EasyVoice: Breaking barriers for people with voice disabilities

Paulo A. Condado and Fernando G. Lobo

DEEI-FCT, University of Algarve
Campus de Gambelas, 8000-117 Faro, Portugal
{pcondado,flobo}@ualg.pt

Abstract. Text-to-speech technology has been broadly used to help people with voice disabilities to overcome their difficulties. With text-to-speech, a person types at a keyboard, the text is synthesized, and the sound comes out through the computer speakers.

In recent years, Voice over IP (VoIP) applications have become very popular and have been used by people worldwide. These applications allow people to talk for free over the Internet and also to make traditional calls through the Public-Switched Telephone Network (PSTN) at a small fraction of the cost offered by traditional phone companies.

This paper presents an application, called EasyVoice, that integrates different systems to allow a person with motor impairment and voice disabilities to talk with another person located anywhere in the world.

Key words: Voice Disabilities, Virtual Keyboards, Text-to-Speech Synthesis, Skype, VoIP

1 Introduction

Information and communication technologies are becoming an integral part of daily life. These new technologies are changing all sectors of our society, and have been used to improve the quality of life of people with physical disabilities. With the advances in computing power, new applications have been developed for helping people with disabilities. Voice synthesizers are a good example of such applications. Voice synthesis quality and the associated computational resources have advanced considerably in the last years [3, 8], and have been a key factor for a better integration of people with voice disabilities in society.

Oftentimes people need to communicate with someone else that is at another location. Unfortunately, people with voice disabilities sometimes can hardly use a regular or mobile phone in order to have a conversation with another person. A minor voice disability might not be a big problem, but for those with severe voice disabilities, a phone conversation is something almost impossible to achieve. This limitation can be surpassed by combining Voice over IP (VoIP) technology with voice synthesizers and text input methods for helping people with disabilities. This solution has been hinted by ourselves in the recent past [4, 6] and is something that to our best knowledge has not been tried before.

This paper presents an application, called EasyVoice, that integrates different systems to allow a person with motor impairment and voice disabilities to talk with another person located anywhere in the world. User interfaces for people with disabilities, speech synthesizers, and VoIP applications, are all technologies that exist for some time. What we did was to marry them together, and by doing that, we believe we have created an innovation, something that did not exist before and that opens a window for a new world of communication, learning, and socialization, for people with voice and motor disabilities.

The paper is organized as follows. The next section describes background material which is relevant to the present work. Section 3 describes the EasyVoice system and its architecture. Section 4 outlines directions for future work, and Section 5 summarizes and concludes the paper.

2   Background

This section is divided into three parts: Text-to-Speech, virtual keyboards, and voice over IP applications.

2.1   Text-to-Speech Technology

Text-to-Speech (TTS) technology is used to convert text to voice using synthesis algorithms. These algorithms generate sounds that simulate those created by human vocal cords. The conversion is a difficult process and various complex techniques are required to produce an intelligible and natural output [9, 13, 8]. The two primary techniques for generating synthetic speech waveforms are concatenative synthesis and formant synthesis. Concatenative synthesis is based on the concatenation of segments of recorded speech. Formant synthesis does not use pre-recorded human speech samples at runtime. Instead, the synthesized speech is generated by an acoustic model.

The utilization of speech synthesis as an assistive technology has been quite popular, allowing barriers to be removed for people with a wide range of disabilities. Application examples include the use of screen-readers for people with visual impairment, as well as to help people with dyslexia and other reading difficulties. Another important utilization is to aid those with severe voice disabilities.

Sometimes, people with voice disabilities also have severe motor impairments. These people need non-standard text input methods to be able to type at a reasonable speed. For example, many individuals cannot control their hands with enough accuracy to use a regular computer keyboard, and sometimes only have the ability to control a single touch button.

2.2   Virtual Keyboards

Virtual keyboards are a reasonable solution to solve some of these limitations. Most virtual keyboards have a set of features to accelerate the writing process.

Keyboard group scanning, word completion, and word prediction systems are among the most common features to achieve that.

With a scanning system, a set of options is presented to the user on the computer screen, and a visual cursor advances through the options, one at time, at a specified time rate. The user responds by pressing a touch button whenever the cursor is on top of the desired option. Sometimes an option is just a container for more options, in which case the options are organized in a hierarchical fashion. Each container option is often referred to as a group option. For example, when a particular group option is selected, the scanning system immediately focuses on the sub-options of that group, and again, advances the visual cursor through each of them [7, 5, 12].

Word completion techniques are also used to accelerate the writing process. Most mobile phones have these features incorporated to allow people to type text messages faster. When the user types a letter, or a sequence of letters, the system shows a list of possible words that have that sequence of letters as a prefix. The system does that by searching in a built-in dictionary, which can also be updated with new words. This feature allows the user to select the desired word without having to type every single letter of it, decreasing the time to write a message [2].

2.3   Communication with Voice over IP

Voice over IP (VoIP) applications have become very popular during the last few years. VoIP has established a commercial niche after the mid-1990s [14]. VoIP is a useful technology for millions of people worldwide. With VoIP applications people can call each other at a very small fraction of the cost of a regular phone call. At present time, one of the most popular VoIP applications is Skype[1], which provides VoIP and instant messaging services, and works behind firewalls and Network Address Translators [10].

3   The EasyVoice System

We have reviewed various technologies in the previous section. Having them in mind, we thought about integrating voice synthesis with Skype, and provide an appropriate user interface for people with voice disabilities as well as motor impairments. We named the system EasyVoice. As opposed to traditional TTS interfaces, EasyVoice injects the sound directly through the network rather than sending it to the computer speakers. EasyVoice achieves this by working together with Skype via its Application Programming Interface (API)[2].

The EasyVoice application is freely available[3]. Any speech synthesizer can be used with it as long as it is SAPI compliant. SAPI stands for Speech Application

---

[1] See http://www.skype.com/

[2] See https://developer.skype.com/

[3] It can be downloaded from http://w3.ualg.pt/~pcondado/easyvoice/

Programming Interface. It is an API developed by Microsoft to allow the use of Speech Recognition and Speech Synthesis within Windows applications.

Notice that one could think that there is no need to do any special integration between a speech synthesizer and a VoIP application. After all, one could simply use a speech synthesizer and let the computer's microphone be close enough to the computer speakers. Such a naive solution, however, yields an excessive amount of echo during the conversation because the person at the other end of the line hears back her own voice. Someone could also think that this system would be unnecessarily complex for a user trying to communicate with a remote user seated at a PC where text-to-text communication would be in principle more efficient. However, we cannot forget an important thing. People seated at a remote PC represent most Skype users, but certainly not all. Many Skype conversations are between a person sitting at a computer and another person holding a regular phone without any text messaging facilities at all. An important aspect of EasyVoice is that it allows a person with voice disabilities to talk with another person who doesn't even know how to use a computer – the other person just needs to have a regular phone.

3.1   User Interface

A person with voice disabilities usually also has motor disabilities. That's often the case of people with cerebral palsy. In such cases, typing at a regular keyboard can also be a difficult thing to do, and that is an obstacle for a smooth utilization of a speech synthesizer. Because of that, special purpose user interfaces need to be considered. To accelerate the writing process, EasyVoice provides four main features:

- an archive of recent messages.
- a word completion system.
- an abbreviation system.
- an optional virtual keyboard.

With EasyVoice, users can choose if they want to write directly on a text box area or to open a virtual keyboard. By making the utilization of the virtual keyboard an optional feature, rather than a mandatory one, the user interface becomes more flexible and useful for a wider audience. Those who have voice disabilities but do not have severe motor problems can choose to type in the normal way. The virtual keyboard, on the other hand, can be selected for helping users with motor disabilities, or to allow the application to be used on touch screen machines.

Figure 1 shows a screenshot of the EasyVoice user interface with all the features turned on. There is a text panel on the top left part where the user can type. On the top right part there are two list boxes. The first one is a list of possible words given by the word completion algorithm, and the other one is a list of recently typed messages. There is also a virtual keyboard on the bottom part.
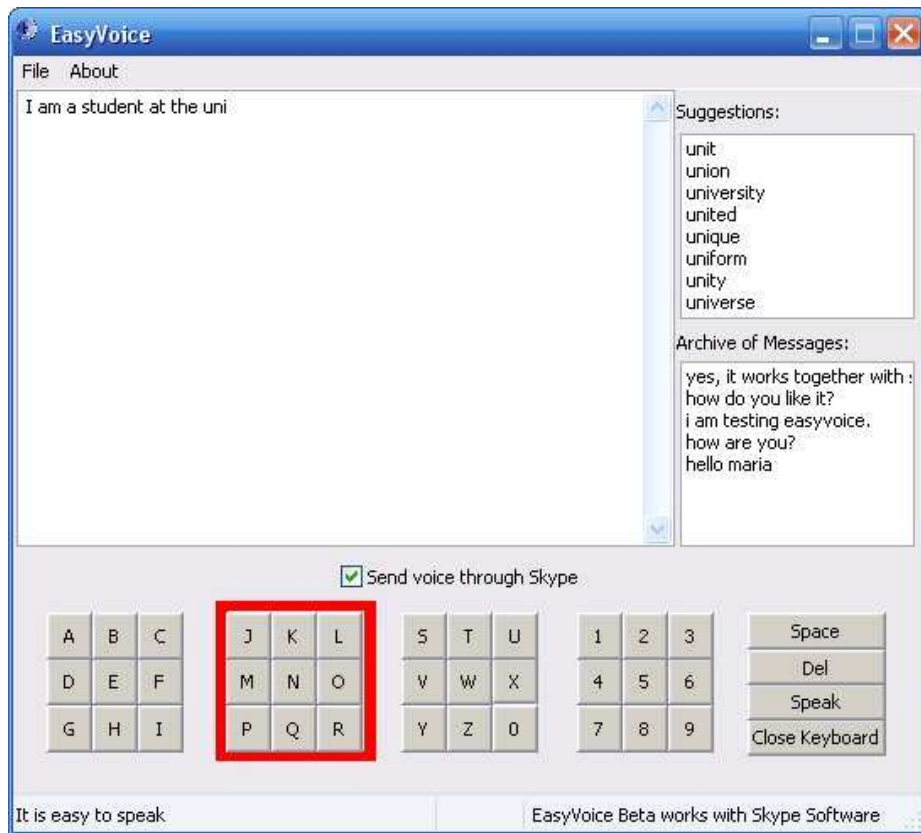
Fig. 1. The EasyVoice user interface. The top left part shows a panel where the input text appears. On the top right part there are two list boxes. The first one is a list of possible words given by the word completion algorithm, and the other one is a list of recently typed messages. Notice the virtual keyboard on the bottom part of the screenshot.

The word completion algorithm searches in a built-in dictionary for those words that have as a prefix the sequence of letters typed so far by the user. The system gives a list of the 8 most frequent words of the language, which were obtained by the British National Corpus[4].

The archive of recent messages is a very important feature because during a conversation it is many times necessary to repeat some words or phrases, because the other person at the end of the line might have not heard the sentence well enough. With the archive in hand, the user does not need to retype the message and can simply pick it from the list again.

Another important feature is the abbreviation system. It is common for people to use abbreviations when writing. It is something that is very popular in

---

[4] See http://www.kilgarriff.co.uk/bnc-readme.html

Fig. 2. An example of the scanning system. There are 5 groups of keys, and the cursor advances through each group at a specified time rate.

Fig. 3. In this example, the second group was selected, and now, the scanning system is focused on the keys of that group, with the cursor advancing through each of the 9 letters (J,K,L,M,N,O,P,Q,R), one at a time.

instant messaging software, especially among young people. For example, in English it is common for people to use "`btw`" as an abbreviation of "`by the way`". Within EasyVoice, the user can define his own abbreviations. The system will automatically replace each abbreviation by the corresponding full spelled words, before sending them to the speech synthesizer.

The EasyVoice TTS interface does not use the QWERTY keyboard layout, a standard for keyboard design, because many studies reveal that such a layout is not optimized [15]. We have used an alphabetic layout that seems to be more natural for people with disabilities, and we have two goods reasons to believe on that. First, disabled people have no experience with standard input devices and their minds are open to explore new concepts. Secondly, this kind of configuration is broadly used in mobile devices.

The keyboard has a group scanning mechanism like the one described in the backgroud section regarding virtual keyboards, and which we describe again schematically in Figs. 2 and 3.

4   Future Work

We have much work ahead to improve the EasyVoice system so that its utilization can be widespread worldwide.

We have done preliminary usability tests with a few people with cerebral palsy, all of them having voice disabilities. For future work, we plan to conduct more extensive usability tests with a larger population of users. We also plan to port EasyVoice to other operating systems besides Microsoft Windows, namely Linux and MacOS.

Finally, there are people with motor impairments which are so severe, that they cannot even press a single switch. Some people are only capable of moving their head, and others can only move their eyes. For those people, a user interface like ours is of little use, and we need to investigate alternative interfaces for those cases. Some research has been done in this area [11].

5   Summary and Conclusions

This paper presents EasyVoice, a system that combines different technologies in a novel way for helping people with disabilities. The development of EasyVoice can be seen as a constructionist learning experience. Just as children use LEGO blocks to create new toys, we have joined technologies to create a new tool with a new functionality.

Sometimes, the technology exists and is simply waiting for someone to have new ideas for its utilization. Most of the time, an innovation emerges from the combination of ideas. A good example of that is the creation of the World Wide Web by Tim Berners-Lee. The Web came upon existence by joining two technologies that already existed for quite some time: (1) the TCP/IP protocol suite, and (2) the notion of Hypertext. The development of TCP/IP, on top of which the Web rides, already happened in the 1970s, providing a general communications infrastructure for linking computers together worldwide. Likewise, the notion of Hypertext, in which the reader is not constrained to read in any particular order, but instead can follow links that point directly to other parts of the document, has also been developed in the 1960s. What Berners-Lee did, according to his own words, was to marry the two notions together [1].

Our invention is nothing compared with the creation of the Web, but it has in itself many things in common with its creation. User interfaces for people with disabilities, speech synthesizers, and VoIP applications, are all technologies that already existed for quite some time. What we did was to marry them together, and by doing that, we have created something that to our best knowledge has not been tried before and that opens a window for a new world of communication, learning, and socialization, for people with voice and motor disabilities.

References

1. Tim Berners-Lee. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor.* Harper San Francisco, 1999.
2. Philippe Boissiere. An overview of existing writing assistance systems. In *Proceedings of the IFRATH Workshop 2003*, Paris, France, 2003.

3. P. Condado, F. Tomaz, H. Shahbazkia, and F. Lobo. Information and communication technologies for special needed people: A case study with a student with cerebral paralysis. In *Advances in Technology-Based Education: Towards a Knowledge-Based Society*, volume 3, pages 1470–1474, Badajoz, Spain, 2003.

4. P. A. Condado and F. G. Lobo. Breaking barriers for people with voice disabilities: Combining virtual keyboards with speech synthesizers, and VoIP applications. UAlg-ILab Report No. 200604, University of Algarve, Faro, Portugal, 2006. Also as arXiv report No. cs.CY/0606088.

5. P. A. Condado, P. F. Miquelina, S. Norte, N. Castilho, F. G. Lobo, and H. R. Shahbazkia. Information and communication technologies for people with disabilities. In *Interactive Computer Aided Learning International Conference*, Carinthia Technology Institute, Villach, Austria, 2004.

6. Paulo A. Condado and Fernando G. Lobo. EasyVoice: Integrating voice synthesis with Skype. In *ASSETS'07: Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 235–236, New York, NY, USA, 2007. ACM Press.

7. P. W. Demasco and K. F. Mccoy. Generating text from compressed input: An intelligent interface for people with severe motor impairments. *Communications of the ACM*, 35(5), 1992.

8. T. Dutoit and Y. Stylianou. Text-to-speech synthesis. In R. Mitkov, editor, *Handbook of Computational Linguistics*, chapter 17, pages 323–338. Oxford University Press, 2003.

9. Paulseph-John Farrugia. Text to speech technologies for mobile telephony services. Master's thesis, Department of Computer Science and A. I., University of Malta, Malta, 2005.

10. S. Guha, N. Daswani, and R. Jain. An experimental study of the skype peer-to-peer VoIP system. In *IPTPS'06: Proceedings of the 5th International Workshop on Peer-to-Peer Systems*, Santa Barbara, CA, USA, 2006.

11. P. Majaranta and Kari-Jouko Räihä. Twenty years of eye typing: systems and design issues. In *ETRA'02: Proceedings of the 2002 Symposium on Eye Tracking Research & applications*, pages 15–22, New York, NY, USA, 2002. ACM Press.

12. P. F. Miquelina, P. A. Condado, C. L. Carvalho, H. R. Shahbazkia, and F. G. Lobo. Toque de voz: sistema de síntese de voz com um teclado virtual para o auxílio de pessoas com necessidades educativas especiais. In *RiBiE-2004: 7th Ibero American Congress in Computers and Education*, pages 650–659, Monterrey, Mexico, 2004.

13. J.P. van Santen, R. Sproat, J. Olive, and J. Hirschberg. *Progress in Speech Synthesis*. Springer, 1996.

14. U. Varshney, A. Snow, M. Mcgivern, and C. Howard. Voice over IP. *Communications of the ACM*, 45(1):89–96, 2002.

15. S. Zhai, M. Hunter, and BA. Smith. The metropolis keyboard: An exploration of quantitative techniques for graphical keyboard design. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 119–128, San Diego, California, 2000.