Usage Scenarios 1: Listening support for NNS
N2F accent conversion can be used to support NNS in conversation with NS. As well as making it technically easier to comprehend what natives are saying, sharing the same accent can take off embarrassment and pressure to speak in a foreign accent.

We expect the following two styles for this application: the press-to-talk style and the hearing-aid style. Hearing-aid style would be more difficult to implement, since utterance detection and accent conversion has to work accurately in real-time.

N2F accent conversion is also appropriate for foreign language learning just as the training wheels for bicycles. For this purpose, we should consider ways to support taking the training wheels off later, for example by features to gradually reduce the level of accent conversion.

Promoting Intercultural Awareness through Native-to-Foreign Speech Accent Conversion

Takeshi Nishida
Graduate School of Intercultural Studies, Kobe University
1-2-1 Tsurukabuto, Nada-ku
Kobe, Hyogo, JAPAN
tnishida@people.kobe-u.ac.jp

CABS'14, Aug 21-22 2014, Kyoto, Japan
ACM 978-1-4503-2557-8/14/08.
http://dx.doi.org/10.1145/2631488.2634058

Abstract
Large difference in pronunciation between languages often causes technical and mental difficulty in listening comprehension as well as in speech production. In spite of the rationale provided for native speakers to support non-native speakers in international contexts, negative reactions to foreign accents are common. To promote intercultural awareness in communication, we implemented a prototype system to convert native speech to have a foreign accent by combining speech recognition and text-to-speech engines.

Author Keywords
Intercultural awareness, accent conversion.

Introduction
Japanese people tend to be described as having serious trouble in English conversation, and difference in pronunciation has been considered to be one of the major obstacles in acquiring conversation skills. Listening to and imitating native speech has been a longtime custom in language learning; however, this is questioned by researchers, reporting that only a small minority acquires accent comparable to native-speakers, and others just lose motivation to language learning [2].

Usage Scenarios 2: Cultural shock experience for NS

N2F accent conversion can provide NS the opportunity to compare international communication with their own accent and with an accent similar to the person he/she is communicating with. Through this shocking experience of accented speech comprehended better than the original speech, people can feel and learn that "natural English" is not necessarily the only best language, and hopefully improve their attitude toward intercultural communication.

N2F conversion can also be used for listening practice of foreign-accented English. Expecting NS to understand accented speech should be a more realistic goal than to expect NNS to speak in native-level accent.

While on the other hand, more people worldwide come to speak English with various accents, which is referred to as "World English." We believe that this provides a strong rationale for "natives" to show willingness to compromise on accented speech and share more burden in communication. Quite a few native speakers actually show willingness to compromise on foreign accents; however, negative reactions to foreign accents are common and deep-rooted [1, 4].

To fill this cultural gap between native speakers (NS) and non-native speakers (NNS) and promote intercultural awareness in communication, we propose to convert native speech to foreign accented speech. This native-to-foreign (N2F) speech accent conversion can make NNS concentrate better on expressing their thoughts during conversation, since it is often easier for them to comprehend utterances similar to their own accent [2]. Moreover, NNS may feel easier to speak if same accent is shared within the group. It can also help NS get rid of their preconceived notions about "natural speech" and enhance their tolerance to accented speech, through experiences where accented speech is comprehended better than their own speech.

Prototype Implementation

We implemented a prototype to explore the possibilities of N2F accent conversion. It works similarly to voice translation applications; users press a button before start talking, and after the user finishes a phrase, the system plays accent converted speech.

Conversion Methods

Speech recognition (SR) and text-to-speech (TTS) are used to convert native speech to foreign accented speech[1]. Converted speech will be played immediately after the input using an ordinary Windows laptop.

We implemented two conversion modes which combine the two engines in a different way. In the first mode (**Figure 1**), speech accent is converted by having Japanese TTS read out the recognized English text. Resulting speech sounds like a novice English speaking Japanese person pronouncing the phrase word-by-word. We expect that converting to other foreign accents is also possible by using TTS of other languages.

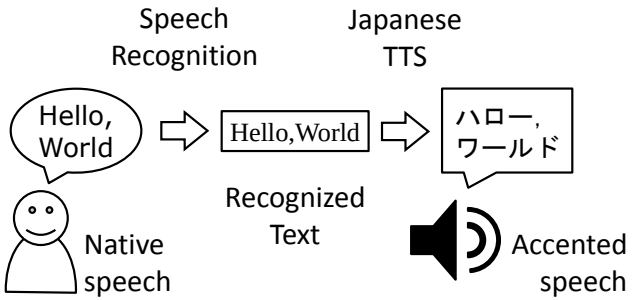

Figure 1. (N2F conversion mode 1) Accent converted by English speech recognition + Japanese TTS.

However, some words will not be pronounced correctly (not like Japanese nor English-native), probably due to the dictionary used to read non-Japanese word with the Japanese TTS.

Conversion mode 2 uses the recognized phonetic symbols instead of text. In this mode, accent is converted by applying string replacement rules to the

[1] We used SR and TTS engine for Microsoft Windows.

Usage Scenarios 3: Contribution balancing in large group meeting

Even for people with sufficient language skills in one-to-one or small group setting, multiparty communication is difficult, where prompt response is required for taking turns among NS. This often results in imbalanced contribution between NS/NNS [6].

N2F converted accent at group meetings can help balance the contributions; because effort to comprehend others' speech will be balanced, and sharing the accent can make NNS feel free from worries regarding their own accents.

Real-time speech detection is not strictly required for this setting because we can expect per-participant microphones with a push-to-talk button. However, accent conversion has to work in nearly real-time to be used in meetings, otherwise it will slow down the meeting too severely.

recognized phonetic symbols and reading out the phonetic symbols using TTS. SR may still have unknown words problems but the conversion works reasonably well for phonetically close guesses.

The second mode allows more precise control on the conversion; for example, "replace R with L" rule can be emitted for people capable of discriminating those two consonants. In addition, the second mode allows the use of English or even TTS of any other languages for accent conversion. For example, speech converted using English TTS sounds like an English native person trying to imitate the pronunciation of a novice English speaking Japanese. We expect that converting to other accents is possible similarly.

One of the major problem of the both modes is that they also convert the voice. We believe that this can be solved by using voice conversion methods.

User Interface

Figure 2 shows the screenshot of the application we implemented to support various type of experiments and real situations. The window consists of conversion parameter setting (upper-left), input selection (lower-left), and recognition history list (right). The prototype supports both real-time input from microphone and input from recorded audio file.

The prototype can handle different input accents of English such as British or American and languages other than English by selecting an appropriate SR engine. Conversion mode 2 is used when "Japanizer" is checked, then TTS engine of any available language can be used for conversion. Japanese TTS has to be selected for conversion without the Japanizer (conversion mode 1). Speed of the output speech can be configured in two different ways: the speech speed slider control and the word-by-word checkbox.

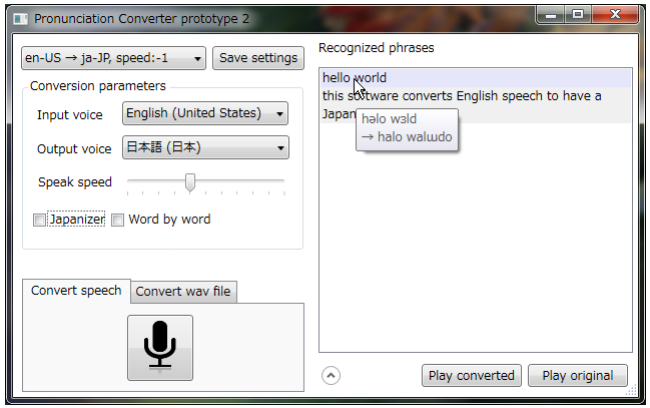

Figure 2. Screenshot of the prototype application.

SR results in text will be displayed on the right side. Users can check the phonetic symbols before and after conversion in the popup. Clicking an item will replay the accent-converted speech. This enables us to easily compare resulting speech by different conversion parameter, and to compare N2F conversion with other supportive methods, for example showing recognized text of what is being said.

Related Work

Previous work has examined multilingual collaboration mediated by a text-chat system equipped with machine translation [7, 8]. Variety of voice translation apps are ready for end-users, where users can speak in their mother tongue to generate translated speech. However, difficulty of multilingual communication remains since even the most accurate systems not infrequently make

Early User Feedback

We collected early feedback from users of various backgrounds, after playing with the prototype: Japanese university students, a native English speaker teaching English conversation at a Japanese university, and Chinese student studying as foreign students in Japan.

All users commented that the accent-converted speech by the first conversion mode simply using Japanese TTS, definitely sounds like a novice English speaking Japanese person. Users also agreed that the speech converted by the second mode was Japanese accented as well, but more difficult to comprehend.

The native English teacher suggested the use of accent conversion software in her English conversation class as an ice-breaker tool.

translation errors. Errors are inherent in N2F accent conversion as well; however, accent conversion errors can be infrequent and less fatal than translation errors. It can produce a phonetically close speech even if the speech recognition fails.

Audio processing technologies has been used to support second language learning mainly by voice conversion, based on the knowledge that imitating speech of their own voice with native accent helps to improve speaking and listening skills [3]. Not as much work has been done on speech accent conversion, and all of them were attempts to generate native-like speech [3].

Simply slowing down the speech or showing recognized text can largely support NNS. In addition, using systems such as Speech Repair [5] to correct speech recognition errors by hand, can resolve the difficulty of using error-prone systems. Yamashita et al. proposed to add artificial lags of 0.2-0.4 seconds only to the channels between NS, to make extra time for NNS to cut into multiparty discussion [6]. N2F conversion also puts extra obstacle to NS, but in a more noticeable way, to bring out gentleness from NS.

Future Directions

Experiments on both NS and NNS are required to examine in detail the effects of N2F accent conversion in listening comprehension and in communication. We would also like to explore N2F accent conversion by direct audio processing, to achieve better accuracy, to convert in real-time, and to maintain voice identity. Based on the work, we will conduct studies on practical systems to support non-native of speakers of English in international communication, and provide cultural shocking experience to native English speakers.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 26870362.

References

1. Derwing, T. M. and Munro, M. J. Accent, intelligibility, and comprehensibility: evidence from four L1s. *Studies in second language acquisition*, 19(01):1-16, 1997.
2. Derwing, T. M. and Munro, M. J. Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly*, 39(3):379-397, 2005.
3. Felps, D., Bortfeld, H., and Gutierrez-Osuna, R. Foreign accent conversion in computer assisted pronunciation training. *Speech Communication*, 51(10):920-932, 2009.
4. Munro, M. J. A Primer on Accent Discrimination in the Canadian Context. *TESL Canada Journal*, 20(2):38-51, 2003.
5. Ogata, J. and Goto, M. Speech Repair: Quick Error Correction Just by Using Selection Operation for Speech Input Interface. In *Proc. Eurospeech '05*, pp. 133-136, 2005.
6. Yamashita, N., Echenique, A., Ishida, T., and Hautasaari, A. Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration. In *Proc. CSCW '13*, 923-934. ACM, 2013.
7. Yamashita, N., Inaba, R., Kuzuoka, H., and Ishida, T. Difficulties in establishing common ground in multiparty groups using machine translation. In *Proc. CHI '09*, 679-688. ACM, 2009.
8. Yamashita, N., and Ishida, T. Effects of machine translation on collaborative work. In *Proc. CSCW '06*, 515-524. ACM, 2006.