Embodying Tone of Voice, Interacting with Tone of Voice

Graham Pullin
DJCAD
University of Dundee
Dundee DD1 4HT, UK
g.pullin@dundee.ac.uk

The author would be fascinated to participate in this workshop. The list of organisers alone promises an interdisciplinary mix that the author is inspired to be part of, yet confident of making a unique contribution to.

Abstract
Exciting developments in speech technology are opening up the possibility of new interactions with the prosody of synthesised speech. Yet our imaginations can be constrained by how elusive a quality tone of voice is. Exploratory design research has embodied these qualities and these interactions, opening up the discussion – most importantly to people using speech technology in their everyday lives. *Six Speaking Chairs* and *Speech Hedge* have illustrated fresh perspectives and provoked new insights, challenging assumptions that more expressive future speech interfaces should be 'natural interactions'. This paper proposes augmentative communication as a crucible (not just a beneficiary) of radical mainstream speech technology.

Author Keywords
Speech synthesis; interaction design; tone of voice; intonation; prosody; augmentative and alternative communication

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User interfaces: Voice I/O.

Introduction
Exciting developments in speech technology are reconciling recorded corpora with flexibility prosody [2].

With this comes the promise of expressive speech synthesis - the term 'emotional speech' often being used interchangeably. Researchers challenging this equivalence [3] are rare. Manufacturers of communication devices, a profound application of speech synthesis, are anticipating emotional speech synthesis involving 'natural interactions' through biosensors [11]. People who use speech generating devices have expressed a desire for more nuanced tone of voice [8]. However, opening up discussion is challenging, tone of voice being an elusive quality that even phoneticians struggle with [5].

Six Speaking Chairs

A first project uses critical design: design not intended to solve a problem directly, rather to ask questions and provoke discussion. *Six Speaking Chairs* is a collection, each of which embodies a different way of conceiving of tone of voice in the first place - and thereby controlling it. These mental models were pre-existing, found in the work of researchers and creative practitioners in other fields, from linguistics to the theatre (and to echo this, the chairs themselves are reclaimed furniture and the controls are repurposed computer keys, radio dials and doorbells - see Figure 1) [12].

17 ways to say yes

The chairs proved a valuable research output in themselves, but also generated further insights. Following a presentation of the *Six Speaking Chairs*, participants were given a sheet of paper and ten minutes in which to list (or otherwise mark) the tones of voice that they could imagine choosing, were they to be limited to 17 for the rest of their lives. For the non-speech-impaired participants, this would be a restriction; for participants using speech generating devices, this represents a widening of choice.

257 ways to say yes

The responses from 40 individuals - including people who use communication devices, speech and language therapists, speech technologists - were collated. Combining equivalent terms resulted in 257 distinct descriptors. It was immediately apparent that whilst some of these - for example 'angrily', 'joyously', 'sleepily' - might be said to describe the emotional state of the speaker, others - for example 'patronizingly', 'reflectively', 'whispering' - did not.

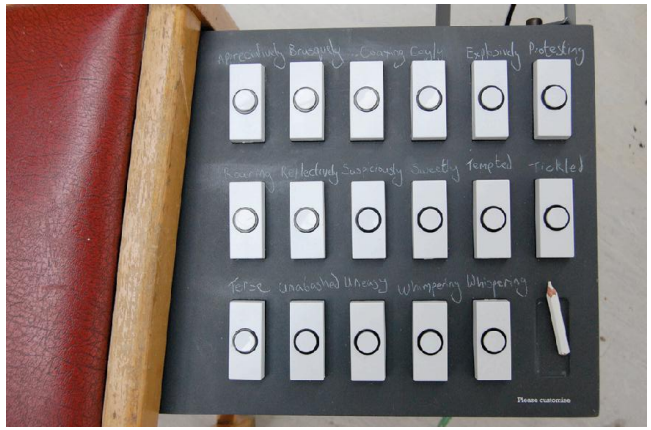

Figure 1. Chair No.6 with 17 doorbells and pencil (inset, left).

Diverse perspectives on tone of voice

Sorting the 257 descriptors, a system of four complementary perspectives on tone of voice was conceived: emotional state; conversational intent; social context; vocal qualities.

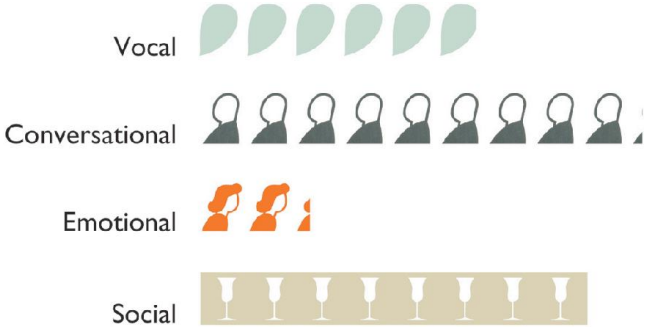

Figure 2. Mapping 257 descriptors against four perspectives - each complete icon represents ten of the 257.

In the context of this research, this is not intended as a definitive set as much as a challenge to the dominance of an emotional perspective. We enlisted an independent researcher to map the 257 tones against these four perspectives: her mapping was even more challenging than our own, with a minority of tones described in terms of emotions (see Figure 2) [10].

Speech Hedge

Having illuminated the diverse perspectives that people intuitively bring to thinking about tone of voice, the second project is an exploratory user interface. *Speech Hedge* is a visualisation, using an Apple *iPhone*™ to control the tone of voice of a speech generating device. Each tone of voice is represented by a 'plant', composed of different coloured 'leaves' (see Figure 3). Each leaf is an elemental everyday tone, described from one of the four perspectives. The plants can be collected into 'hedges' offering limited palettes of tones appropriate for different social contexts. The motif of plants [6] is not important, but the modularity is.

Figure 3. *Speech Hedge* concept

A million ways to say yes

Following a presentation of Speech Hedge, participants were given a form and ten minutes to speculate as to how they might synthesise a few complex tones of voice by combining elemental tones (in practice this would be a matter of trail and error). 'Coaxing' and 'sarcastically' were suggested, along with any additional tones of their own choosing (see Figure 4).

Responses from 39 participants - again including people who use communication devices, speech and language therapists and researchers - were collated and visualised. They illuminated a widespread ability to engage with even such an abstract model, and the ability of this model to accommodate personal differences in describing tone of voice.

Different ways to say yes sarcastically

Other comparisons revealed a deeper issue though: one participant synthesised 'sarcastically' by combining 'bored' and 'emphasising', whilst another assembled 'sarcastically' from 'loudly', 'energetically' and 'emphasising'. The first reads as a dead-pan delivery of sarcasm; the second an ironically exaggerated delivery. Two different sounds are being described.

It would be easy to see this finding as being a liability, the field of speech technology, in its quest for so-called 'natural' interaction, being traditionally averse to ambiguity and inaccuracy. But who is to say what any other tone of voice should sound like?

Toward an open library of tones of voice

What if this cultural subjectivity were embraced? Leading augmentative and alternative communication (AAC) researchers have identified increased demands

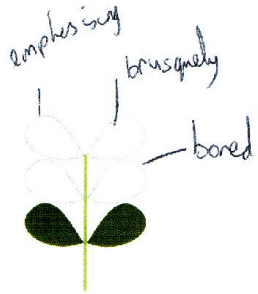

Figure 4. 'Sarcastically' response to *a million ways to say yes*

for more culturally responsive AAC [7]. Increasingly there are precedents for such creative collectives [1]. *Speech Hedge* includes this interaction (as a concept). Research is planned to actually build and deploy such a platform.

Discussion

A wider implication of this research is the substantive contribution that design could make to interdisciplinary research. Not only by helping to conceive radical new interfaces (which might be expected of interaction designers) but at a more fundamental level: by embodying elusive qualities of speech to provoke discussions about what our priorities should be.

AAC could be a crucible of appropriate but radical approaches, not just the beneficiary of mainstream technology [9]. AAC always involves interpersonal interaction - which could have wider implications for HCI as it aspires to become more 'natural'.

Acknowledgements

Six Speaking Chairs was in partnership with Dr Andrew Cook, whose PhD [4] was funded by the EPSRC through the School of Computing, University of Dundee. *Speech Hedge* was in collaboration with Ryan McLeod, as part of his MDes at DJCAD, University of Dundee. Thanks to all participants in the *17 ways to say yes* [11] and *a million ways to say yes* [10] exercises.

References

[1] Adobe *Kuler*. https://kuler.adobe.com

[2] Astrinaki, M., d'Alessandro, N., Picart, B., Drugman, T., and Dutoit, T. 2012. Reactive and continuous control of HMM-based speech synthesis. *Proc. IEEE SLT 2012*.

[3] Campbell, N. 2005. Getting to the heart of the matter: speech as the expression of affect, rather than just text or language. *Language Resources & Evaluation* 39(1):109–118.

[4] Cook, A. 2013. Studying interaction design by designing interactions with tone of voice. Unpublished PhD thesis, University of Dundee.

[5] Fox, A. 2000. *Prosodic features and prosodic structure: the phonology of suprasegmentals*. Oxford University Press.

[6] Orla Kiely. http://www.orlakiely.com

[7] Light, J., and McNaughton, D. 2012. The changing face of augmentative and alternative communication: past, present and future challenges. *AAC* 28(4):197–204.

[8] Portnuff, C. 2006. Augmentative and alternative communication: a user's perspective. http://aac-rerc.psu.edu/index-8121.php.html

[9] Pullin, G. 2009. *Design meets disability*. Cambridge, MA: The MIT Press.

[10] Pullin, G. 2013. 17 ways to say yes, exploring tone of voice in augmentative communication and designing new interactions with speech synthesis. Unpublished PhD thesis, University of Dundee.

[11] Pullin, G., and Cook, A. 2008. Six speaking chairs to provoke discussion about expressive AAC. *Proc. ISAAC 2008.*

[12] Pullin, G., and Cook. A. 2010. Six Speaking Chairs (not directly) for people who cannot speak. *Interactions* 17(5):38–42. New York: ACM.

[13] Pullin, G., and Cook, A. 2013. The value of visualizing tone of voice. *Logopedics Phoniatrics Vocology* 38(3):105–1.