Xiaoyun Wang

Doshisha University
Department of Information and
Computer Science
1-3 Tatara Miyakodani, Kyotanabe-shi,
Kyoto, JAPAN 610-0394

ou.gyouun@gmail.com

1 Research Interests

My research interests lie generally in the area of **spoken dialogue systems** aim at **foreigner languages learning**. The main theme of my research is exploring the effective and customized **acoustic modeling** for the **non-native speech** used in a dialogue-based computer-assisted language learning (CALL) system. These kinds of dialogue systems act as automated interlocutors that prompt learners to elicit speech in the target language and provide informative feedback that is of enormous educational value in terms of improving the learners' language communication skills (Kawai, 1997; Ito, 2008). But somehow, the quality of non-native speech usually differs significantly from native one in terms of phonemes, prosody, lexicon, disfluencies, and so on. Its automatic speech recognition (ASR) can be a great challenge. My research goal is to build a recognition system with adapt models that suit the character of second language (L2) speakers.

1.1 Dialogue-based computer-assisted language learning (CALL) system

I am interested in computational linguistics, and modeling based on the characteristics of both acoustic and linguistic features for speech, such as language acquisition based on human-computer/robot interaction; such as phonetic acoustic features of speech, and lexical or grammatical information of transcription. In addition I design a dialogue-based English CALL system for Japanese, and modeling the customized acoustic models for Japanese-English speakers.

Wang et al. (2014) described a work on developing a human-machine dialogue-based English CALL system. The system aims at eliciting more speech production from Japanese learners in order to improve their speaking skills, by involving them in man-machine dialogues. The system interface provides prompt text in Japanese, in order to help learners construct utterances which are easy to be recognized. Figure 1 illustrates a screenshot of the man-machine interface of the developed CALL system.

Figure 1: *A screenshot of the CALL interface: (1) dialogue scenario selection (shopping, restaurant and hotel); (2) prompt by system; (3) hint stimulus; (4) recognition result; (5) corrected feedback*

1.2 Customized acoustic modeling with reduced phoneme set considering integrated acoustic and linguistic features of second language speech

Most of the ASR technologies have been developed to handle the subject of pronunciation variations in terms of acoustic modeling or extended lexicon and grammatical relations in terms of language modeling for non-native speech ASR. However, there are almost no methods that handle the difference between acoustic and linguistic features of non-native and native speech in a unified way, even if both features share a close relation and should be simultaneously taken into consideration.

Wang et al. (2016) proposed a novel phoneme set design method, based on the research results obtained with the previously proposed reduced phoneme set from the perspective of handling the acoustic and linguistic features of non-native speech in a unified way. The previously proposed reduced phoneme set was created with a phonetic decision tree (PDT) based top-down sequential splitting method (Wang et al. 2015a) that utilizes the phonological knowledge between mother and target languages and their phonetic features, delivering a better recognition performance for non-native speech. The recent approach considers acoustic and linguistic discrim-

inating performance in a unified way and optimizes the weighted total of both discriminating performances.

1.3 Examine the relation between proficiency of second language speakers and a reduced phoneme set customized for them

The proficiency of L2 speakers varies widely, as does the influence of the mother tongue on their pronunciation. As a result, the effect of the reduced phoneme set is different depending on the speakers' proficiency in L2. Wang et al. (2015c) examined the relation between proficiency of speakers and a reduced phoneme set customized for them.

The experimental results are then used as the basis of a novel speech recognition method using a lexicon in which the pronunciation of each lexical item is represented by multiple reduced phoneme sets (Wang et al. 2015b), and the implementation of a language model most suitable for that lexicon is described. Then multiple-pass decoding using a lexicon represented by multiple reduced phoneme sets is proposed for speech recognition of second language speakers with various prociencies. The relative error reduction obtained with the multiple reduced phoneme sets is 26.8% compared with the canonical one.

2 Future of Spoken Dialog Research

In the future of dialogue research, there are several challenges that are in need of attention. The first is how to collect enough and effective dialogical data for large scale statistical analysis, not only the adult speech and native speech, but also the children speech and non-native speech. The second is how to make better use of paralinguistic phenomena of human conversation characteristics, such as eye gazes, emotions and so on, to improve the performance of spoken dialogue systems.

I also work on collecting a multimodal corpus of human-to-human and human-to-robots conversations in English as second language by Japanese speakers in order to develop natural dialogue turn-taking model. Their voices were recorded together with the gaze-tracking data which is assumed to provide close cues for dialogue turns.

3 Suggestions for Discussion

- What's the core competence for basis research of dialogue systems in universities in comparison to that in industries?
- Standardization: How to collect enough and effective dialogical data for large scale statistical analysis? How to rapid development and standardization?
- Evaluation: How to evaluate the dialogue system (human-machine/robots)? Is there more standard one?
- Perspectiveness: How far can we go in deep learning for the research of dialogue systems?

References

Goh Kawai, and Keikichi Hirose. 1997. *A CALL system using speech recognition to train the pronunciation of Japanese long vowels, the mora nasal and mora obstruents*, EUROSPEECH, Rhodes, Greece.

Akinori Ito, Ryohei Tsutsui, Shozo Makino, and Motoyuki Suzuki, 2008. *Recognition of English utterances with grammatical and lexical mistakes for dialogue-based CALL system*, INTERSPEECH, pp. 2819–2822, Brisbane, Australia.

Xiaoyun Wang, Jinsong Zhang, Masafumi Nishida, and Seiichi Yamamoto, 2014. *Phoneme set design using English speech database by Japanese for dialogue-based English CALL systems*, LREC, pp. 3948–3951, Reykjavik, Iceland.

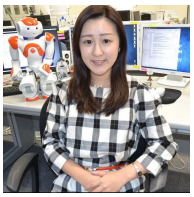
Xiaoyun Wang, Jinsong Zhang, Masafumi Nishida, and Seiichi Yamamoto, 2015. *Phoneme set design for speech recognition of English by Japanese*, IEICE Transactions on Information and Systems, 98(1): 148–156.

Xiaoyun Wang, and Seiichi Yamamoto, 2015. *Second Language Speech Recognition Using Multiple-Pass Decoding with Lexicon Represented by Multiple Reduced Phoneme Sets*, INTERSPEECH 2015, Dresden, Germany.

Xiaoyun Wang, and Seiichi Yamamoto, 2015. *Speech Recognition of English by Japanese using Lexicon Represented by Multiple Reduced Phoneme Sets*, IEICE Transactions on Information and Systems, 98(12): 2271–2279.

Xiaoyun Wang, Tsuneo Kato, and Seiichi Yamamoto, 2016. *Phoneme Set Design Considering Integrated Acoustic and Linguistic Features of Second Language Speech*, INTERSPEECH 2016, San Francisco, USA. (accepted)

Biographical Sketch



Xiaoyun Wang is a PhD candidate at the graduate school of Science and Engineering, Doshisha University, Kyoto, Japan, working under the supervision of Professor Seiichi Yamamoto. She is also a collaborative researcher in National Institute of Information and Communications Technology (NICT), Japan. Her research interests include speech recognition, spoken dialogue system, language acquisition, spoken language processing, and speech processing. She received a B.S. in Information Science from Yamanashieiwa University, Japan in 2012 and an M.S. from the graduate school of Science and Engineering, Doshisha University, Japan in 2014.