

# Voice Control on Smart TV Using Farsi Speech Recognition

Woo Kyeong Seong, Jang Won Lee, Dae Hyun Kim, Hong Kook Kim,  
Sung Hwan Shin, Ji Hun Park, and Myung Jae Kim

**Abstract** We demonstrate a Farsi voice control system which enables a user to operate a smart TV through a microphone mounted in a remote controller. In particular, users can easily control 451 smart TV commands such as power on/off, channel/volume change, and TV application access. The developed Farsi voice control system is based on Farsi automatic speech recognition (ASR) in a server-client scenario with TCP/IP protocol [1]. At a client side, a smart TV captures user's speech and sends raw speech to an ASR server. Then, at the server side, the transmitted speech is decoded using a Farsi ASR system and the smart TV performs actions corresponding to the recognized word. The Farsi ASR system is developed in a weighted finite state transducer (WFST) using the Kaldi open source speech recognition toolkit [2]. Here, mel-frequency cepstral coefficients with cepstral mean normalization are used as recognition feature, and sub-space Gaussian mixture models (SGMMs) [3] are trained with 120 hours Farsi speech data.

## References

1. Forouzan, B. A.: TCP/IP Protocol Suite. McGraw-Hill, Inc., New York, USA (2002)
2. Povey, D., et al.: The Kaldi speech recognition toolkit. In: Proceedings of ASRU, Honolulu, HI, pp. 1-4 (2011)
3. Qian, Y., Povey, D., Ju, J.: State-level data borrowing for low-resource speech recognition based on subspace GMM. In: Proceedings of INTERSPEECH, Florence, Italy, pp. 553-560 (2011)

---

Woo Kyeong Seong, Jang Won Lee, Dae Hyun Kim, Hong Kook Kim  
School of Information and Communications, Gwangju Institute of Science and Technology (GIST), e-mail: {wkseong, ljw, daehyun, hongkook}@gist.ac.kr

Sung Hwan Shin, Ji Hun Park, Myung Jae Kim  
Visual Display Business, Samsung Electronics Co. Ltd, e-mail: {sean.shin, jh1227.park, koong.kim}@samsung.com