**Abstract of**
**Sound Oriented CAPTCHA**

*Nancy Chan*
*City University of Hong Kong*

CAPTCHA stands for Completely Automatic Public Turing Test to tell Computers and Humans Apart. This test can be graded by computer but only human can pass. For more details, you can go to http://www.captcha.net.

The original CAPTCHA in CMU mainly based on the ability of human to read from distorted text and the inability of computer to read from distorted text. However, this has a shortcoming that people weak at their eyes may not pass the test. For this reason, sound oriented CAPTCHA is now being designed, and everything is still in the research status.

Up to now, there are 2 versions: BYAN-I and BYAN-II
These 2 versions of sound-oriented CAPTCHA mainly based on speech perception. According to the email from Prof. Moreno and [1], adding another speaker at the background is a comparatively harder noise for ASR. When 2 speeches at the same volume mix together, human can still concentrate on one voice and exclude the other one, however, this is very difficult for machine to distinguish.

In both versions, user is asked to choose a language they used to speak, and then computer will randomly generates 6 digits and creates a sound file including 6-digits spoken in the chosen language with background noise, which is another speaker. User should be able to input what digits he/she heard from the sound file.

The main difference between 2 versions is the background noise.
In BYAN-I:
The background noise is the same 6-digits spoken in another language.

In BYAN-II:
The background noise is some randomly picked vocabularies spoken in another language.

Nevertheless, according to the current technology for ASR, digit recognition has been developed very well [4] and the recognition performance can be improved by training the recognizer in a similar environment [1], [2], [3]. It may be possible to break this test by a trained recognizer. In order to know the answer, cooperation with speech recognition experts for these 2 versions will be needed

**Reference:**

[1] Jean-Claude Junqua & Jean-Paul Haton (1996), *Robustness in Automatic Speech Recognition*
[2] Hans-Gunter Hirsch & David Pearce (2000), *The Aurora Experimental Framework for the performance evaluation of speech recognition systems under noisy conditions*, ASR2000, Paris, France, Sept. p.18-20
[3] P.D.Green, M.P. Cooke & M.D. Crawford (1995), *Auditory Scene Analysis and Hidden Markov Model Recognition of Speech in Noise*, IEEE p.401-404
[4] J. Barker, M.Cooke and P.Green, *Robust ASR Based on Clean Speech Models: An Evaluation of Missing Data Techniques for Connected Digit Recognition in Noise*, Eurospeech 2001, Special Event on Noise Robust Recognition, Denmark, September