Don’t Say Yes, Say Yes: Interacting with Synthetic Speech Using Tonetable

Don't Say Yes, Say Yes: Interacting with Synthetic Speech Using Tonetable

Matthew P. Aylett
University of Edinburgh and
CereProc Ltd.
Edinburgh, UK.
matthewa@inf.ed.ac.uk

Blaise Potard
CereProc Ltd.
Edinburgh, UK.
blaise@cereproc.com

Graham Pullin
University of Dundee
Dundee, UK.
g.pullin@dundee.ac.uk

Shannon Hennig
Inclusive Communication LTD
Wellington, NZ.
shannon@inclusive-communication.co.nz

David A. Braude
CereProc Ltd.
Edinburgh, UK
dave@cereproc.com

Marilia Antunes Ferreira
University of Dundee
Dundee, UK.
mariliaferreiraa@gmail.com

CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2890245

Abstract
This demo is not about what you say but how you say it. Using a tangible system, Tonetable, we explore the shades of meaning carried by the same word said in many different ways. The same word or phrase is synthesised using the Intel Edison with different expressive techniques. Tonetable allows participants to play these different tokens and select the manner they should be synthesised for different contexts. Adopting the visual language of mid-century modernism, the system provokes participants to think deeply about how they might want to say *yes, oh really*, or *I see*. Designed with the very serious objective of supporting expressive personalisation of AAC devices, but with the ability to produce a playful and amusing experience, Tonetable will change the way you think about speech synthesis and what *yes* really means.

Author Keywords
Speech Synthesis; Interactive Media; AAC

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Background and Rationale
At CHI 2014 Aylett et al [1] argued that HCI and speech technology faced significant problems in terms of collab-

oration and understanding. The presentation was given a best talk award, and argued for a new era of collaboration setting out a series of ways the two communities could work together. Tonetable is about realising this objective. The speech technology community needs the expertise of interaction designers and engineers to realise a new generation of eyes-free, hands-busy interfaces, as well as offering better, more flexible speech solutions for for people with severe speech impairments. These two objectives are not independent. The more mainstream and flexible speech interfaces become, the more scope there is for innovative and ground breaking functionality for assistive technology applications. *Don't say yes, say yes*, is the result of collaboration between speech technologists, interaction designers and speech therapists. It is fun, provocative, with serious intent, and offers a real opportunity to support new forms of expressive communication using speech technology.

Its Not What You Say It's How You Say It.
Technology has become part of our social fabric and as such it needs to be able to engender playfulness, and enrich our sense of experience. Furthermore applications which perform a key role in mediating technology for social good require a means of interacting with users in complex social and cultural situations. Speech technology offers a means to extend technology from the mundane to reflect the ambiguity, beauty and complexity of life. HCI has always taken up the challenge of not just looking at what is now, but trying to envisage what is next. Speech technology is key in ambiguous and ludic systems because of the capacity of natural language to be both playful and ambiguous.

Pullin [8] demonstrated the importance of using tangible design to provoke and develop participants understanding and opinions of communication using speech technology. The Six Speaking Chairs, created with Andrew Cook, were a surprising and engaging embodiment of different ways of thinking about tone of voice. Six approaches to generating tone of voice were taken from different academic and creative disciplines. Pullin argues that design can play an important role in provoking new conversations, not just directly solving problems that have already been identified [7].

The exploration of the subtle nuances and differences in speech that can communicate crucial information on the emotional state and intentions of the speaker was continued with the speculative design concept Speech Hedge, created with Ryan McLeod, in which complex tones of voice were represented by abstract plants with different coloured leaves in different combinations [9]. With Tonetable, the objective is more concrete. We have designed and built a tool that allows people with speech impairments to contribute to the design process in the hope that the next generation of AAC technology allows users to craft expressive messages with an appropriate tone of voice for a given message and social context.

To close the gap in performance between biological and synthetic speech researchers are investigating the intelligibility of these artificial voices [3]. Progress is also being made in creating personalised voices that better match gender, age, dialect, and other vocal markers of identity [5, 10][1]. Some work looks to increase the very reduced rate of communication typically observed with most AAC systems. However, even if these issues (i.e., intelligibility, personalisation, and communication rate) are resolved, the question of vocal expressiveness, which increasingly is being understood to play a fundamental role in successful social interaction [4], will remain.

[1] See also CereVoiceMe www.cereproc.com/en/products/cerevoiceme

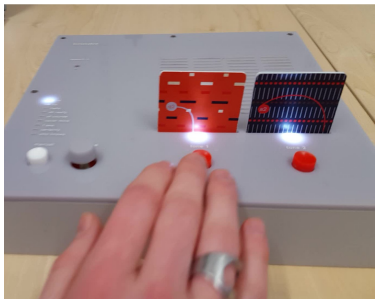

Figure 1: The current Tonetable prototype.

Simply put, we suspect that until those AAC users who rely on speech synthesis can effectively express a given linguistic message in multiple ways, the performance gap will continue to exist.

To a large extent the functionality of speech communication systems for people with significant speech impairments has lagged behind the state of the art. The constraint is not the speech synthesis systems which are a component of an AAC solution. In these modern system a word can be synthesised automatically and with the help of XML mark-up, can be produced in many different ways [2]. Rather it is because of the real difficulty in building interfaces to access and use personalised expressive speech.

Tonetable is envisaged as a participatory research tool. Which allows individuals to craft and select speech tokens for different scenarios such as my enthusiastic yes, my diffident yes, my seductive yes, and so on.

The physical design of Tonetable has been prototyped. It involves a deck of twenty-two cards, each one representing a different tone of voice. Participants can select a card and insert it into either of two card readers, allowing comparisons between two contrasting tones of voice in a particular conversational context (see Figure 1). So as not to influence perceptions, the cards are unnamed except for an alphanumeric identifier. Abstract patterns aid recognition and are intended to support a growing conversation between researcher and participant about each tone. Each card also includes a blank area that can be written on, should participants wish to name or label a specific tone. The Tonetable also comes with a notebook pre-formatted with tables in which experimental results and observations could be recorded. Researchers' notes would be hand-written, but could later be uploaded onto a communal research portal, where results could be shared and discussions hosted.

The participation we are aiming for is therefore twofold: between researchers and people who use augmentative communication, and between different research centers. The latter is important if the cultural nuances of tone of voice are to be recognised – and embraced. As an AAC user, Colin Portnuff lent authority to calls for participation of both kinds when he said, *"Spend time with us. Learn from us, and teach us. Share what you learn freely and openly with your colleagues."* [6] Making the investigative processes accessible to the entire community would catalyse new and deeper lines of inquiry.

About the demo

The *Don't say yes, say yes* demo has two phases. Phase one is based a participatory design process which provokes, familiarises and challenges the participant to explore the different potentially subtle and ambiguous meaning of four renditions of a set of common phrases: yes, no... This process has been designed to help AAC users develop there personalised set of nuanced responses to a variety of scenarios. For example:

Scenario 1: You are a young single person who is asked: "Would you like to go out to an Italian restaurant?" by a work colleague.

Scenario 2: You are eating alone in a restaurant when the waiter asks: "Would you like the bill?"

The participant and demo leader explore the meanings of the four renditions of a constrained set of responses - Yes, No, Hello, Of Course, Certainly etc - to these questions. In addition the dimensions of the responses are mapped out to give an insight in to the user requirements of simple synthesised answers to such questions.

The process helps the participant become familiar with the different renditions, the technology that makes such variation in responses possible, and allows a creative discussion of when, where and how such variation adds to the ludic

and experiential nature of interacting with speech. Not just a game, the results of these phases will be annotated and help define and design speech variation functionality for AAC users, as well as for personal digital assistants and embodied agents.

In phase two we play 20 questions, with the participant writing a famous name of a celebrity of fictional character onto a post it note. This is placed on the demo leader's head and the demo leader asks up to twenty yes/no questions to the participant in order to guess the name. The participant can use any of the vocal renditions available to the Tonetable but will not speak themselves. The objective is to show the playful as well as the functional use of variation in synthesised vocal renditions.

Objectives

To summarise, the objectives of the demo are as follows:

1. To demonstrate the use of embedded systems such as Edison for speech technology provocations and the modern scope of speech synthesis.
2. To demonstrate the potential range of interactive speech systems outside the conventional systems on smart phones or on automatic voice response telephone systems.
3. To understand how important speech variation can be in an interactive setting, and how important they may be for AAC users.
4. To seed new conversations about future developments in speech-enabled interactions.

References

[1] Matthew P Aylett, Per Ola Kristensson, Steve Whittaker, and Yolanda Vazquez-Alvarez. 2014. None of a CHInd: relationship counselling for HCI and speech technology. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 749–760.

[2] Matthew P. Aylett and Christopher J. Pidcock. 2007. The CereVoice Characterful Speech Synthesiser SDK. In *AISB*. 174–8.

[3] Kathryn D. R. Drager, Joe Reichle, and Carrie Pinkoski. 2010. Synthesized Speech Output and Children: A Scoping Review. *American Journal of Speech-Language Pathology* 19, 3 (2010), 259–273.

[4] D. Jeffery Higginbotham, Kyung-Eun Kim, and Christine Scally. 2007. The effect of the communication output method on augmented interaction. *Augmentative and Alternative Communication* 23, 2 (2007), 140–153.

[5] Camil Jreige, Rupal Patel, and H. Timothy Bunnell. 2009. VocaliD: Personalizing Text-to-speech Synthesis for Individuals with Severe Speech Impairment. In *SIGACCESS (Assets '09)*. ACM, 259–260.

[6] Colin Portnuff. 2006. Augmentative and alternative communication: a user's perspective. http://aac-rerc.psu.edu/index-8121.php.html, *lecture delivered at the Oregon Health and Science University* (18 aug 2006).

[7] Graham Pullin. 2009. *Design meets disability*. MIT press.

[8] Graham Pullin and Andrew Cook. 2010. Six speaking chairs (not directly) for people who cannot speak. *interactions* 17, 5 (2010), 38–42.

[9] Graham Pullin and Shannon Hennig. 2015. 17 Ways to Say Yes: Toward Nuanced Tone of Voice in AAC and Speech Technology. *Augmentative and Alternative Communication* 31, 2 (2015), 170–180.

[10] Junichi Yamagishi, Christophe Veaux, Simon King, and Steve Renals. 2012. Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. *Acoustical Science and Technology* 33, 1 (2012), 1–5.