Non-visual Human-Computer Interaction

By Aaron Bowen, ECE '17

Introduction

There is a rich, under-utilized set of non-visual human-computer input and feedback interactions that complement the ubiquitous visual smartphone interface. This note investigates a series of fundamental and novel audio and tactile tools for smartphone Human-Computer Interaction (HCI), seeking to demonstrate the rich variety of non-visual feedback available to augment the modern smartphone interface.

Audio

Text-to-Speech

Text-to-speech, as it sounds, is written text read aloud by a computer, as distinct from a recording a human reading or speaking. The core functionality of text-to-speech is speech synthesis, the artificial production of human speech. Synthesizing human-like speech is an immensely difficult problem involving "linguistic theory [...], perceptual psychology, mathematical modeling of speech production, structured programming, and computer hardware design" (Klatt, 1987). Many modern operating systems (including Android, iOS, macOS, and Windows) are shipped with a text-to-speech engine, which can be leveraged through by programs through an API, or by the user themselves to have on-screen text read aloud.

Speech-to-text and Emotion Recognition

Speech-to-text, the complement of text-to-speech, (also called automatic speech recognition) is human speech synthesized into text. Speech recognition lives in a similar problem-space as text-to-speech where linguistic theory, etc. are still relevant along with new areas such as machine learning, natural language processing, and paralinguistics (Cieri et al., 2004) (Ververidis et al., 2006).

A primary challenge within speech recognition is emotion recognition. Emotion recognition is not a new concept, but has undergone significant advancements in recent years with the application of hidden Markov models and neural networks (Schuller et al., 2003). A group out of the Aristotle University of Thessaloniki investigated detecting emotional "features" in speech, most interesting of which were pitch, formants (resonant frequencies of the vocal tract, which vary with mood), short-term energy, and the Mel-frequency cepstral coefficients (a more refined look at vocal frequencies that take into account the human auditory frequency response) (Ververidis et al., 2006). The study concluded with several models of short-term emotion in speech, separated by gender that could be used for further investigation.

Tactile

Tactons

A research group from the University of Glasgow led by Stephen Brewster investigated the use of tactile icons, or Tactons. Tactons are analogous to icons (a visual symbol, such as a printer icon) and earcons (an audio symbol, such as the falling tone that may sound when a file is deleted), in that they all concisely represent complex, abstract concepts and together form a "simple, efficient language to represent concepts at the user interface" (Brewster et al., 2004). Brewster posits that presenting information through speech, written text, and braille text is "slow because of its serial nature; to assimilate information the user must hear a spoken message from beginning to end and many words may have to be comprehended before the message can be understood." Icons and earcons are effective at conveying complex information extremely quickly, but until Brewster's group, Braille had no iconic counterpart (Brewster et al., 2004).

A tacton uses a subset of the parameters for tactile perception to encode information (a subset also found in auditory perception): frequency, amplitude, waveform, duration, and rhythm. A tacton may be compound (morphing from an initial to a final state), hierarchical (layered, rhythmic frequencies), or transformational (where tacton attributes map directly to the subject's attributes) (Brewster et al., 2004).

With huge variability in the size of possible tacton displays, the backs or edges of smartphone and tablets could be converted into tacton displays, bringing this information channel to the mobile industry.

Mobile Force Feedback

Force feedback is the simulation of physical forces such as weight, acceleration, and pressure. Current research applications focus on an enhanced sense of realism in augmented and virtual reality, the aided comprehension of complex information, and the improved efficiency of more conventional visual interfaces (Brave et al., 2001).

These research interests have, until recently, resisted entrance into the mobile world because of the requirement for motors too large to be conveniently portable. In 2013, Pedro Lopes from the Hasso Plattner Institute in Germany published a study detailing a miniaturized force feedback system that "[actuates] the user's [arm] muscles using electrical stimulation" (Lopes et al., 2013). Lopes demonstrated that electrically actuating the user's muscles in response to on-screen activity "creates sufficient force and that its effect is indeed perceived as force feedback". Furthermore, the force feedback was preferred to vibrotactile feedback because of its novelty, increased realism, and relatability to the physical world (Lopes et al., 2013).

Bringing force feedback to the mobile scale is an important advancement that opens the door to more advanced, wearable augmented reality tools.

Conclusion

Human-computer interaction is presently weighted heavily toward visual interfaces, an overdependence that limits the rate of complex information that a person can intake from a smartphone. Capitalizing on the alternative, underused information channels of audio and tactile communication offers a high capacity link one's smartphone. This link gives smartphones new and higher utility in a broader span of applications.

References

Brave, S., Nass, C., & Sirinian, E. (2001, August). Force-Feedback in computer-mediated communication. In HCI (pp. 145-149).

Brewster, S., & Brown, L. M. (2004, January). Tactons: structured tactile messages for non-visual information display. In Proceedings of the fifth conference on Australasian user interface-Volume 28 (pp. 15-23). Australian Computer Society, Inc..

Cieri, C., Miller, D., & Walker, K. (2004, May). The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text. In LREC (Vol. 4, pp. 69-71).

Klatt, D. H. (1987). Review of text-to-speech conversion for English. The Journal of the Acoustical Society of America, 82(3), 737-793.

Lopes, P., & Baudisch, P. (2013, April). Muscle-propelled force feedback: bringing force feedback to mobile devices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2577-2580). ACM.

Schuller, B., Rigoll, G., & Lang, M. (2003, April). Hidden Markov model-based speech emotion recognition. In Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on (Vol. 2, pp. II-1). IEEE.

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. Speech communication, 48(9), 1162-1181.