

INVITED REVIEW

MyVoice: Rescuing voices of ALS patientsShigeto Kawahara^{1,*}, Musashi Homma², Takaki Yoshimura³ and Takayuki Arai^{4,†}¹*The Keio Institute of Cultural and Linguistic Studies,
2–15–45 Mita, Minato-ku, Tokyo, 108–8345 Japan*²*Department of Rehabilitation, Tokyo Metropolitan Neurological Hospital,
2–6–1 Musashidai, Fuchu, 183–0042 Japan*³*Pasobora,
957–7 Shimo-utagaura, Shikamachi-cho, Sasebo, 859–6204 Japan*⁴*Faculty of Science and Technology, Sophia University,
7–1 Kioi-cho, Chiyoda-ku, Tokyo, 102–8554 Japan***Keywords:** Speech synthesis, QOL, Social welfare, ALS, Japanese phonetics education**PACS number:** 43.10.Ln, 43.10.Sv, 43.70.Jt, 43.71.Bp, 43.72.Kb [doi:10.1250/ast.37.202]**1. INTRODUCTION**

This paper is an up-to-date overview of our collaborative project on “MyVoice,” a system that is intended to rescue voices of those patients who may end up not being able to speak with their own voice. This is an English translation of the accompanying Japanese version, although we have decided to change the exposition in some parts of the English version in order to increase readability.

In everyday communication we use our own voice without appreciating its value too much. However, there are many factors which can cause us to lose our voice. One typical case is due to a motor neuron disease, such as ALS (Amyotrophic Lateral Sclerosis). ALS patients slowly but steadily lose control of their muscular activities, and end up not being able to swallow or to speak. In many cases, a patient has to go through gastric fistula surgery in order to directly inject nutrition into his/her stomach. As a consequence of this surgery, the patient has a hole in his stomach, which makes spontaneous utterances difficult, eventually resulting in the complete loss of his voice.

This problem of losing one’s voice is not specific to ALS patients. One can have one’s larynx removed due to laryngeal cancer, and thus lose one’s voice.

We have been making efforts to rescue the voices of such ALS and other patients in Japan; this article provides the updated overview of what we have achieved as of 2016. The third author, Yoshimura, has developed the free software “MyVoice,” which allows patients to keep

communicating with their own voices by recording basic Japanese CV-moras, before their voices are lost [1–4]. This software is made available for free, and is designed to make the user-interface as simple as possible in order to minimize the burden on users. Since the release of the first version, we have now recorded and rescued the voices of more than 220 patients. Homma (the second author) primarily works directly with the patients, and phoneticians, including Kawahara and Arai (the first and fourth authors), have been helping to improve the quality of MyVoice. This article is intended to introduce MyVoice to a wider general public audience. The major topics to be discussed in this paper include: (i) why it is important to keep communicating with one’s own voice, (ii) how phonetics can contribute to a Japanese speech synthesis system which is based on CV-moras, (iii) how phonetic research can contribute to the general social welfare, (iv) how our activity has been used in our education programs, and how effective it can be.

2. HOW MYVOICE WORKS

MyVoice builds upon HeartyLadder, which is an independently-usable keyboard inputting device [5]. This software allows one to use a keyboard without actually typing on the keyboard; i.e. without using hands. HeartyLadder shows a keyboard screen on a PC’s monitor (Fig. 1), and with this, one can use a computer mouse to select a particular Japanese letter or English alphabet on the screen. For those patients who are not physically able to use a mouse, it also allows them to select a letter by gradually narrowing the selected field. The latter function makes keyboard inputting possible for those patients who can barely move their muscles. For example, HeartyLadder

*e-mail: kawahara@icl.keio.ac.jp

†e-mail: arai@sophia.ac.jp



Fig. 1 An example screen of HeartyLadder [2].

is being used by those ALS patients who can only move a single cheek muscle.

As of 2016, Yoshimura has combined an eye-tracking system (distributed by Tobii) and HeartyLadder. This combination allows patients to use a keyboard by just looking at the letters that they want to enter. This system is named “HeartyAI” (where AI here has a double meaning: “love” in Japanese and “eye” in English). HeartyLadder and HeartyAI can be used both for Japanese letters and English alphabets (although MyVoice is exclusively for Japanese.)

MyVoice is software that plays the sentences inputted in HeartyLadder. To use it, sound files should be stored in a specific folder. The sound files should ideally consist of all of the CV-mora combinations in Japanese, of which there are only about 140 sounds [2], since the overall syllable structure of Japanese is simple. Having these sound files alone allows ALS patients to communicate with their caretakers, or anybody, using their own voices.

There are two fundamental philosophies that underlie MyVoice. One is to make it freely available. The other is to make it as simple as possible. Let us first expand on the first point—both HeartyLadder and MyVoice are made available for free, and in that sense, they differ from other types of similar software [6]. We find it important to keep this principle, since ALS patients suffer from financial pressure due to their medical and other associated expenses. (We are not criticizing other commercial, non-free software in this regard; we are saying that MyVoice has this advantage.)

The second philosophy is to make MyVoice as simple as possible. This aspect of MyVoice is important because we would like our users to be able to use it without requiring much of our help. Our hope is that ultimately, its simple user-interface will allow patients and their caretakers to use MyVoice without our help at all. (In reality, although it is possible to create and use MyVoice without any of our help, many patients do ask for help.) We also would like to emphasize the fact that it is very often the case that ALS patients, by the time they decide to record

their voice for MyVoice, have lost much muscular control due to ALS, and therefore it is not easy for them to spend a lot of time recording. That is why MyVoice is CV-mora based, since this style of recording reduces the number of sound files that need to be recorded. We are aware that speech synthesis is much improved by incorporating VC formant transition information. However, this step requires the recording of VCV sequences, and our honest opinion is that many ALS patients will find this to be too much of a burden. In extreme cases, one can just record vowels and a coda nasal, because the speaker-specific features most clearly appear in these sounds; then onset consonants, recorded by family members or friends can be concatenated later. With this strategy, only six sounds need to be recorded by the patient. We can create CV-moras artificially based on these six sounds, which is sufficient to be able to use MyVoice.

Generally speaking, speech synthesis has two major goals: one is to maximize the fluency or clarity of the synthesized speech; the other is to keep the speaker-specific features of the speech. MyVoice is CV-mora based, so it is inferior to other software in terms of fluency or clarity [6,7]. However, no other similar software that we know of pursues the second goal as much as MyVoice does. According to a comment by Prof. Hiroaki Kato from NiCT at one of our workshops (discussed in further details in Sect. 6), “general speech synthesis software almost always pursues the clarity of the synthesized speech, and considers speaker-specific features to be secondary. MyVoice has the exact opposite philosophy.”

So far, we have stated that MyVoice is based on CV-moras. However, this is a simplification, since MyVoice considers higher levels of prosodic structure, especially syllables. For example, Japanese has a so-called “coda nasal” (*hatsuon*, as in [kanda] “bit”), which constitutes its own mora, but is syllabified with the preceding vowel [8]. We recommend to our patients that if possible, coda nasals should be recorded with their preceding vowels (see Sect. 4 for more on this). Moreover, syllables with long vowels are not synthesized based on a sequence of two short vowels, but by lengthening the short vowel (MyVoice also leaves the option of playing two short vowels). MyVoice creates geminates by lengthening the consonant duration of singleton counterparts, but it also manipulates other features to capture the acoustic characteristics of Japanese geminates. For example, preceding vowels are longer before geminates than singletons in Japanese by about 15–20 ms [8,9], and therefore, MyVoice can automatically lengthen preceding vowels before geminates, which results in better acoustic replications of Japanese geminates. In short, we try to minimize the burden on the patients’ side; at the same time, we incorporate insights from phonetic and phonological research on Japanese in order to make the



Fig. 2 Interval specifications within a CV-mora.

outcome of MyVoice as close as possible to the actual Japanese sounds.

MyVoice also allows users to make subtle adjustments within a CV-mora unit. As shown in Fig. 2, users can specify (i) the onset of the consonant, (ii) the onset of the vowel, (iii) the center of the vowel. By doing so, users can manipulate the duration of a particular interval within a CV-mora. This feature is particularly useful for lengthening or shortening the vowel while keeping the consonant duration constant.

MyVoice can also assign word-level accent. MyVoice contains an accent dictionary of Tokyo Japanese; based on the input information, it assigns the accent patterns of the inputted words. For example, it can distinguish [sake] (unaccented) “liquor” and [sa’ke] “salmon,” or [kurumade] (unaccented) “by car” and [ku’rumade] “until one comes.” Section 3 discusses this mechanism in more detail.

Furthermore, the users are encouraged to record phrases that they often use; for example, the names of their family members, the names of places that they often go, or phrases that they like. MyVoice has a mode in which inputted words are matched by searching from the longest file names in its inventory. For example, if there is a sound file called “arigatoo.wav,” with the input “arigatoo,” MyVoice can play that sound file rather than creating a new sound file based on separate CV-moras.

This feature of MyVoice has an important consequence, beyond being able to play naturally-produced phrases. It leads to collaboration between patients and care-takers, as it encourages them to think together about which phrases and words will be important for them after the patient loses his voice. We have witnessed many cases in which patients and care-takers actually enjoy thinking about these phrases. In a sense, they create their own version of MyVoice together. This collaboration is made possible by the fact that the user interface of MyVoice is simple and user-friendly.

Yet another feature that is worth mentioning is its function to automatically attach the sounds created by

MyVoice to emails. This feature is often used between patients (in fact, this feature was added in response to a patient’s request), and has served as an important communication tool. This feature is also often used to introduce MyVoice to those who are not already familiar with it, for example in university classroom settings, as we will discuss more in Sect. 7.

Finally, we have been focusing on those patients who have lost their voice, but the use of MyVoice is not limited to those people. For example, the utterances of people with cerebral palsy are sometimes hard to understand to those people who are not used to them. However, by recording each mora carefully, they too can use MyVoice, which often leads to better clarity of their utterances (see Fig. 4). In fact, Yoshimura (third author) is a person with cerebral palsy, and he uses his own MyVoice. There are also those people who are self-conscious about their speech; e.g. those with hearing loss or those with a stutter. Those patients sometimes avoid oral communication with other people, because they are not confident about their own utterances, but MyVoice can be used to lower their psychological boundaries. They can make recordings at their own pace when they do not have much psychological pressure, and can create their MyVoice based on these recordings.

3. ITS RAISON D’ETRE: WHY USE YOUR VOICE?

In this section, we would like to consider the “raison d’etre” of MyVoice, by considering why it is important to keep communicating with our own voices. There are alternative ways for patients to communicate with care-takers. For example, they can do so by writing, or by pointing to a list of letters. Moreover, there are other text-to-speech synthesizing software which are, in terms of clarity, better than MyVoice. For example, there is commercial Voiceroid software, which patients can use. To reiterate, why should they keep using their own voice?

We believe that this question is best answered by quoting our patients’ opinions. Let us take the case of ALS again. When an ALS patient starts having difficulty with breathing spontaneously, there are only two options: (i) to have a tracheotomy, which entails the loss of one’s voice, or (ii) to wait to die. However, MyVoice offers the third possibility—to have a tracheotomy without losing their voice. We know a number of ALS patients who have decided to have a tracheotomy (partly) because of MyVoice. One patient was choosing to die, if a tracheotomy entailed losing his voice, but MyVoice changed his mind. We think that we are justified in saying that MyVoice saved his life.

Let us quote one patient’s remark to further illustrate our point: “For me, recording my own voice for MyVoice

was tantamount to getting ready to live longer. By practicing using MyVoice and thinking about the phrases that I wanted to record, it necessarily got me thinking about what the life after a tracheotomy was going to be like. That experience allowed me to decide to have a tracheotomy in 2011 without too much hesitation.” For this patient, it is clear that MyVoice encouraged him to have a tracheotomy in order to live longer.

We are tempted to quote a remark from every other patient of ours, but we limit ourselves to some representative examples: “My voice is really my identity”; “my joy of life comes from the fact that I can keep speaking with my own voice.” ALS is an irreversible, incurable disease. ALS patients keep losing their ability to do things on their own day by day. In this situation, being able to generate new messages with their own voices can make them think that “my life is not just about losing [10].” Another patient appreciated the fact that with MyVoice, they can create messages with their own voice to give to their children, as they grow up. Their children may get married after they lose their voice. In that case, “congratulations” created with their own voice would mean a lot, both for the parent and the child.

Some MyVoice users report that “using MyVoice makes me feel like I am really speaking.” This statement is not groundless, according to a recent experiment. An experiment using NIRS (Near Infra-Red Spectroscopy) shows that when the participants hear their own voice, activity is observed in the premotor cortex area [11]. This result might imply that hearing our own voice may make our brain think that we are actually speaking.

ALS patients can end up not being able to move any parts of their body at all, but research has shown that their consciousness remains clear, even in the late stages of the disease [12]. Given that situation, being able to talk with their own voice can be a big help for surviving, especially given its effect on our brain.

Using the patient’s own voice is not just beneficial for the patient, but also beneficial for the care-taker. Taking care of an ALS patient is not an easy task. But being able to hear the patient’s voice during such difficult times can reduce the emotional burden. We have heard opinions such as “I can keep listening to sounds from MyVoice, as long as they are based on my husband’s voice,” “I feel very encouraged to keep taking care of my partner, thanks to MyVoice” or even “I feel like I can argue with my husband thanks to MyVoice, because I don’t have to feel bad that he cannot speak—he can.”

4. MYVOICE AND LINGUISTIC RESEARCH

We would now like to move on to the discussion of how phoneticians—or linguists in general—have been contributing to MyVoice.

Let us start by briefly reviewing the history of MyVoice. Yoshimura and Homma started developing MyVoice in 2005. It was not until 2013 when Kawahara joined this project. Yoshimura’s expertise is in programming, and Homma is an occupational therapist. Neither of them had explicit training in phonetics; they nevertheless kept trying to improve the phonetic quality of MyVoice. With hindsight, the “discoveries” that they had made make phonetic sense. In other words, if the collaboration with phoneticians had started earlier, the development of MyVoice would have been easier; but on the other hand, it seems safe to say that Yoshimura and Homma both had “the right sense” about Japanese phonetics and phonology.

Let us discuss some examples. As discussed in Sect. 2, it is much better to record the coda nasal with the preceding vowel. Yoshimura and Homma “discovered” this fact by themselves. For example, for a word like *shinkansen* “bullet train,” it is better to record [ʃiN], [kaN], and [seN]—which is syllable-based—than to record [ʃi], [N], [ka], [N], [se], and [N]—which is mora-based. Phonetically speaking, this is most likely due to the fact that Japanese vowels followed by a tautosyllabic nasal consonant are nasalized [8], and the syllable-based recording would reflect this nasalization. More generally speaking, the coda nasal and the preceding vowel are grouped into the same syllable, and hence show several phonetic interactions [8,9]. Therefore, the general conclusion is that it is better to record syllables, if we can, than to record moras (although it is not practical to record syllables with geminates in isolation.)

Yoshimura and Homma also realized that word-initial sounds and word-internal sounds are acoustically different, and they had thus recommended that, if possible, these sounds are separately recorded. This observation also makes phonetic sense. Word-initial sounds are usually more robustly articulated than word-internal words. Word-initial syllables in Japanese usually bear a L-tone [13], and in that sense too, they may differ from word-internal syllables.

They also “discovered” that Japanese is a pitch-accent language, the primary acoustic correlate being differences in fundamental frequencies. They noticed that by preparing H-toned sounds and L-toned sounds, they can reproduce Japanese accent systems. Furthermore, they even “discovered” that the syllable before the pitch fall is higher than the general H-toned syllables, an observation that is well known in the phonetic literature [13]. These examples show that as of 2013, MyVoice had incorporated many phonetic and phonological features of Japanese. Nevertheless, our collaboration resulted in further improvement.

For example, the version of MyVoice before 2013 adjusted the duration of each mora by specifying how

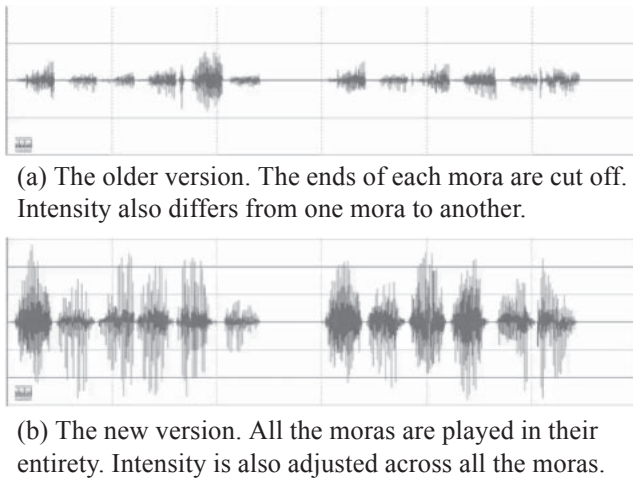


Fig. 3 Two methods for adjusting duration.

many ms each mora should last, and by cutting off the rest. This method has a non-negligible problem, however. For example, [s] is longer than other consonants. Therefore, by setting a certain ms threshold—say 200 ms—which would be suitable for moras with other onset consonants, it could result in not playing the vowel after [s], because the [s] is long. This problem was solved by making use of the overlap-and-add method, which allows us to lengthen or shrink the whole CV-mora to the specified target duration (Fig. 3). (MyVoice also still allows an option of adjusting duration in the old way, and in this mode, if a particular mora is too short, it is automatically lengthened.)

The second improvement has to do with how to generate H-toned sounds and L-toned sounds. In order to realize the Japanese pitch accent system, we at least need H-toned and L-toned sounds. For example, [a'me] “rain” is HL, whereas [ame] “candy” is LH. To generate L- and H-toned sounds, MyVoice used to manipulate wavelength (because frequency f is an inverse function of wavelength λ). However, this method has the drawback of changing the duration of sounds. Applying the overlap-and-add method to manipulate frequency resulted in generating H-toned and L-toned moras in a more natural manner.

Also, it is impossible for patients to adjust the intensity of all of the moras at the time of recording. This problem was solved by artificially adjusting the intensity of all moras after the recording. Consider Fig. 3 again. In the old version (3a), the intensity of the moras is inconsistent, but in the new version (3b), the intensity of all of the moras is adjusted.

Kawahara first used Praat [14] to automate these processes. Scripts [15] were written so that they processed all the sounds in a folder specified by MyVoice. As of 2016, however, MyVoice can run these processes inter-

nally. This is an important improvement, because using Praat can be difficult for patients and care-takers, since Praat’s interface is available only in English.

Phoneticians also contributed to the improvement of the acoustic realizations of geminate consonants. For example, an old version of MyVoice generated geminates by inserting silence. This method works well for stop geminates. However, for fricative geminates, it is better to lengthen the frication than to insert silence. Also, when the coda nasal appears before a nasal onset consonant, these two sounds realize as a geminate nasal consonant, instead of a sequence of a coda nasal and an onset consonant. These examples show that suggestions from phoneticians can help improve the quality of MyVoice.

Phoneticians also contributed by making suggestions about how to extract each mora from a long recording file. There was once a complaint that it was difficult to extract moras with onset /h/. Phonetically speaking, indeed, Japanese /h/ has non-strident, weak frication (realized variably as [h], [ç], or [ϕ] depending on the following vowel). Since MyVoice’s interface to extract moras uses waveforms, not spectrograms, it is easy to miss /h/’s frication. However, just a little advice to watch out for the weakness of /h/’s frication can help avoid the problem of missing it. Another possibility is to incorporate spectrograms in the sound editing window, because /h/’s frication would be visible in high frequency components, but this has not happened as of 2016.

Arai also made a suggestion about how to record /r/ in Japanese. The phonetic realization of Japanese /r/ can differ rather substantially between word-initial positions and word-internal positions [16]. Recording /ra/, /ri/, /ru/, /re/, /ro/ with a pause between each mora results in all the moras recorded as they would be pronounced in word-initial position. Given the fact that Japanese /r/ often appears word-internally, Arai suggested that /r/ sounds be recorded intervocalically instead, as in /ara/, /iri/, /uru/, /ere/, /oro/.

5. REMAINING ISSUES

Although we believe we have made substantive progress, there are many remaining issues. One of the biggest issues is, to the extent that MyVoice is based on CV-mora recordings, it is unable to incorporate VC formant transition information. In order to do so, it would be necessary to record VCV units, but that would make the recording list much longer, which goes against the basic philosophy of MyVoice (recall that we want to keep MyVoice as simple as possible, so as to minimize the burden on the patients). This dilemma is yet to be solved. Arai suggested one possible solution to this problem. We can potentially reverse the time course of recorded CV-moras to create VC formant transition, and use that as an

approximation of VC formant transition in MyVoice's speech. We have yet to implement this suggested method, but it can potentially offer a (partial) solution to the above mentioned dilemma.

Another thorny problem is that while MyVoice can currently realize word-level accent, it is incapable of realizing phrasal or sentence-level intonation. An exception is the rising intonation for questions, which can be done if the patients are able to record moras with rising intonation. But no other types of sentential intonation can be produced, and this is a task yet remaining for phoneticians and other types of linguists. For phrase-level intonation, it is likely that we need some system for syntactic-parsing, which goes beyond the realm of phonetics *per se*.

Many patients also inquired whether MyVoice can realize accent patterns for dialects other than Tokyo Japanese. The short answer is that it cannot do so now. The long answer is that it will be able to, as long as there is an accent database for other dialects, in a format that can be read by a computer program like MyVoice. This is a task that is suitable for Japanese phonologists, because accentology in different Japanese dialects is an actively investigated area. Relatedly, Japanese dialects also differ at the level of segmental phonology, for example in terms of the presence/absence of high vowel devoicing. Many Japanese phonologists work on dialectal differences, and our hope is that contributions will be made in such a way that they can be incorporated into MyVoice, and other similar software.

Another big challenge is bone conduction. Anybody who has recorded their own voice and listened to it knows that what we usually "hear" when we speak is different from what is recorded. This is because bone conduction enhances the low frequency energy of our speech. Many patients are, unfortunately, disappointed to hear their first version of MyVoice, thinking that the sounds produced by MyVoice do not sound like their own voice; we often hear that "the sounds are too high-pitched." This is partly due to the effect of bone conduction. We have tried to enhance low-frequency energies while weakening high-frequency energies to mimic the effect of bone conduction, but have not succeeded in this task.

The issue of bone conduction is more complicated than just solving the technical issue of how to reproduce the enhanced low-frequency energies. If MyVoice is for care-takers, then the problem of bone conduction does not arise, because sounds without bone conduction are what the care-takers are used to hearing. At any rate, the more options, the better, and therefore, we continue in our effort to develop a system to reproduce the effect of bone conduction.

Another general issue to be addressed is what kind of microphone should be used for MyVoice recording. If the

only aim of MyVoice is to pursue the best quality of sounds recorded, we should use a dynamic microphone or shotgun microphone to exclusively target the patient's voices. However, when we speak and listen to our own voice, our speech is not directly conveyed to our ears. Thus, some patients actually prefer sounds recorded by a condenser microphone, saying that sounds recorded by a condenser microphone sound more like their own voice.

Also, by the time the patients have decided to record their voice, often their vocal folds are weakened so much that they can produce only a whisper voice. In such cases, it should be possible to extract their vocal tract information based on the whisper voice, and synthesize it with an artificially produced modal voice. This technique should be able to replicate the patients' modal voice.

On the other hand, however, in response to our proposal of this kind, one patient responded that "the whispered voice is my own voice, because that's the voice that I have been using while I am suffering from my disease." This response suggests, again, that fluency is not all that is expected of MyVoice. After all, if fluency is all we want, there is no need to use MyVoice. To summarize, the remaining task for phoneticians and speech scientists is to improve the quality of MyVoice sounds, while keeping each patient's speaker-specific characteristics. As discussed throughout this paper, we have been making efforts to improve the quality of MyVoice, but as of yet, not all patients are completely happy with it.

6. PUBLICITY

So far, we have been discussing how phonetic research can and should contribute to the improvement of MyVoice. Another way in which researchers can contribute to MyVoice is publicity. To that end, we are regularly hosting a workshop on MyVoice at Keio University. The workshop is open to everybody, and the major participants include the patients who are using MyVoice, their care-takers, and occupational therapists who want to learn how to use MyVoice [17,18]. Topics vary from one meeting to another; often, we have a tutorial on how to use MyVoice; we at the same time discuss various topics on MyVoice. We also ask the patients and care-takers to tell us how they feel about MyVoice, some of which has been quoted in this paper. In addition, sometimes our students present their research on MyVoice.

Figure 4 presents a slide presented by a group of occupational therapists in November 2015. The presentation was about the process of how they collaborated to create MyVoice for a person with cerebral palsy.

Our efforts in this area have had a positive impact in that more and more people are interested in MyVoice. However, we need more man-power. Ideally, we should have more people who can record and edit sounds to create

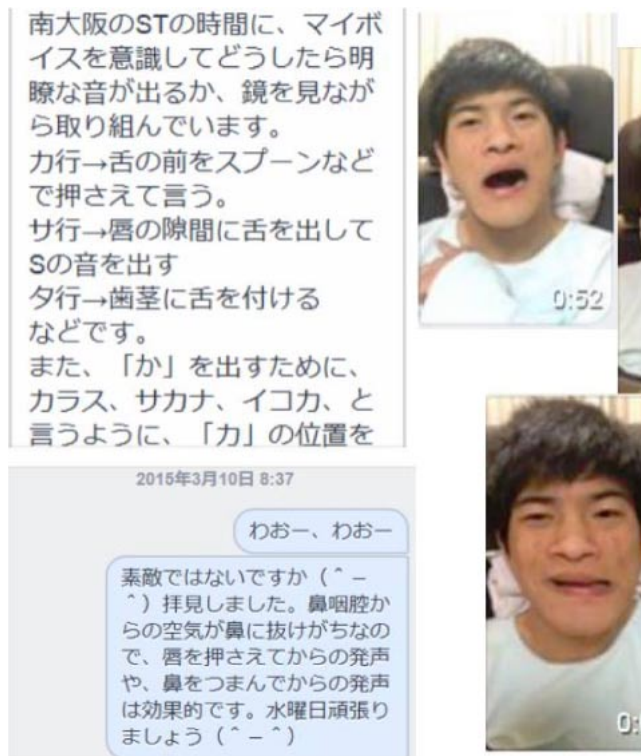


Fig. 4 A slide from the presentation at one of the Keio Workshop meetings on MyVoice. The left panel shows the discussion thread on how the occupational therapists tried to have the person with cerebral palsy produce Japanese consonants. The first comment includes a statement, “In order for him to produce [k], it would help to depress his tongue tip with a spoon.” The other person responds that “sounds great, he tends to have an air flow leakage from the nasal cavity, so it would be effective to pinch his nose while recording.” The right panel is the picture of the person with cerebral palsy (quoted with permission).

MyVoice. We plan to continue holding workshops about MyVoice, and hope that more people will become involved with this project.

7. MYVOICE IN HIGHER EDUCATION

We have found that MyVoice can provide good teaching materials in university education. Except for those who go to a medical school or related fields, undergraduate students rarely get a chance to interact with those who are involved with actual medical scenes. However, MyVoice makes it possible for students to interact with ALS patients directly, which is definitely an influential experience.

For example, we took some of our students to visit Homma at Tokyo Neurological Hospital, and had them interact with ALS patients. The workshops on MyVoice also offer opportunities for students to get to see ALS patients who use MyVoice. Interacting with those patients

who actually use MyVoice can be one of the best ways to appreciate its value.

Kawahara and Arai also discuss MyVoice in their undergraduate classes, with the goal of encouraging students to think about the importance of speaking with their own voices. Kawahara started asking students to write a letter to Yoshimura and Homma about how they have felt about the MyVoice project, and Yoshimura and Homma actually responded to these letters. This interactive approach to undergraduate teaching has turned out to be very effective as well.

Many undergraduate students who are majoring in linguistics do so simply because “linguistics is fun.” However, some of them do wonder at the same time how linguistics can be applied in actual life. MyVoice can provide a concrete answer to such a question, connecting their study at a university to actual societal welfare. One student was interested in computational linguistics, and was very inspired by Yoshimura’s work. He wrote to him in his letter saying that he wants to become like him. Another student, who has been helping with this project of ours for more than two years now, conducted a questionnaire survey at one of the workshops, and presented about MyVoice at the Keio Academic Skills Presentation Competition. She won the best presenter award [19].

To summarize, our impression is that MyVoice has the power to directly appeal to students. Many students wonder why they have to learn phonetics; MyVoice is one very good way to show them why. It also helps them to appreciate the relationship between science, technology and society. Indeed, students from different universities — Sophia University, Keio University, Osaka University, and Tokyo Metropolitan University — have written BA theses on MyVoice [10,20–22].

We may even be able to turn around and say that we can use MyVoice to teach phonetic and phonological notions. We have witnessed many cases in which those people who had never been interested in sounds became attracted to phonetics, thanks to MyVoice. The reality is that phonetics is not necessarily the easiest subject to study, or teach, but MyVoice may be able to lower the psychological boundary by providing concrete motivations. We hope that we can make more use of MyVoice in higher education in Japan.

8. CONCLUSION

MyVoice is a general project to rescue the voices of ALS and other patients. Our own voices are in a sense our own identity. We have argued that it is non-trivial that we offer free software by which patients can save their own voices and play sentences based on them. We have also argued that MyVoice offers an opportunity for the field of phonetics to relate itself to actual society; this is especially

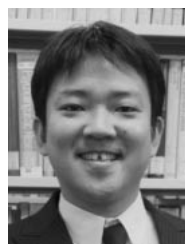
meaningful for students. We strongly hope that more and more people will join the club, and help develop the MyVoice project.

ACKNOWLEDGEMENTS

We would like to thank all those patients, care-takers, and occupational therapists, who are using, and who have used, MyVoice. Yuko Imazeki has been offering indispensable help with recording and setting up MyVoice. Prof. Ken-ichiro Yabu has offered great advice and assistance from the view point of phonetics from the outset of this project. This project is supported by Fukuzawa Academic Development Funds to the first author and a grant to the Center for Life-Span Development of Communication Skills (PI: Yasuyo Minagawa) from Keio University. We also would like to thank Mai Michishita, Tomoko Monou, and Akiko Takemura who provided comments on the Japanese version of this paper, and Donna Erickson and Helen Stickney for their careful proofreading of this English version.

REFERENCES

- [1] M. Homma and M. Nagao, "Jibun-no koe-o nokosu (saishougen-no hito-no koe-no rokuon-niyoru kikitariyasui onsei saisei)," *Toukyouto Byouin Keiei Honbu Rinshoo Kenkyuu Houkokusho*, pp. 71–77 (2013) (in Japanese).
- [2] S. Kawahara, M. Homma, Y. Imazeki, T. Yoshimura, M. Ogiwara, H. Fukazawa, H. Masuda, K. Shinohara, Y. Sugioka and Y. Sugiyama, "MyVoice: How linguists can (and should) do," *Phonological Stud.*, **18**, 127–136 (2015) (in Japanese).
- [3] M. Yoshioka, T. Arai, K. Yasu and H. Otsuki, "Hatsuwasha jishin-no moora onsei-ni motozuku tekisuto onseigousei-ni okeru kentou: Onsei-no kihonshuuhasuu-ya supekutoru tokusei-ni kanshite," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 1239–1242 (2015) (in Japanese).
- [4] S. Kawahara, *Oto-to Kotoba-no Hushigina Sekai*, Iwanami Science Library (Iwanami Shoten, Tokyo, 2015) (in Japanese).
- [5] <http://takaki.la.coocan.jp/hearty/> (accessed 2016-07-24).
- [6] <http://hits.kkhts.com/Solution/Voistar/home/> (accessed 2016-07-24).
- [7] J. Yamagishi, C. Veaux, S. King and S. Renals, "Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction," *Acoust. Sci. & Tech.*, **33**, 1–5 (2012).
- [8] S. Kawahara, "Japanese has syllables: A reply to Labrune (2012)," *Phonology*, **33**, 169–194 (2016).
- [9] S. Kawahara, "The acoustics and perception of Japanese special moras," *J. Acoust. Soc. Jpn. (J)*, **69**, 191–196 (2013) (in Japanese).
- [10] M. Ogiwara, "Shissei kanousei-no aru kanjya-no tame-no komyunikeeshon kiki "MyVoice"-no sakusei-to sono hitsuyousei-ni kansuru kentou," *Senior Thesis, Tokyo Metropolitan University* (2013) (in Japanese).
- [11] Y. Tamura and Y. Minagawa, "Koe-o kikukoto niyoru noukatsudou-no henka: Jibun-no koe-to jinkou onsei-no hikaku," Presentation given at *Keio MyVoice Workshop 3* (2015) (in Japanese).
- [12] H. Koizumi, *Nou-wa Deai-de Sodatsu: Noukagaku-to Kyouiku Nyuumon* (Seidosha, Tokyo, 2005) (in Japanese).
- [13] J. Pierrehumbert and M. Beckman, *Japanese Tone Structure* (MIT Press, Cambridge, Mass., 1988).
- [14] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott Int.*, **5**, 341–345 (2001).
- [15] <http://user.keio.ac.jp/~kawahara/myvoice.html> (accessed 2016-07-24).
- [16] T. Arai, "On why Japanese /r/ sounds are difficult for children to acquire," *INTERSPEECH 2013*, pp. 2445–2449 (2013).
- [17] <http://user.keio.ac.jp/~kawahara/MyVoiceMeetings.html> (accessed 2016-07-24).
- [18] <https://www.youtube.com/watch?v=nA1YH7ZyKZE&list=PLWXQYx-RCmeNgKE2P2R4M7eqiJBwXWnLz> (accessed 2016-07-24).
- [19] M. Tsuchiya, "Jibun-no koe-o riyoo-shita jyuudo shougai-shayou ishidentatsusouchi-no yuukousei-to hukuuyujitai: MyVoice-to sono riyoo-sha-o rei-ni," *Keio University Academic Skills Presentation Competition* (2016) (in Japanese).
- [20] H. Otsuki, "Hatsuwashajishin-no onsei-ni motozuku tekisuto onseigousei niokeru firuta shori-no kouka: Hatsuwa konnansha shien-o mezashite," *Senior Thesis, Sophia University* (2014) (in Japanese).
- [21] M. Yoshioka, "Kihon shuuhasuu-no baratsuki-ga tekisuto onseigousei-no shutsuryokuon-ni ataeru eikyuu," *Senior Thesis, Sophia University* (2014) (in Japanese).
- [22] Y. Arayama, "A study of Japanese and English synthesized speech," *Senior Thesis, Osaka University* (2016).



Prof. Shigeto Kawahara obtained his PhD in linguistics from the University of Massachusetts 2007. He was an assistant professor at the University of Georgia (2007–2008) and Rutgers University (2008–2013). Currently, he is an associate professor of the Keio Institute of Cultural and Linguistic Studies. His research centers around phonetics, phonology and their interface. He also works on how phonetic studies can and should contribute to social welfare. He has published many academic articles in peer-reviewed journals, and recently published an introductory phonetics textbook from Iwanami Publisher. As of 2016, he is the project leader for the Phonetics Society of Japan, the young scholar network for the Science Council of Japan.



Musashi Homma is an occupational therapist. After getting his degree in 1986, he worked in different divisions of various hospitals. In 2003, he moved to his current hospital, Tokyo Neurological Hospital, where he started working on helping patients with incurable diseases. He has been developing ways to assist ALS patients to compensate for the loss of their muscular controls, and more generally keep their QOL high by inventing new creative devices. He is mostly interested in making use of recent IT technologies to facilitate communications with the patients and caretakers. He also enjoys working with patients and their families directly.



Takaki Yoshimura graduated from the Department of Social Welfare at Bukkyo University. He is currently working for LaboTec in Sasebo City, and is also an active member of a volunteer group “Pasobora Kokorono Kakehashi.” He developed HeartyLadder in 2000, which allows users to use a keyboard without typing. He is also the main developer of MyVoice. He keeps updating HeartyLadder

and MyVoice in response to comments and requests by the actual users. He also published a book in 2001 from Kobunsha about his life, focusing on how he became interested in developing software like HeartyLadder.



Prof. Takayuki Arai received the B.E., M.E. and Ph.D. degrees in electrical engineering from Sophia Univ., Tokyo, Japan, in 1989, 1991 and 1994, respectively. In 1992–1993 and 1995–1996, he was with Oregon Graduate Institute of Science and Technology (Portland, OR, USA). In 1997–1998, he was with International Computer Science Institute (Berkeley, CA, USA). He is currently Professor of the Department of

Information and Communication Sciences, Sophia Univ. In 2003–2004, he is a visiting scientist at Massachusetts Institute of Technology (Cambridge, MA, USA). His research interests include signal processing, acoustics, speech and hearing sciences, spoken language processing, and acoustic phonetics.