

Progress to a VOCA with Prosodic Synthesised Speech

Jan-Oliver Wülfing and Elisabeth André

Human Centered Multimedia University of Augsburg,
Universitätsstrasse 6a, 86159 Augsburg, Germany,
[wuelfing]/[andre]@hcm-lab.de,
WWW home page: <http://www.hcm-lab.de>

Abstract. Providing people, who cannot or almost not articulate themselves vocally, with a Voice Output Communication Aid (VOCA) with prosodic features would allow them to express their daily needs and intentions emotionally. We outline first steps towards such a prosodic VOCA, the EmotionTalker.

Keywords: AAC, Synthetic Speech, Prosody, Emotion

1 Introduction

If individuals with Complex Communication Needs (CCN) are not able to express themselves vocally, they mostly rely on methods of Augmentative and Alternative Communication (AAC). Methods to communicate that are given to them at hand typically include high-tech devices, such as Voice Output Communication Aids (VOCA). Customary VOCAs sound natural and comprehensible, but what they do not provide to individuals with CCN is the option to change the emotional content of speech dynamically (Higginbotham 2010), (Hoffmann & Wülfing 2010).

Our overall objective is to provide individuals with CCN with a means to convey their ideas and needs in an expressive manner during their daily routine (e.g. at work, at school, or in leisure). Such a prosodic VOCA is supposed to elicit a higher degree of attention from the interlocutors. Based on these considerations, we implemented a VOCA with prosodic features, the EmotionTalker.

2 Motivation

Attitudes towards individuals with CCN using an conventional VOCA are discussed by Mullenix and Stern (2010). Their work suggests that non-disabled people have negative attitudes towards individuals with CCN (e.g. less liked and less accepted). Since they have to invest more effort when listening, they tend to be more reserved towards individuals with CCN. This puts the individual with CCN in a more excluded situation than her or his non-disabled peer.

Breen (2014) addresses the importance of an expressive voice in Text-to-Speech (TTS) systems. He has shown that, at least, for conversational agents, the interlocutor would prefer a more dynamic style of the TTS when listening. Also Portnuff (2006) - a VOCA user - puts forward that a prosodic VOCA, which is, in its basics, an expressive TTS, would have benefits for both the individual with CCN and their interlocutors.

Although the topic of emotional utterances is hardly addressed in research on AAC (Pullin & Hennig 2015), there are a couple of innovative applications to be mentioned:

ExpressivePower^{TM1} co-developed by AssistiveWare B.V. and Acapela Group Babel Technologies SA enables individuals with CCN to create buttons with special emotive expressions and sounds, such as whining or questioning tones. This application has been developed with a particular focus on children using ProLongo2Go, a symbol-based communication application.

The VOCA, Tango², developed by BlinkTwice Inc - it is no longer being manufactured - offered the option to convey tones as well, i.e. it allowed individuals with CCN, for example, to select a whispering or shouting voice. Shouting is a very rare VOCA feature despite the fact that vocally speaking individuals often have to speak in a louder manner (for example, in a cafeteria or pub). Tango gave individuals with CCN also the possibility at hand to save the pronunciation for each word in a dictionary.

3 User Sensitive Inclusive Design

Our research target is the design of a prosodic Voice Output Communication Aid for individuals who use high-tech devices in order to communicate as they want to modulate their utterances. Since our target group is very diverse in respect to the characteristics of their impairments, a traditional user-centered design approach does not apply. Inspired by an approach developed Newell (Newell et al. 2011) called the User Sensitive Inclusive Design, we tried to develop a more empathic view - being more sensitive - while working with the target group, rather than treat them as “subjects” in experiments. Due to the limited mobility of individuals with CCN, arranging meetings with them was a complex task often requiring a significant amount of travel by the experimenter. Furthermore, during the interviews, we had to take into account the great variety of impairments and communication aids. A particular challenge was to find a way to present information about the envisioned prosodic Voice Output Communication Aid most effectively to the participants.

As a first step towards an prosodic VOCA, we investigated how individuals with CNN communicated emotions in their daily life and whether a system that produces prosodic speech could be of benefit to them. To shed light on this question, we recruited five participants (see Table 1) that also helped us testing a first version of the prosodic VOCA.

¹ www.assistiveware.com/innovation/expressivepower (accessed 11/09/17)

² www.spectronics.com.au/product/tango-2 (accessed 11/09/17)

Table 1. Overview of the participants

	Sex	Age	Disability	Communication Method
P1	f	8	CP	Tobii C15
P2	f	10	CP	Accent1000
P3	m	15	CP	Tobii I12
P4	f	45	CP	EcpTalker
P5	f	56	ASD	Facilitated Communication

While the speaking ability of the five participants is limited to single sounds, they do not have disorders in language understanding. Four of our participants (P1 - P4) are suffering from cerebral palsy (CP) and use VOCAs to communicate, especially when talking to foreigners. Three participants (P1 - P3) use symbol- and letter-based software on their devices, except P4 who has a reading and writing disorder and relies on symbol-based software only. In addition, she employs a grid to operate her VOCA, a keyboard finger guide for better fine-motor coordination. P1 and P3 control their VOCA via eye tracking technology. The fifth participant (P5) has an autistic spectrum disorder (ASD), and instead of using a VOCA, she uses methods of Facilitated Communication (Table 1). This is a type of communication where a disabled person is supported by a facilitator who leads her or his hand across a communication board, for example. The muscles of people with CCN are often weak and therefore they are only capable of initiating the input and require help to complete it.

Due to the limited mobility of our users, the conduction of a focus group study where several people gather ideas at the same place was no option. Instead, we offered our participants a meeting in an environment that was most convenient to them. The interview with P2 and P4 was conducted at home. During the conversation with P2, her sister and her mother were present as well. P1 and her mother were met in a special education center. The interview with P3 and his speech and language therapist was conducted in conjunction with a logopedics session. P5 was interviewed at the university. She was accompanied by her personal assistant.

The participants were asked whether there are situations in which they would like to be able to communicate emotions. They mentioned situations, such as watching a movie (to express fear) or having a meal or a drink. Furthermore, they would find it useful to express emotions, such as sadness or anger, when somebody does not understand them. Also they would like to be able to communicate emotions when talking about school (dislike and like of peers). Based on the input provided by the participants, we generated a list of prestored utterances for the VOCA to be developed that could be easily accessed by the participants.

During our communication with the participants, we found out that the participants relied on conventional VOCAs to communicate the content of speech and employed additional modalities, such facial expressions, gestures, and sound, to convey the emotions associated with the content of speech. Despite other

means of communicating emotions, our participants found it desirable to be able to express emotions via speech as well and welcomed our idea to develop a VOCA with prosodic speech.

4 The R&D Work

Our VOCA named EmotionTalker was designed as a standalone application for PC and tablet. Its front-end (see Fig. 1) consists of an ordinary keyboard in QWERTZ- or ABC-layout. In order to annotate utterances with the emotion to be conveyed, we placed three emoticons in the upper right corner. These emoticons enable individuals with CCN to annotate the typed utterance as happy, sad, or angry before synthesising. The grey button is clicked to open up this selection and double clicked to switch back to the neutral style. The users can see their utterances in the description field left to the emoticons.

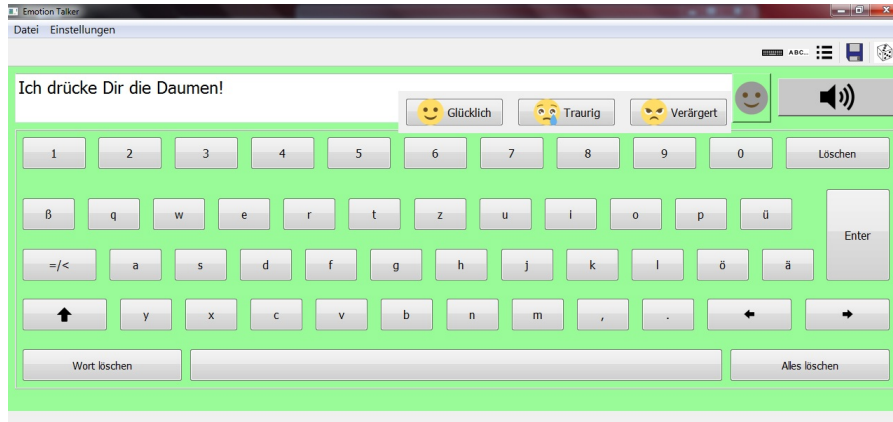


Fig. 1. The front-end of the EmotionTalker showing 'I cross the fingers for you'

The VOCA also allows individuals with CCN to switch between different menus and layouts. The upper five buttons in the right corner offer users to switch from QWERTZ- to ABC-layout. The 'category'-menu gives users fast access to prestored phrases in categories, such as school, work, or leisure. To adapt the VOCA to their needs, users have the possibility to save phrases themselves. In addition, we decided to implement a dice since individuals with CCN mostly have movement disorders and cannot roll a dice (for example, when playing ludo).

To synthesise prosodic speech, we used Cereproc's Ltd. CereVoice synthesiser (Aylett & Pidcock 2007). Text can be annotated with SSML³. A sample sentence, such as 'I cross the fingers for you', looks in a happy tone in SSML as

³ www.w3.org/TR/speech-synthesis11 (accessed 11/28/2017)

follows:

```
<?xml version='1.0'?>
<parent>
<prosody pitch="high" rate="fast" volume="+60">
"I cross the fingers for you"
</prosody>
</parent>
```

The attributes 'pitch', 'rate' etc. of the tag 'prosody' enable us to modulate different characteristics of the tone. We produced several variants of utterances by modifying these attributes and asked our participants to assess how well a variant portrayed a happy, sad and angry tone. The best variants were implemented.

5 Our Study

A first version of EmotionTaker has been tested with the participants mentioned above. Depending on their skills, participants had the chance to control EmotionTaker themselves or with the help of the experimenter.



Fig. 2. A participant playing ludo when she uses the EmotionTaker

The graphical user interface of EmotionTaker was easy to use. Text input via the touch keyboard was intuitive and did not require further explanation. P5

appreciated the distance between the single keys. Due to the limited motor skills of individuals with CCN, keys in the neighborhood of the target are often hit by mistake. One recommendation by her was to display a list of word proposals after pressing a key to speed up communication. The option to store own predefined sentences was well received by our participants. P3 suggested to maintain a list of sentences along with the emotional state to be expressed.

The participants had no problems to map the symbols onto the intended emotional state. They were not in favour of using a higher number of symbols since the selection in daily life should not take too much time.

We also conducted perception tests with the speech output produced by EmotionTalker. In order to make sure that vocal emotions were conveyed in a natural manner, we abstained from the communication of extreme emotions. It turned out, however, that participants were not always able to recognise the subtle emotional states EmotionTalker was supposed to convey. A particular challenge was to manipulate the EmotionTalker voice in such a way that it conveyed the intended emotion with great expressivity without resulting into a voice that appeared to belong to a different person.

Overall, our participants agreed that a prosodic VOCA like EmotionTalker would be of great benefit to them. For example, P2 said: *"I enjoy to tell my mom that I love her - in a happy sound."* And P3 told us: *"I love it to talk with friends about football in an excited and louder manner,"* - he was referring to the angry voice that was presented to him.

With P2, we also tested EmotionTalker during two daily life situations: having a meal and playing ludo (Fig. 2). Even though she had found EmotionTalker easy to use, she rarely used it in these two situations. For example, P2, who was able to eat by herself, did not use EmotionTalker during a meal with her family because using the VOCA while eating was hard. During the game of ludo, she liked to use the dice integrated into EmotionTalker. However, emotions were communicated mostly through facial expressions, gestures and sound since typing took too long. These observations show that the design of a VOCA that may be used during daily activities still remains a challenge.

6 Exploring the Impact

Our study indicates that a prosodic featured VOCA bears great potential to improve the communicative competence of individuals with CCN. Such a VOCA would help satisfy, at least, some pragmatic conversational goals of such users (Todman & Alm 2003). For example, Todman and Alm emphasise the need to incorporate pragmatic features in VOCAs to enable smoother interactions of individuals with CNN. A VOCA like EmotionTalker would provide individuals with CCN with capabilities to deal with unexpected situations, such as confusion or irritation, in a more efficient manner by choosing appropriate emotional backchannels.

Another factor, which is mentioned by Wickenden (2011), is that individuals with CCN often get no attention when they want to say something. Emo-

tionTalker would also mitigate this handicap since an emotional voice helps capture a listener's attention.

7 Conclusion

In this paper, we presented first results for EmotionTalker, a VOCA that provides people with CCN with an additional channel to communicate their emotions. Due to their limited mobility, the recruitment of participants for our study was a challenge. We therefore met our participants at locations that were most convenient to them including their private homes. This way we also got a realistic impression of the participants' physical condition and the environment in which EmotionTalker could be employed. Overall, the participants were very positive towards EmotionTalker. They thought that the ability to communicate emotions outweighs the additional effort required to select the appropriate icons. Nevertheless, more work is required to enable an easy selection of the emotional states to be conveyed in naturalistic environments.

In the future, we will investigate how to speed up the input of emotional states in order to enable users to produce prosodic speech in daily environments. One idea would be to exploit other modalities, such as facial expressions, to determine the user's emotional state based on our previous work on automated emotion recognition (Wagner et al. 2015) and enhance the user's emotional expression by prosodic speech. Furthermore, it would be desirable to offer speech output to the participants that conveys not only emotions in a convincing manner, but also matches their personality.

Acknowledgement

We would like to thank Franziska Kerstiens for her help with the preparation and the conduction of the studies. The work presented here is partially supported by PROMI - Promotion inklusive and the employment centre.

References

- Aylett, M.P., Pidcock, C.J. The cerevoice characterful speech synthesiser sdk. AISB, 174178 (2007)
- Breen, A. Creating Expressive TTS Voices for Conversation Agent Application. In: A. Ronzhin et al. (eds), SPECOM 2014, LNAI 8773, 114. Switzerland: Springer (2014)
- Higginbotham, J.: Humanizing Vox Artificialis: The Role of Speech Synthesis in Augmentative and Alternative Communication. In: J.W. Mullennix, S.E. Stern (eds.), Computer Synthesized Speech Technologies Tools for Aiding Impairment, 50-70. Hershey, PA: IGI Global (2010)
- Hoffmann, L., Wülfing, J.-O. Usability of Electronic Communication Aids in the Light of Daily Use. Proceedings of the 14th Biennial Conference of the International Society for Augmentative and Alternative Communication, 259. Spain: Barcelona (2010)

- Mullennix, J.W., Stern, S.E. Attitudes toward Computer Synthesized Speech. In: J.W. Mullennix, S.E. Stern (eds.), *Computer Synthesized Speech Technologies Tools for Aiding Impairment*, 205-218. Hershey, PA: IGI Global (2010)
- Newell, A.F., Gregor, P., Morgan, M. et al. User-Sensitive Inclusive Design. *Univ Access Inf Soc* (2011) 10: 235-243. <https://doi.org/10.1007/s10209-010-0203-y>
- Portnuff, C. Aac: A users perspective, Webcast available as part of the AAC-RERC Webcast Series. <http://aac-rerc.psu.edu/index.php/webcasts/show/id/3> (accessed 01/10/2018) (2006)
- Pullin, G., Hennig, S. 17 Ways to Say Yes: Toward Nuanced Tone of Voice in AAC and Speech Technology. *Augmentative and Alternative Communication*, 31:2, 170-180 (2015)
- Todman, J., Alm, N. Modelling conversational pragmatics in communications aids. *Journal of Pragmatics*, 35, 523-538 (2003)
- Wickenden, M. Whose voice is that? : Issues of identity, voice and representation arising in an ethnographic study of the lives of disabled teenagers who use Augmentative and Alternative Communication (AAC). *Disability Studies Quarterly*, 31:4 (2011)
- Wagner, J., Lingenfelter, F., André, E. Building a robust system for multimodal emotion recognition. In Konar, A., Aruna Chakraborty, A.: *Emotion recognition: A pattern analysis approach*, John Wiley & Sons, Inc., 379-410 (2015)