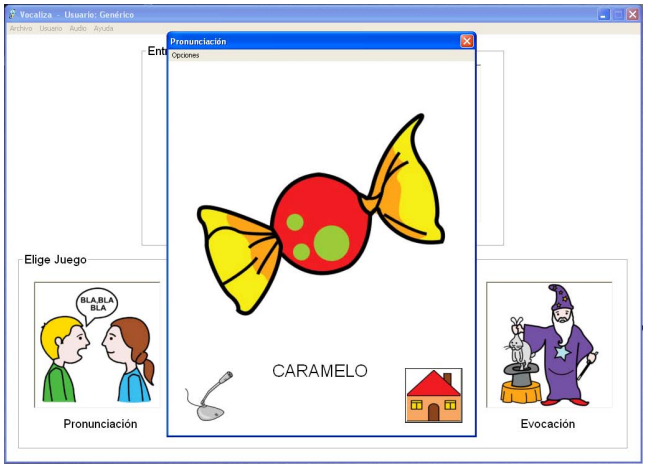

Figure 4: "Vocaliza" Application.

ASR constitutes the main technology of the application. Speech therapy games require ASR to decode user utterances, and to decide which word sequence has been pronounced so that the application will be able to let the user know if the game has been completed successfully. ASR is hence the first way in which pronunciation verification takes place in "Vocaliza" as it accepts or rejects utterances of the user. A good performance of the ASR in the application is, then, strongly required.

Speech synthesis provides a way to show the user how a word or sentence should be pronounced, which is useful in speech therapy games. As soon as a speech therapist adds a new word, sentence or riddle to the application, it is able to synthesize a correct Spanish utterance of the corresponding word, sentence or question. For further functionality, "Vocaliza" allows speech therapists to record their own utterances to be used instead of speech synthesis.

Speaker adaptation enables the application to estimate speaker dependent acoustic models adapted to each user. "Vocaliza" uses Maximum A Posteriori (MAP) [1] estimation. Speaker adaptation is strongly required for obtaining the full performance of the application since impaired speech can reduce dramatically the performance of ASR, so that users suffering severe speech impairments the system would not be able to obtain good results.

Pronunciation verification is the way in which the application provides an evaluation in the improvement of user communication skills. "Vocaliza" uses a Likelihood Ratio (LR) based Utterance Verification (UV) procedure [2] to assign a measure of confidence to each hypothesized word in an utterance.

To achieve this, the application uses a speaker independent acoustic model, which is assumed to model correct speech, as target hypothesis, and a speaker dependent acoustic model, which is assumed to be adapted to impaired speech, as alternate hypothesis. Therefore, this measure of confidence involves a relative evaluation method to quantify improvement of the user articulation skills.

2.2.3 User interface and AAC systems

Augmentative and Alternative Comunication (AAC) systems and an easy user interface are the way in which "Vocaliza" motivates children to enjoy the application while practicing their speech. The use of text, images and sounds (via synthesized speech or pre-recorded audio) reinforces in the patient the concept and correct pronunciations of the word or sentence that is presented.

The final evaluation given after every act within a therapy session (every word or sentence pronounced or every riddle correctly answered in a round) also relies on these three interaction ways. A grade in text will be presented to the user in a six levels scheme from 'average' to 'perfect' altogether with an animated gift with sound that reinforces the feeling of 'goodness' in the way the user has gone through the game.

The use of these three interactive ways of presenting the game proposal and the outcome of it are aimed to achieve a total accessibility by every user:

- Image and sound reinforce the games interface for users who haven't developed reading yet due to their development disabilities.
- Image and text provide feedback to children who may suffer hearing impairments.
- Finally, the use of sound and big font text and images will allow children with visual impairments to take full advantage of the potential of the application

Also, configuration of the application to be used with different patients is made in a simple way to help speech therapists work with different patients. Every patient can have a different user profile that stores the words or sentences that he needs to work with as well as acoustic information about his disorder, reflected in an speaker dependent model obtained with his speech in a previous adaptation phase and necessary to make full use of the possibilities of "Vocaliza".

2.3 "Cuéntame"

"Cuéntame" is the latest of the speech and language therapy tools developed under "Comunica". This tool aims to help children with delays in the acquisition of the oral language to improve their communicative skills and follows the same philosophy that "Vocaliza". Hence, this application also relies strongly in a robust but simple user management and configuration and in the need of using AAC systems to reinforce the correct use of language in the patients, as well as in the proven performance of Speech Technologies in speech and language therapy. As "Vocaliza", "Cuéntame" is intended to allow children to interact with the application in an unsupervised way after a short time of configuration within the application by the speech therapist.

2.3.1 Speech training and games

Three games are designed into the application. All of them share the same vision on the initial approach that consists in scenarios of growing difficulty that the user has to solve via speech.

Figure 5: "Cuéntame" Application.

The first game is a question answering game, where the used is asked a question about an image and has to give the correct answer to progress within the game. The questions are inserted into the application by a therapist who also configures the correct answer to the question. The second game asks the user to make a description of an object shown in the screen accordingly to a given group of attributes (shape, color, etc...); the user has to utter his description of the object until filling up all the attributes. The proposed scenario and the attributes that the patient has to use to correctly describe the object or situation are configured by the speech therapist. The third game is designed as an action-adventure game in which the user interacts freely with a scenario to complete a given goal like in Figure 5. The interaction is purely by speech where the user utters the desired actions and the application gives feedback about the outcome of those actions. Possible actions are taking and using objects, moving around the scenario and carrying out pre-defined actions. This game resembles an oral command control interface in which the user had to be acting over different elements in the environment.

With these games, the patient is taught in the usefulness of language to learn things, describe the environment or achieve goals. Within this idea, the application also aims to stimulate the correct production of full sentences. Hence, a grade is given to the user in every game when he achieves the goal in every scenario (answering or describing), but a better grade is achieved by the user the more the sentence uttered is close to following the structure of $Subject+Verb+Predicative$ in the sentence.

2.3.2 *Speech Technologies for speech and language therapy*

"Cuéntame" share with "Vocaliza" a similar use of Speech Technologies regarding ASR, speech synthesis and speaker adaptation. But, further technologies are required to deal with two new issues in "Cuéntame"

On one side, the control and rejection of Out-Of-Vocabulary (OOV) words is one of the key points. As the user is encouraged to utter freely any sentence, an open ASR system is required inside the application than just isolated words as in "Vocaliza". Hence, rejection of OOV words is necessary via a confidence measure. Every word decoded by the ASR system is evaluated under the confusion network (using a zerogram model) and words whose phonemes obtain an averaged low score are rejected. On the other side, the use of a correct language modeling is also required to avoid ASR errors during the games procedure.

The language model obtained in the question answering and description games is obtained directly from all the information introduced in the configuration process (question and answer or attributes and description) (missing parts) and requires a syntactic analysis of the question in itself to obtain the subject and the verb of the question and create a whole sentence using the answer introduced by the therapist.

3. RESULTS AND CONCLUSIONS

In this paper, a set of tools and technologies for speech and language therapy have been presented under the framework of "Comunica". These tools are oriented to children who need to train their articulatory ("Vocaliza"), descriptive ("Cuéntame") and phonatory ("PreLingua") abilities.

All the applicactions in "Comunica" are in use at a special education school in Zaragoza. They have been very successful among the children and speech therapists, the evaluations by the group of speech therapists shows that these informatic applications are a useful tool for the training of children with speech disorders at several levels of the language.

The therapists also evaluate positively the easiness of use of the applications. The results are very encouraging to keep working in this direction as it is planned improve the functionality and robustness of informatic applications. The adequate development of pre-Language and Language improves the quality of life of individuals with speech disorders and enable them to use computers by means of multimedia applications.

4. ACKNOWLEDGMENTS

This work was supported under TIN-2005-08660-C04-01 from MEC of the Spanish government and Santander Bank scholarships. The authors want to acknowledge José Manuel Marcos, César Canalís, Pedro Pegero and Beatriz Martínez from the CPEE "Alborada" for introducing them to the world of the handicapped community and for their ongoing collaboration in this work.

5. REFERENCES

[1] J.-L. Gauvain and C.-H. Lee. Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291–298, 1994.

[2] E. Lleida and R.-C. Rose. Utterance verification in continuous speech recognition: Decoding and training procedures. *IEEE Transactions on Speech and Audio Processing*, 8(2):126–139, March 2000.

[3] W.-R. Rodríguez, C. Vaquero, O. Saz, and E. Lleida. Speech technology applied to children with speech

disorders. In *Proceedings of the 4th Kuala Lumpur International Conference on Biomedical Engineering*, Kuala Lumpur, Malaysia, June 2008.

[4] R.-C. Snell and F. Milinazzo. Formant location from lpc analysis data. *IEEE Transactions on Speech and Audio Processing*, 1:129–134, 1993.

[5] C. Vaquero, O. Saz, E. Lleida, and W.-R. Rodríguez. E-inclusion technologies for the speech handicapped. In *Proceedings of the 2008 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas (NV), USA, April 2008.