

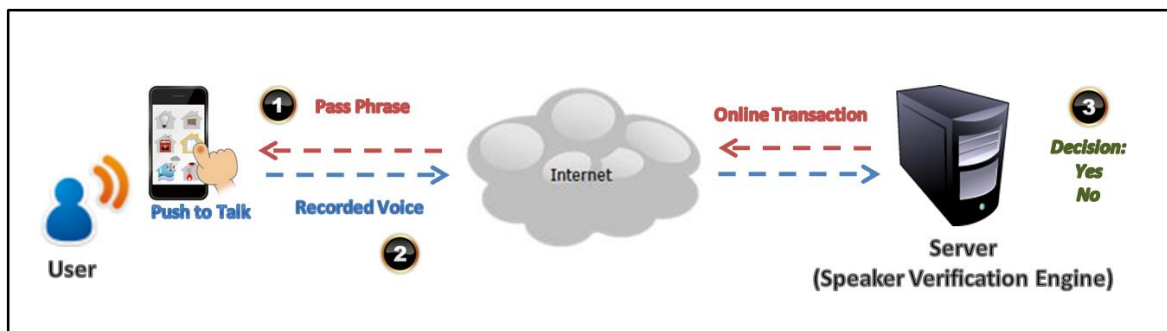
# IMAGINATION TO REALITY

## Voice Biometrics – My Voice Tells Who I Am



### Technology Overview

Over the past few decades, speaker recognition technology has seemed to reach high accuracy in the context of conversational telephony speech as evaluated in the past few NIST (i.e., US National Institute of Standards and Technology) speaker recognition evaluations (SREs). However, the accuracy can easily drop below 90% when a duration constraint has to be imposed (where speech samples are restricted to be 3 seconds or less) for the ergonomic use of voice biometrics in daily context. In addition, a multilingual solution (where the users could use pass phrases in different languages) has also shown an increasing market demand. The focus of the technology is therefore to offer high accuracy voice biometric using short pass phrases of 3 seconds and in different languages.



- The above diagram shows the overall architecture of a voice biometric system used for an online access control scenario.
- For any attempt to access the system, the server will respond to the request with a pass phrase randomly selected from a database.
- The user reads the pass phrase displayed on the screen.
- The recorded voice will be sent to the server for verification.
- The recognition engine performs the verification and responses with an accept or reject decision.

### Technology Features

- Verify the identity of a speaker using just 3 seconds of speech with high accuracy.
- Use pre-defined set of pass phrases, randomly prompted for each access to prevent recording attack.
- Language independency, allows the use of pass phrases in different languages.

### Potential Applications

1. Second-factor authentication for credit card, internet and telephone banking.
2. Online authentication for transactions and services involving large payment amount or high risk (e.g. request for the resetting of passwords).
3. Physical access control. Voice is used as the “soft key” in conjunction with a smart phone functioning as the “hard key” and front-end recording device.

### Benefits

Text-dependent but language-independent speaker recognition: The technology models and compares speakers' characteristics under specific lexical context (i.e. the text content) to gain high recognition accuracy. We train the statistical models in a progressive manner from general to speaker specific and to text-speaker specific model based on previous stages. This technique allows us to train a sufficiently robust model for each short pass-phrase, and leads to a language-independent implementation enabling the use of pass phrases in different languages.

### Contact Us

Industry Development Department  
Institute for Infocomm Research (I²R)  
1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632  
Tel: (65) 6408 2000 Fax: (65) 6776 1378 Email: inddev@i2r.a-star.edu.sg  
www.i2r.a-star.edu.sg