INNOVATION IN LANGUAGE LEARNING AND TEACHING
https://doi.org/10.1080/17501229.2024.2315101

I Can Speak: improving English pronunciation through automatic speech recognition-based language learning systems

Muzakki Bashori [a,e], Roeland van Hout [a], Helmer Strik [a,b,c,d] and Catia Cucchiarini [b]

[a]Centre for Language Studies, Radboud University, Nijmegen, the Netherlands; [b]Centre for Language and Speech Technology (CLST), Radboud University, Nijmegen, the Netherlands; [c]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands; [d]NovoLearning BV, Nijmegen, the Netherlands; [e]Faculty of Social and Political Sciences, Semarang State University, Semarang, Indonesia

ABSTRACT

Practicing pronunciation through language learning systems incorporating Automatic Speech Recognition (ASR) technology has been effective in helping improve foreign language pronunciation. One of the ASR affordances is that it can provide immediate, personalized feedback on learners' pronunciation. We investigated the effects of two ASR-based language learning systems, *I Love Indonesia* (ILI) and *NovoLearning* (NOVO), on learners' word-level and sentence-level pronunciation. ILI offers global corrective feedback, while NOVO is equipped with corrective feedback on phonetic details. 117 Indonesian high school students participated in a five-week-long experiment; 52 students used ILI and 65 NOVO. Three pronunciation measures were calculated on a pre and post reading test: phonetic edit distances plus accentedness and comprehensibility ratings. Results indicate significant improvements in learners' pronunciation, confirming that both systems are promising learning tools, with NOVO leading to more progress. Future studies should examine the long-term effects of ASR-based global and phonetic feedback on learners' pronunciation quality.

ARTICLE HISTORY
Received 22 March 2022
Accepted 1 February 2024

KEYWORDS
Automatic speech recognition; pronunciation; phonetic distance; accentedness; comprehensibility

1. Introduction

Research has shown that Automatic Speech Recognition (ASR) technology can be fruitfully employed to develop Foreign Language (FL) learning systems that provide speaking practice with instantaneous and individualized feedback. ASR-based practice can help improve learners' speaking performance, especially pronunciation (Cucchiarini and Strik 2017; Cucchiarini, Neri, and Strik 2009; McCrocklin 2016; Mroz 2018a), even when ASR accuracy is not 100 percent (Daniels and Iwago 2017; Golonka et al. 2014).

ASR technology can provide corrective feedback on pronunciation in different ways. Corrective feedback can be given globally by indicating whether a word or sentence was pronounced correctly (Evers and Chen 2020; Mroz 2018b; Neri et al. 2008) without further details. On the other hand, the feedback can return phonetic details by highlighting phonetic features that show precisely which speech sounds were mispronounced (Castelo 2022; Tejedor-García et al. 2020). Both global and specific phonetic types of corrective feedback effectively enhance learners' pronunciation.

CONTACT Muzakki Bashori muzakkibashori90@gmail.com Centre for Language Studies, Radboud University, 6525 XZ Nijmegen, the Netherlands; Faculty of Social and Political Sciences, Semarang State University, Semarang 50229, Indonesia

The need to adopt and personalize ASR for pronunciation teaching is particularly felt in – including, but not limited to – Asian countries, where English Foreign Language (EFL) classrooms are crowded, and most learners generally do not receive sufficient oral practice and positive feedback. Besides, teachers feel insecure about teaching pronunciation (e.g. Breitkreutz, Derwing, and Rossiter 2001; Foote, Holtby, and Derwing 2011), and feedback to language learners is limited (Burns 2012) and frequently erratic (Iwashita 2003).

ASR also helps create a less anxiety-provoking learning environment, which can benefit learners with higher levels of Foreign Language Speaking Anxiety (FLSA), who do not need to worry about being humiliated or laughed at because of their mispronunciations (Bashori et al. 2020). This is especially useful for EFL learners in Indonesia, where contextual constraints such as students' higher levels of speaking anxiety (Bashori et al. 2021), large classroom size, and limited teaching time for speaking (Ariatna 2016) seem to hinder learners in increasing their speaking proficiency. As reported by the Education First English Proficiency Index (EF EPI), over the last six years (2017-2022), Indonesia was always included in the category of low-proficiency countries (Education First 2022). However, relatively little research has explored the potential effects of ASR-based practice on secondary school students' pronunciation in such contexts. In this study, we intend to bridge this gap by providing evidence and perspectives that shed more light on how ASR-based practice specifically triggers improving students' word-level and sentence-level pronunciation.

The present study investigates to what extent two ASR-based language learning systems with two different types of feedback, *I Love Indonesia* (ILI) and *NovoLearning* (NOVO), used in the *I Can Speak* program, contribute to learners' actual pronunciation achievements. ILI and NOVO were selected because of their promising affordances in EFL learning when applied in the Indonesian context (Bashori et al. 2020; 2021). In addition, by adopting both ILI and NOVO in our study, we get a more general picture, by not depending on only one system. Moreover, there is an essential distinction between the two systems concerning pronunciation. ILI generates global feedback, while NOVO offers phonetic feedback. We formulated four research questions:

1. To what extent do ILI and NOVO help students improve their pronunciation?
2. Are there differences in pronunciation improvement between ILI and NOVO?
3. How and to what extent are our three pronunciation measures (phonetic distance, accentedness, comprehensibility) related?
4. What do experts (nine raters and a phonetic transcriber) say to observe and experience when giving ratings or making phonetic transcriptions?

2. Automatic speech recognition for pronunciation learning

Automatic Speech Recognition (ASR) is a product of technological advancement that can perform the function of '… decoding and transcribing oral speech' (Levis and Suvorov 2013, 1). LaRocca, Morgan, and Bellinger (1999) explain that investigations of ASR use for Second Language (L2) learning were started in the 1990s under research programs conducted by the U.S. military, aiming at discovering ways for cadets to practice speaking by themselves. Since then, the number of studies on the development of ASR and the effects of ASR-based FL/L2 learning have increased rapidly (e.g. Ehsani and Knodt 1998; Eskenazi 1999; Chen 2011; Chiu, Liou, and Yeh 2007). Neri et al. (2008) found that ASR-based practice in which global feedback such as 'well done' is employed improves learners' pronunciation quality. Even a short duration of ASR-based training can significantly affect learners' pronunciation achievements (McCrocklin 2016). ASR can also be employed in designing so-called speaking practice partners that stimulate learners to make their pronunciation more comprehensible (Mroz 2018a).

ASR-based applications in FL pronunciation learning have mushroomed in recent years, incorporating various teaching techniques (e.g. individual work, pair work) and feedback-providing methods

(e.g. ASR feedback alone, peer plus ASR feedback). Discussions have then advanced to which approach is more successful in helping learners enhance their pronunciation quality. Elimat and Abu-Seileek (2014) investigated learners in three experimental groups that practiced through an ASR-based system (individual work, pair work, and group work). Results indicate that individual work is the most effective teaching technique in improving learners' pronunciation. Liakin, Cardoso, and Liakina (2014) also found that learners who practice with an ASR-based system and receive immediate feedback (no human interaction noted) surpass those supported with teacher-based pronunciation feedback and those without teacher feedback. Studies such as Hincks (2003) and McCrocklin (2016) support the effectiveness of individual practice, but point out that pronunciation practice through ASR-based systems alone can cause frustration (McCrocklin 2016), which may, in turn, reduce learners' motivation.

The results by Dai and Wu's (2021) mixed-methods study provide a different perspective. Learners who received peer plus ASR-based feedback and those who only obtained peer feedback (without ASR) outperformed those who followed an autonomous self-practice ASR-based session. ASR-based practice supported by peer assistance also helps learners reduce anxiety (Nakazawa 2012). Evers and Chen (2020) observed differences in adults' pronunciation attainment when using an ASR system supported with or without peer feedback. They found that embedded peer feedback was more effective than individual practice for correcting pronunciation.

Using ASR for pronunciation learning also offers considerable benefits to FL teachers. Couper (2017) interviewed 19 EFL teachers and found that they suffered from a lack of knowledge of phonetics, including the production of sounds, stress, rhythm, or intonation, and how to teach these complex aspects of pronunciation. To provide effective pronunciation teaching, phonetic feedback is of the utmost importance (Baker and Burri 2016), but teachers have limited teaching time in their classrooms (Ariatna 2016). Breitkreutz, Derwing, and Rossiter (2001) and Jenkins (2007) discovered that many non-native EFL teachers feel uncertain about teaching pronunciation. In Bashori et al.'s (2020) interviews, one of the teachers' concerns was that students who sit at the back sometimes fail to hear the teacher's words correctly. This indicates that it may not be possible for teachers to provide adequate feedback on pronunciation to every student. These abovementioned situations suggest that technologies such as ASR can be a helpful partner that enables today's teachers to support their students' pronunciation learning, both in the classroom and outside the class, and both in individual work (without peer feedback) and pair/group work (with peer/group feedback).

3. Methodology

3.1. Participants

One hundred twenty-eight students from four class groups at a vocational high school in Indonesia participated in this study. All students were 10th graders, aged between 14 and 17 years, and were from three different study programs: Mechanical Engineering (two groups), Mechatronics Engineering, and Industrial Electronics Engineering. Most participants (n = 125) were male, probably because boys prefer these study programs.

Two class groups were asked to use ILI, while the others practiced on NOVO (see Table 1).

Table 1. The school's majors, research treatment, English teachers, and participants.

Class	Treatment	Teacher	Participants
Mechanical Engineering 1	ILI	ET01	31
Mechanical Engineering 2	NOVO	ET01	32
Mechatronics Engineering	ILI	ET02	31
Industrial Electronics Engineering	NOVO	ET03	34

A small survey distributed before the experiment revealed that most students (56.3%) started to learn English between five and ten years. They rated their English speaking skills as *poor* (n = 87), followed by *moderate* (n = 41). They stated that they spoke English (not including at school) *seldom (only one or two days a week)* (47.7%) or *never* (45.3%) and indicated finding it moderately difficult (74 students) or difficult (53); only one student chose *easy*. Most of the students (59.4%) reported having *moderate* motivation in learning English, followed by *high* (31.2%) and *low* (9.4%).

3.2. Procedure and instruments

This study followed four steps: (1) administering an English proficiency test, (2) recording pronunciation (pre-test), (3) conducting two ASR-based experiments, and (4) recording pronunciation (post-test).

3.2.1. English proficiency test

We employed a 15-minute Standard English Test by Education First (EF) (https://www.efset.org/quick-check/) to estimate the students' levels of English proficiency. The test consisted of 20 questions, including reading (grammar and vocabulary) and listening skills. Results were aligned to CEFR levels.

3.2.2. Pronunciation test and evaluation

The *Audacity* software was used to record the students' speech on a one-by-one basis at the school's computer laboratory by the first author before and after the experiment. Twelve students who did not join the pre-test and ten students who were absent from the post-test session were contacted personally and asked to send their audio files via *WhatsApp*.

There were 28 words and 28 phrases/sentences (see Appendix A for the complete list) selected from seven speaking units adapted from the speaking materials provided freely by the British Council (https://learnenglishteens.britishcouncil.org/skills/speaking) that the participants had to read aloud and record through the software. The rationale behind this selection is that these words and phrases/sentences contain problematic segments for many Indonesian EFL learners (Swan and Smith 2001). The read-aloud technique to assess pronunciation before and after the intervention was also employed in other studies, such as Evers and Chen (2020) and Saito (2011).

To analyze the word-based recordings, we invited *Rahayu* (pseudonym), a highly proficient Indonesian speaker of English (with an overall IELTS score of 8.5) currently working as a university lecturer. She made phonetic transcriptions of all the recorded word utterances (n = 28 target words x 2 pre–post x 128 participants = 7,168 transcribed words) and was paid an incentive. At the beginning stage, the first author checked her transcription samples from two participants (n = 2 participants x 28 target words = 56 transcribed words) and discussed the findings with her. Overall, their consensus was nearly perfect. They kept discussing two challenging aspects, the complicated distinction between long and short vowels and word stress. The short-long distinction is unfamiliar to Indonesian learners since long vowels are absent in the Indonesian sound system (Perwitasari 2013). The transcriptions agreed by the first author and the transcriber contain only short vowels, except for some rare utterances that sounded relatively long.

Additionally, (word) stress in Indonesian is essentially free, not fixed (Van Zanten and Van Heuven 2004). The expert and the first author agreed to employ the expert's final judgment. Finally, the first author rechecked a subset of the phonetic transcriptions (n = 8 participants x 28 target words x 2 pre–post = 448 transcribed words). The first author agreed with these transcription samples.

Next, we measured the phonetic distances between the learners' realizations and the target pronunciations by comparing the IPA Reference Transcription (RT) and the Actual Transcription (AT). The RT was based on the Oxford Learner's Dictionaries American English pronunciation as this pronunciation is more familiar to teachers and learners in the educational settings in Indonesia. At the same

time, the AT is a phonetic transcription of the learners' pronunciations. The distances were measured using the Levenshtein algorithm and normalized by the alignment length; the IPA transcription of the target pronunciations defines the alignment length.

To examine the sentence-based recordings, we collected the judgments of nine raters (three males and six females) (see Appendix B for demographic information). Seven raters were professional teachers of English in Indonesia (non-native speakers of English). The other two raters were a Master's student and a website translator. The raters' English proficiency levels varied between B2 and C1 (according to their TOEFL/IELTS scores). The recordings ($n = 28$ target sentences x 2 pre–post x 128 participants = 7,168 recorded sentences) were anonymized and rated independently. The raters were instructed online through *WhatsApp* and email exchanges to give ratings at their own pace, place, and time (with a maximum of three weeks). They received written rating protocols and explanations and were unaware of a pre- and post-design. They were paid an incentive.

We employed nine-point rating scales on *accentedness* and *comprehensibility*, as proposed by Munro and Derwing (1995; 2020): accentedness refers to *linguistic nativelikeness* or *sounding native-like* (Saito, Trofimovich, and Isaacs 2016), while comprehensibility means *perceived ease of understanding* (Munro and Derwing 1995). In this study, we 'reversed' the rating scales following Evers and Chen's (2020) study; higher rating scores mean a degree to which speech samples indicate less local/foreign accent and are more native-like (accentedness), and are easier to understand (comprehensibility). We also asked the raters about this reversed version; they all said to prefer the reversed one.

Among 128 participants, 11 participants (10 from the ILI groups and one from the NOVO) joined the pre- and post-test but did not participate in the experiment at all. We obtained this information from the participants' log files and excluded them from our final analysis.

3.2.3. I Can Speak *program*

We conducted two ASR-based experiments under the *I Can Speak* program, which lasted about five weeks. During the first four weeks, the participants were instructed to use either *I Love Indonesia* (ILI) or *NovoLearning* (https://www.novo-learning.com/) (NOVO) from home due to the Covid-19 pandemic. In the final week, the participants were invited to practice speaking through the websites at the school's computer laboratory following health protocols.

The seven units used in the experiment were adapted from the speaking materials provided freely by the British Council. We used the video materials as Lesson 1 in each of the units. In Lesson 1, the participants were instructed to watch the videos that contained specific topics (see Table 2). We also created two lessons, Lesson 2 and 3, that used ASR technology. In Lesson 2, the participants were asked to pronounce the target words and phrases derived from the videos in Lesson 1. In Lesson 3, the participants were required to respond to utterances using available prompts (phrases/sentences) within specific contexts referring to the videos.

Table 2. Participants' activities during the experiment.

Week	Activity	Practice Mode
1	• As the facilitator, the first author shared tutorials through students' *WhatsApp* groups. The tutorials explained how to sign in and use ILI or NOVO. • Students performed Unit 1 (topic: *Meeting people*).	From home
2	Students performed Unit 2 (*Homework problems*) and Unit 3 (*Not feeling well*).	From home
3	Students performed Unit 4 (*At the Shop*) and Unit 5 (*Buying new shoes*).	From home
4	Students performed Unit 6 (*Making plans*) and Unit 7 (*Talking about your family*).	From home
5	Students came to school twice in this final week – each lasting for about an hour – to complete the units.	At school

We selected the speaking activities for *Beginner A1* (Units 1, 2, 3, and 4) and *Elementary A2* (Units 5, 6, and 7). The rationale behind this selection was that most participants were considered beginners (in the A1/A2 range). This was confirmed by the results of the Education First English proficiency test ($M_{ILI} = 44.15$, $SD_{ILI} = 16.512$; $M_{NOVO} = 38.79$, $SD_{NOVO} = 16.445$; 95% CI: ILI [39.95, 48.34], NOVO [34.75, 42.83]).

3.2.3.1. ASR-based language learning systems: ILI and NOVO. In this study, ILI was upgraded with more personalized feedback. To help build a tailored automatic speech recognition system on ILI, *ResponsiveVoice* (https://responsivevoice.org/) and *RecordRTC* (https://recordrtc.org/) were employed. *ResponsiveVoice* was utilized on ILI to convert speech to text. *RecordRTC* is an open-source WebRTC JavaScript audio and video recording library. This software enabled recording and replaying the users' utterances on ILI.

Percentage scores (ranging from 0% to 100%) were displayed on ILI immediately after the utterances had been spoken. The participants could practice pronunciation multiple times, but only their four latest scores would be displayed (see Figure 1). The scores were calculated as a percentage of orthographic (characters) similarity between the participants' utterances (recognized by the ASR-based system) and their reference orthographic transcriptions (see Figure 2).

The scores above 80% received 'You got it!' feedback (indicated in a green bar), while for those below 80%, the learners were given feedback such as 'You could use some practice' (indicated in a red bar) (see Figure 1 and 2). This feedback was inspired by and adapted from *Speechace* (https://www.speechace.com/).

ILI also provides two special features: *I Can Speak Chart (ICSC)* and *Top 20 Students* (T20S) (see Figures 3 and 4). ICSC shows users their learning activity through a simple chart (in their accounts). T20S displays a list of 20 users with their highest accumulated scores.

The second system we used, NOVO (see Bashori et al. (2020; 2021) for more information), provides detailed feedback on phonetics that identifies which phonemes were pronounced correctly and even provides phoneme-level pronunciation practice.

Three colors represent how close a user's 'spoken' phoneme is to the reference phoneme; green (*excellent*), yellow (*almost*), and pink (*we heard / … /*) for a deviant phone (see Figure 5). General feedback is also given, such as 'Good job!', 'Yes, that's it!', 'Right on!' or 'That doesn't seem right' (see Figure 6). Detailed written descriptions on how to correctly pronounce specific phonemes are also provided, and users can listen to target pronunciation audio samples, record and replay their voices, and are updated on their learning progress.

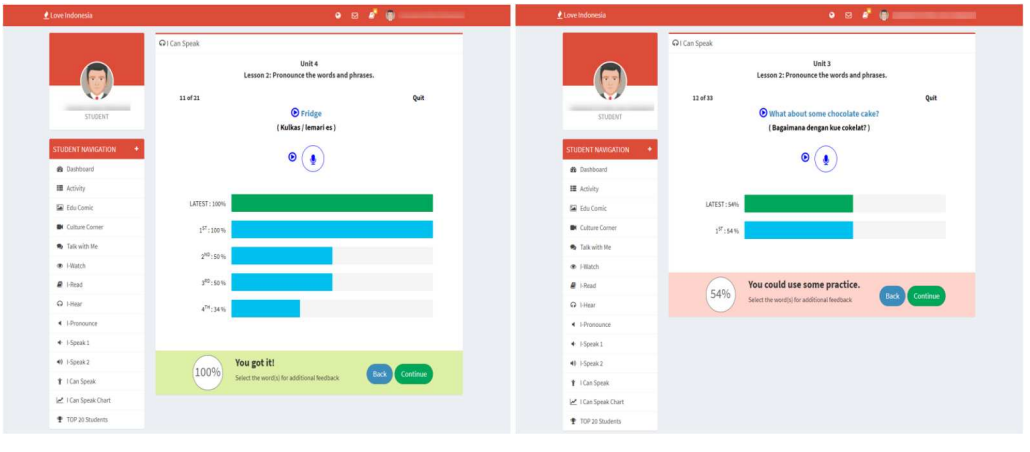

Figure 1. Pronunciation feedback on ILI.

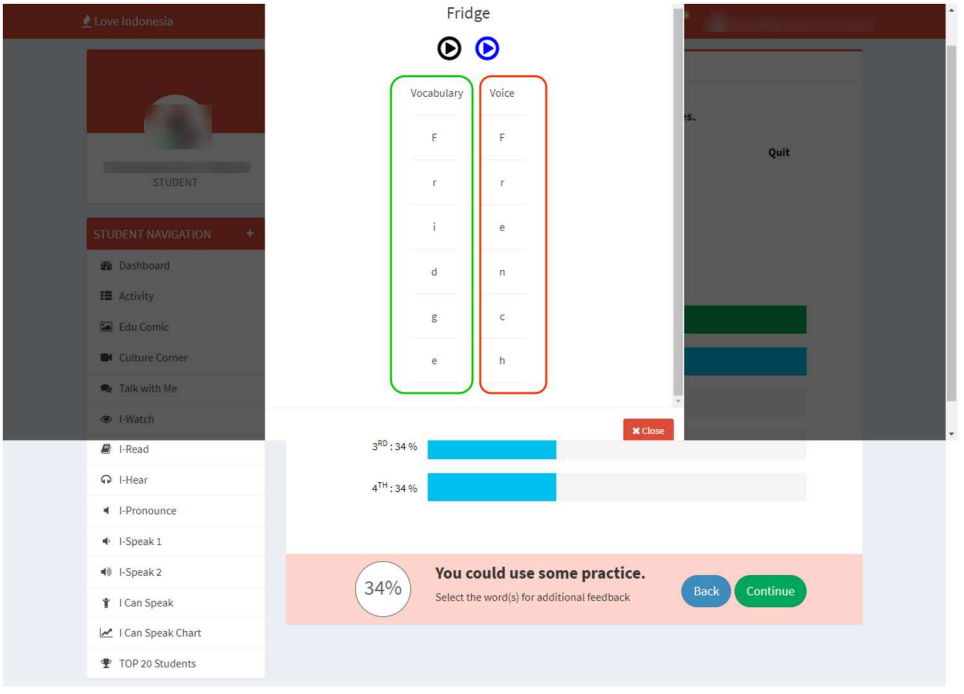

Figure 2. The orthographic similarity between a user's utterance recognized by the ASR system (right, in red square) and its reference orthographic transcription. (left, in the green square) as feedback.

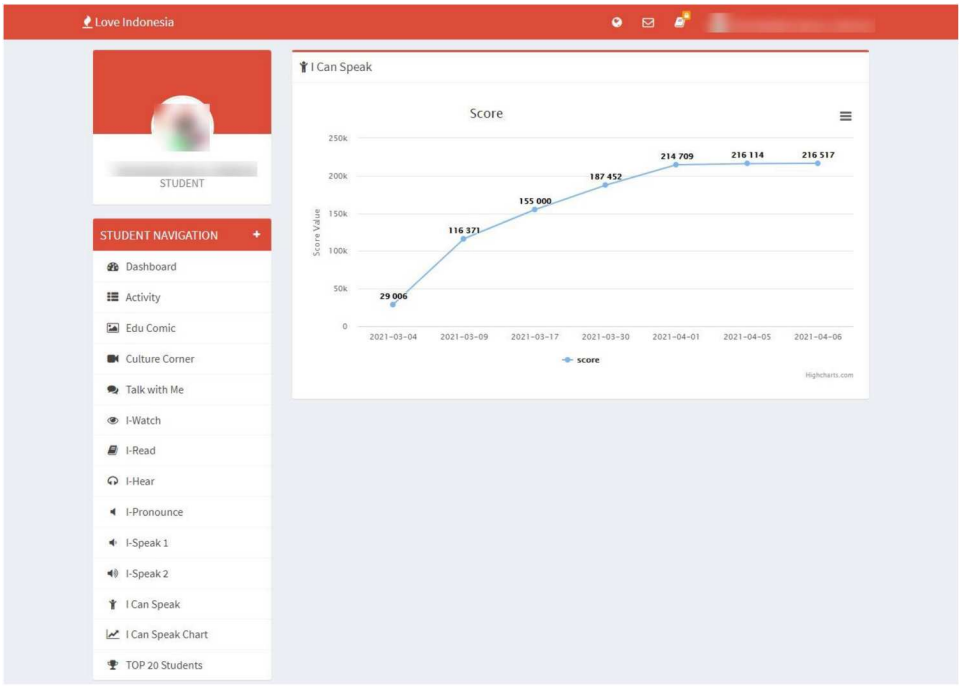

Figure 3. The feature *I Can Speak Chart*.

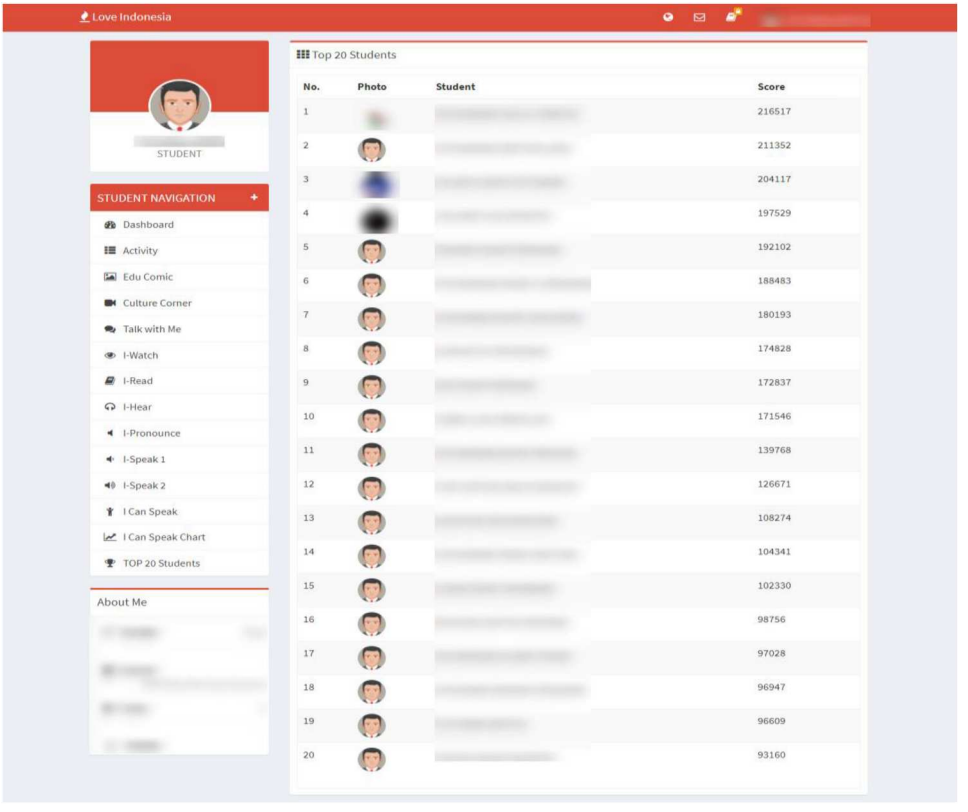

Figure 4. Feature Top 20 Students.

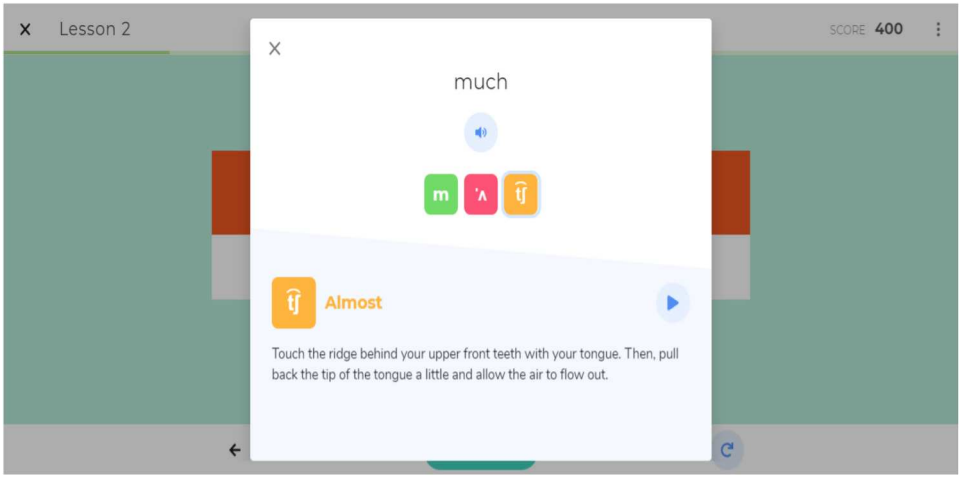

Figure 5. Detailed pronunciation feedback (phonetics; phoneme level) in NOVO.

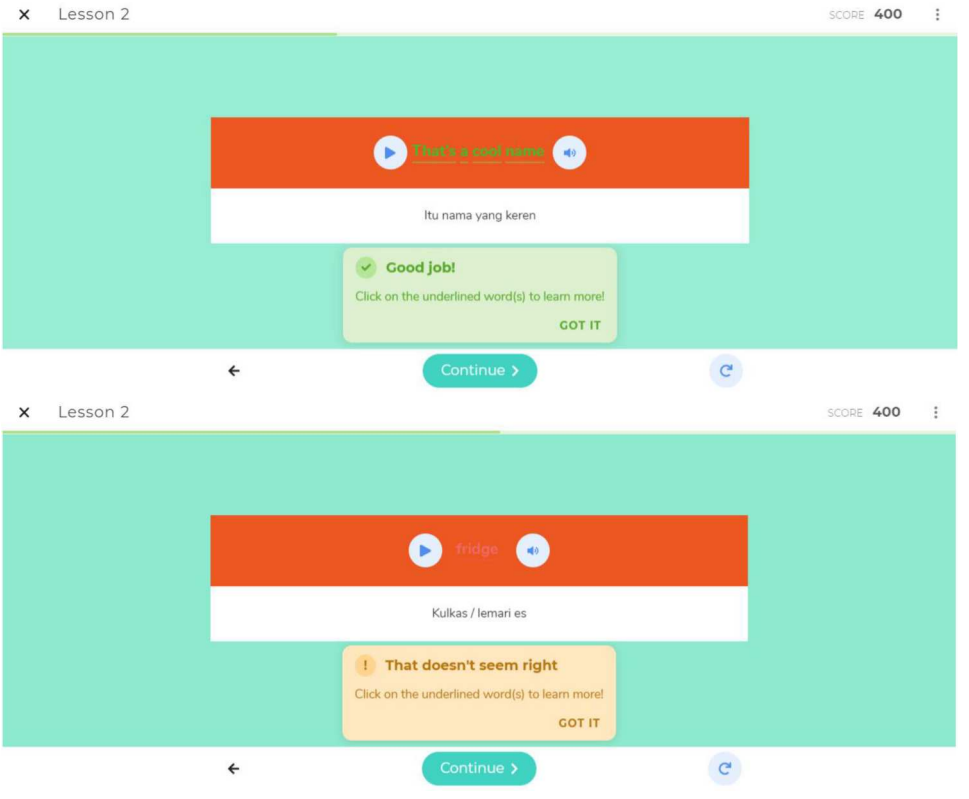

Figure 6. Pronunciation practice in NOVO.

4. Results

4.1. Word-level pronunciation

To investigate possible improvements in word-level pronunciations, a repeated measures ANOVA was employed with time (pre- and post-test) as a within-subject factor and treatment (ILI and NOVO) as a between-subjects factor.

The phonetic distance scores differed significantly between the two-time points ($F(1,115) = 82.405$, $p < .001$, $\eta_p^2 = .417$). There was a main effect of treatment ($F(1,115) = 9.325$, $p = .003$, $\eta_p^2 = .075$). Finally, the interaction between time and treatment was significant ($F(1,115) = 4.33$, $p = .040$, $\eta_p^2 = .036$), indicating a difference in progress between the treatments. The average phonetic distance gain score for ILI was .0145, and for NOVO, it was .0232. The difference was significant ($t'(101.4) = -2.048$, $p = .043$). To ascertain that the higher initial distance scores in the NOVO group were not the source of a larger decrease, we analyzed with treatment as a fixed factor and the pre-test as a covariate. The difference between ILI and NOVO was still significant ($F(1,114) = 4.752$, $p < .05$).

As smaller distance scores indicate more progress, Figure 7 visualizes the students' word-level pronunciation improvement, with NOVO leading to an overall better improvement.

Aggregated results over the 28 target words show that the students improved their pronunciation of most of the target words. Based on the alignment distance (mean) gain scores, only five words showed no improvement: *black* (mean gain score = 0), *size* (.06), *bottle* (0), *cool* (0), and *enough* (0).

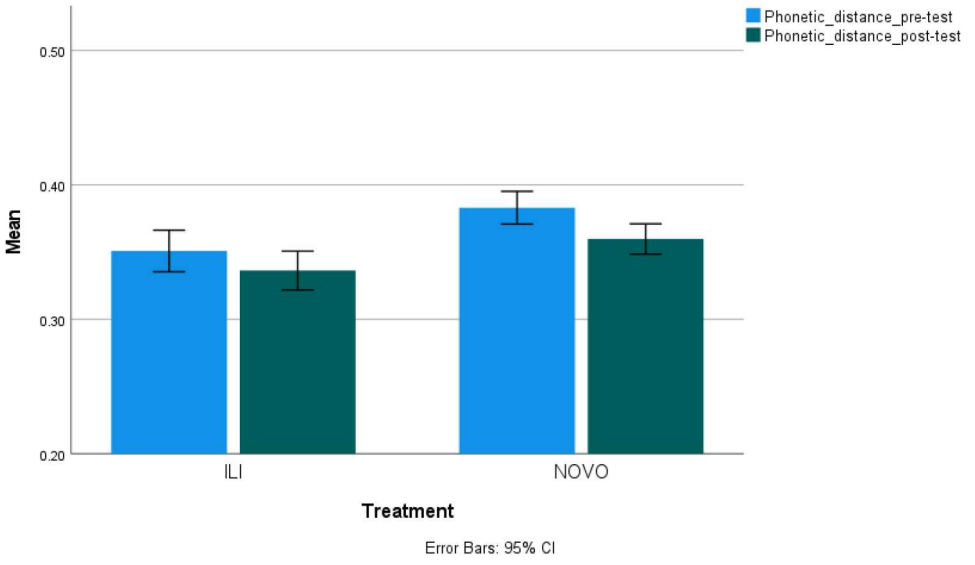

Figure 7. Mean scores and confidence intervals of the overall phonetic distance scores in the pre-and post-test in the two treatment groups.

4.2. Sentence-level pronunciation

Nine raters rated the 28 sentences spoken by 117 students in the pre-and post-test on accentedness and comprehensibility. We first investigated the reliability of these ratings using the intraclass correlation coefficient (ICC) and found high reliability on accentedness (pre = .863; post = .860) and comprehensibility (pre = .934; post = .939), which shows strong agreement among the raters. We inspected whether individual raters were deviant in their ratings. This was not the case, despite their English proficiency levels (varying between B2 and C1).

We then examined the reliability of the 28 sentence items using the ICC. The items were highly reliable (accentedness, pre = .989, post = .987; comprehensibility, pre = .980, post = .976). There were no deviant sentences. All the sentences had high item-total correlations, and their differences were limited.

The correlations between accentedness and comprehensibility were .889 in the pre-test and .913 in the post-test. At the pre-test, the averages of comprehensibility and accentedness were 4.97 and 3.40, respectively, which is a significant difference ($t(127) = 53.433$, $p < .001$). At the post-test, their averages were 5.60 and 4.17, respectively, again a significant difference ($t(127) = 59.674$) (see Figures 8 and 9).

Both ILI and NOVO lead to pronunciation improvement. To investigate the differences between the sentence ratings in the pre-and post-test on accentedness and comprehensibility, a repeated measures ANOVA was employed with time (pre- vs. post-test) and treatment (ILI vs. NOVO) as explanatory variables.

The accentedness scores differed significantly between the two-time points ($F(1,115) = 1332.32$, $p < .001$, $\eta_p^2 = .921$). There was a main effect of treatment ($F(1,115) = 17.58$, $p < .001$, $\eta_p^2 = .133$). Finally, the interaction between time and treatment was significant ($F(1,115) = 32.48$, $p < .001$, $\eta_p^2 = .220$), indicating a difference in the degree of progress between the treatments. We investigated the difference between ILI and NOVO progress by computing and analyzing the gain scores, the post-test score minus the pre-test score. The average gain score for ILI was .650, and for NOVO, it was .891. The difference was significant (t' $(114.98) = -5.851$, $p < .001$). To ascertain that the lower initial scores on accentedness in the NOVO group were not the source of a larger increase, we analyzed

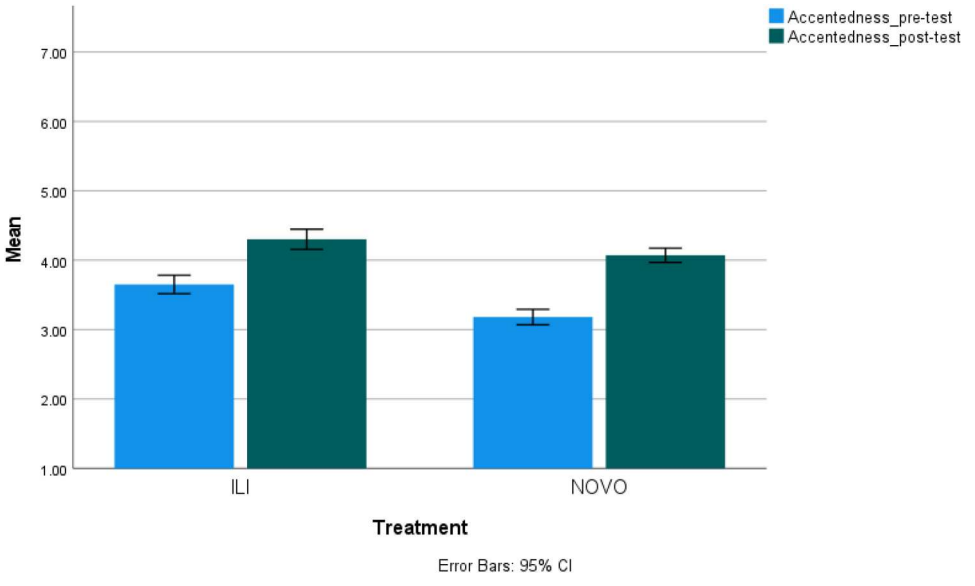

Figure 8. Mean scores and confidence intervals of the average sentence ratings on accentedness in the pre-and post-test in the two treatment groups.

with treatment as a fixed factor and the pre-test as a covariate. The difference between ILI and NOVO was still significant (F(1,114) = 16,127, $p < .001$).

The comprehensibility scores differed significantly between the two-time points ($F(1,115) = 470.96$, $p < .001$, $\eta_p^2 = .804$). There was a main effect of treatment ($F(1,115) = 14.95$, $p < .001$, $\eta_p^2 = .115$). Finally, the interaction between time and treatment was significant ($F(1,115) = 15.38$, $p < .001$, $\eta_p^2 = .118$), indicating a difference in progress between the treatments. The average comprehensibility gain score for ILI was .524, and for NOVO, it was .755. The difference was significant (t' (110.82) = −4.103, $p < .001$). To

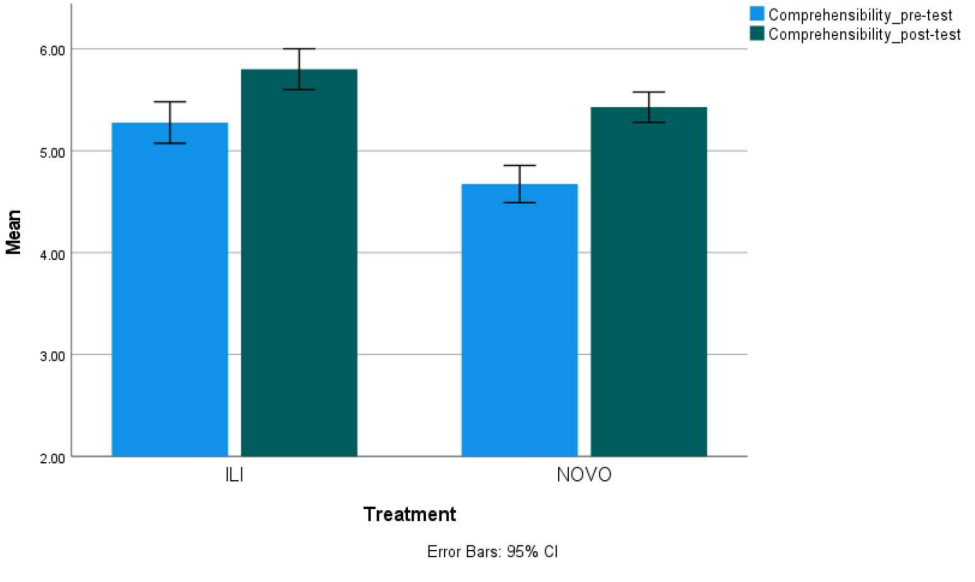

Figure 9. Mean scores and confidence intervals of the average sentence ratings on comprehensibility in the pre-and post-test in the two treatment groups.

ascertain that the lower initial scores on comprehensibility in the NOVO group were not the source of a larger gain, we analyzed with treatment as a fixed factor and the pre-test as a covariate. The difference between ILI and NOVO was still significant ($F(1,114) = 3,946$, $p < .05$).

4.3. The correlations between the pronunciation measures

The correlations between the pre-and post-test scores were high for all three tests (phonetic distance $r = .907$; accentedness $r = .864$; comprehensibility $r = .906$). All correlations between the measures were high as well, the absolute value of the correlations always being larger than .750. The correlations between accentedness and comprehensibility scores were consistently higher than .860, indicating two strongly related, but distinct dimensions.

4.4. Results of raters' written comments and expert's overview on phonetic transcriptions

We employed a thematic analysis (cf. Braun and Clarke 2012) on the written comments we obtained from the nine raters. We identified five central themes from the raters' written comments on sentence ratings: local (non-native) accent, segmental and suprasegmental errors, rating challenges, and language learning recommendations for learners. We translated (and adjusted) the excerpts into English (see Table C1 in Appendix C).

We also asked for a brief overview from the expert who helped making the 7,168 phonetic transcriptions on three main points: most frequent pronunciation errors, transcription challenges, and suggestions for speakers (see Table C2 in Appendix C for the overview).

5. Discussion

5.1. First RQ: To what extent do ILI and NOVO improve students' word-level and sentence-level pronunciation?

We found that ILI and NOVO successfully helped enhance the students' pronunciation of 28 words and 28 phrases/sentences. This adds a significant contribution to what ILI and NOVO can offer, after previous research had shown that these programs can help learners reduce their speaking anxiety and increase their language enjoyment and vocabulary (Bashori et al. 2021).

A few target words did not show any improvement at all, maybe because these words are (more or less) familiar to the students. For example, *black* is an A1-level word that seems easy and familiar to the students, but most of them mispronounced it as /blek/ instead of /blæk/. The phoneme /æ/ does not exist in the Indonesian sound system (Swan and Smith 2001), and replacing /æ/ with /e/ in this context cannot cause misunderstandings. We included several familiar words as target words in our pronunciation test and training to help reduce students' speaking anxiety, as students would feel too overwhelmed with too many unfamiliar words to learn (Horwitz, Horwitz, and Cope 1986).

The results of the sentence ratings indicate that overall, the students significantly improved their sentence-level pronunciation. We did not collect judgments of native speakers, but Crowther, Trofimovich, and Isaacs (2016) found that non-native and native speakers showed no differences in ratings. Moreover, the raters were reliable. The speaking program helped the students reduce their foreign accents and increase their comprehensibility.

The positive effects of learning through ASR-based language learning systems on students' pronunciation might be due to (1) the longer *time on task* and (2) the personalization aspect, as demonstrated by, among others, individualized feedback. The former has been considered critical in learning (Anderson 1981), and the latter complements the 'smart' integration between language learning and technology, which profoundly affects students' linguistic performance in a positive way (Colpaert 2018a; 2018b).

5.2. Second RQ: Are there differences in pronunciation improvement between ILI and NOVO?

There were significant differences between ILI and NOVO in helping improve the student's pronunciation. With its phonetic corrective feedback, NOVO outperformed ILI with its global corrective feedback in terms of lowering phonetic distances (for word-level pronunciation), reducing accentedness and enhancing comprehensibility (for sentence-level pronunciation).

Schmidt's (1990) 'noticing hypothesis' underlines the importance of learners becoming aware of discrepancies between their speech production and the L2 because this is necessary to acquire specific L2 features. Corrective Feedback (CF) can help draw the learners' attention to their specific problems and help them improve (Havranek 2002). Although various studies on feedback forms have produced mixed results (Norris and Ortega 2000), there are indications that explicit CF is more effective than implicit, potentially ambiguous CF (Lyster 1998), that CF does not work when it is erratic, and inconsistent (Chaudron 1988), that CF should be intensive (Han 2002), and that it should be immediate, when the procedural 'knowledge' that led to the error is still active in memory (DeKeyser 2007).

This all suggests that the phonetic feedback provided by NOVO was more effective because it was more explicit and thus more conducive to improvement than the global feedback provided by ILI. The NOVO feedback stimulated awareness of discrepancies in line with Schmidt's (1990) 'noticing hypothesis'. Through the detailed information provided by NOVO, students became aware of what they needed to improve in their pronunciation and could practice in pronouncing the speech sounds they found problematic. The positive impact of phonetic feedback has also been found in studies by Castelo (2022) and Tejedor-García et al. (2020). Although the treatment effect indicated a significant difference, ILI, with its global, corrective feedback, remained effective for pronunciation learning. This finding echoes previous studies such as those by Neri et al. (2008), Evers and Chen (2020), and Mroz (2018b).

Another explanation might be that NOVO could be accessed through laptop/PC and Android-based applications (smartphones). In contrast, ILI could only be accessed through a laptop/PC, which might have led NOVO users to practice more.

5.3. Third RQ: How and to what extent are our three pronunciation measures (phonetic distance, accentedness,d comprehensibility) related?

All correlations between the three measures were high. The correlations between accentedness and comprehensibility scores were consistently higher than .860, indicating two strongly related, but distinct dimensions. We also found that better accentedness scores accurately predict comprehensibility levels. Interestingly, we discovered that the raters systematically gave lower accentedness than comprehensibility scores. This is congruent with findings in other studies (Munro and Derwing 2020; Crowther, Trofimovich, and Isaacs 2016). Munro and Derwing (2020) argued that raters tend to focus on even the smallest deviances in judging accentedness.

As our study discusses two widely employed and essential constructs in the field of pronunciation, accentedness, and comprehensibility, it is crucial to obtain information on how these two constructs 'relate' to each other, such as whether raters are 'stricter' when it comes to giving ratings on accentedness or use other criteria compared to giving ratings on comprehensibility. Knowing whether (heavily) accented speech can still be comprehensible and whether reduced accentedness may yield and relate to better comprehensibility is helpful for researchers and educators/teachers. These concepts are central in all discussions on the consequences of L2 accents.

5.4. Fourth RQ: What do the experts say to observe and experience while giving ratings or making phonetic transcriptions?

The expert who made the phonetic transcriptions found that most students seemed to start learning English from reading instead of listening. As a result, they tended to apply spelling pronunciation (Swan and Smith 2001), which can lead to 'negative transfer' or 'interference' (e.g. Bardovi-Harlig and Sprouse 2018). This can be seen from the pronunciation of the word *enough*. The phoneme /f/ in the last syllable was mainly pronounced as /g/. Even after the treatment, no improvement was found for this word. Another reason is that most of the students were beginners, so perhaps they needed more practice to learn this difficult target word.

In our study, the raters noted the students' local (Javanese) accent called *medhok*. Studies by Purwaningsih and Nurdiawati (2020) and Khusniyah, Ismiatun, and Sholihah (2021) indicate that this *medhok* accent 'negatively' influences the English pronunciation of Indonesian learners. Interestingly, Zentz (2015) reported that one male learner explained that *medhok*-ness helped him sound 'masculine' and feel more secure when speaking English.

Our raters also noticed some of the problematic aspects of an accent, which were also observed in previous research: (1) monotone and inappropriate primary-stress placement (Hahn 2004), (2) segmental errors with a high functional load (Munro and Derwing 2006), and (3) speech rate (too slow or too fast) (Munro and Derwing 2001). Although recently there has been growing support for a paradigm shift in pronunciation (from native-likeness to comprehensibility/intelligibility), it is, however, still necessary to identify the above-mentioned aspects of an accent that may have a deleterious effect on comprehensibility/intelligibility (Munro and Derwing 2020).

5.5. Limitations and future implications

One of the limitations of this study is that we employed read speech for pronunciation testing, which is less representative of authentic communication than extemporaneous and spontaneous speech (Munro and Derwing 2020). Various speaking contexts and research designs are indeed necessary to help shed more light on pronunciation learning and assessment.

Interestingly, despite the relatively short period of oral practice, the students successfully improved their pronunciation. This aligns with, for instance, McCrocklin's (2016) experimental study that even a short duration of ASR-based training can positively impact learners' pronunciation. The relatively short amount of time also seems to imply that studying from home is less extensive (but still effective) than studying in the classroom. We hope that if students can practice longer with ASR-based language learning systems, they might achieve high pronunciation performance. Additionally, perhaps a combination of the systems' feedback (phonetic and global) can be taken into consideration by educators and researchers to yield better results.

6. Conclusions and future perspectives

This study investigated to what extent two ASR-based language learning systems (ILI and NOVO) can support learners in improving their pronunciation. ILI and NOVO turned out to have a positive impact on learners' accurate word-level and sentence-level pronunciation, as indicated by smaller phonetic distances (in reading words), reduced accentedness, and increased comprehensibility (in reading sentences). With its phonetic corrective feedback, NOVO appeared to have a stronger learning effect than ILI's global corrective feedback.

Future studies should investigate the effectiveness of various forms of ASR-based feedback (phonetic versus global or a combination) compared with human feedback (peer-based and teacher-based) in improving learners' pronunciation. Additionally, examining to what extent ASR-based practice can impact learners' long-term pronunciation performance would be relevant.

Acknowledgment

The project *I Love Indonesia* was funded by *Lembaga Pengelola Dana Pendidikan* (LPDP), the Indonesia Endowment Fund for Education from the Ministry of Finance, Indonesia. We would like to thank the students and teachers who participated in this research for their cooperation and *NovoLearning* for their valuable support. We also wish to acknowledge the assistance of John Huisman for data analysis support in computing the phonetic distances.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Lembaga Pengelola Dana Pendidikan [grant number:].

Notes on contributors

Dr. Muzakki Bashori obtained his Ph.D. from Radboud University Nijmegen. He currently works at the Faculty of Social and Political Sciences, Semarang State University, Indonesia. His research interests include automatic speech recognition in EFL learning, foreign language speaking anxiety, and local folklore in ELT. Email address: muzakkibashori90@gmail.com or muzakkibashori@mail.unnes.ac.id.

Prof. Roeland van Hout is an emeritus professor in applied linguistics and variationist linguistics at the Centre for Language Studies of the Radboud University Nijmegen. He publishes in sociolinguistics, dialectology, and second language acquisition and is interested in research methodology and statistics. Email address: roeland.vanhout@ru.nl.

Dr. Helmer Strik is an associate professor at the Centre for Language Studies of the Radboud University Nijmegen. His fields of expertise include computer-assisted language learning (CALL), phonetics, speech production, speech processing, automatic speech recognition (ASR), spoken dialogue systems, e-learning, and e-health. Email address: helmer.strik@ru.nl.

Dr. Catia Cucchiarini is a senior researcher at the Centre for Language Studies of the Radboud University Nijmegen. Her research activities address speech processing, computer assisted language learning (CALL), and the application of automatic speech recognition (ASR) to language learning and testing. Email address: catia.cucchiarini@ru.nl.

ORCID

Muzakki Bashori http://orcid.org/0000-0002-8899-6791
Roeland van Hout http://orcid.org/0000-0002-8870-1631
Helmer Strik http://orcid.org/0000-0003-1722-3465
Catia Cucchiarini http://orcid.org/0000-0001-5908-0824

References

Anderson, L. W. 1981. "Instruction and Time-on- Task: A Review." *Journal of Curriculum Studies* 13 (4): 289–303. https://doi.org/10.1080/0022027810130402.

Ariatna. 2016. "The Need for Maintaining CLT in Indonesia." *TESOL Journal* 7 (4): 800–822. https://doi.org/10.1002/tesj.246.

Baker, A. A., and M. Burri. 2016. "Feedback on Second Language Pronunciation: A Case Study of EAP Teachers' Beliefs and Practices." *Australian Journal of Teacher Education* 41 (6): 1–19. https://doi.org/10.14221/ajte.2016v41n6.1.

Bardovi-Harlig, K., and R. A. Sprouse. 2018. "Negative Versus Positive Transfer." The TESOL Encyclopedia of English Language Teaching, 1–6.

Bashori, M., R. van Hout, H. Strik, and C. Cucchiarini. 2020. "Web-based Language Learning and Speaking Anxiety." *Computer Assisted Language Learning*, 1–32.

Bashori, M., R. van Hout, H. Strik, and C. Cucchiarini. 2021. "Effects of ASR-Based Websites on EFL Learners' Vocabulary, Speaking Anxiety, and Language Enjoyment." *System* 99: 102496. https://doi.org/10.1016/j.system.2021.102496.

Botes, E., J. M. Dewaele, and S. Greiff. 2020. "The Foreign Language Classroom Anxiety Scale and Academic Achievement: An Overview of the Prevailing Literature and a Meta-Analysis." *Journal for the Psychology of Language Learning* 2: 26–56. https://doi.org/10.52598/jpll/2/1/3.

Braun, V., and V. Clarke. 2012. "Thematic Analysis." In *APA Handbook of Research Methods in Psychology*, edited by H. M. Cooper, P. M. Camic, D. L. Long, and A. T. Panter, 57–71. Washington, DC: APA.

Breitkreutz, J., T. M. Derwing, and M. J. Rossiter. 2001. "Pronunciation Teaching Practices in Canada." *TESL Canada Journal* 19 (1): 51–61. https://doi.org/10.18806/tesl.v19i1.919.

Burns, A. 2012. "A Holistic Approach to Teaching Speaking in the Language Classroom." Symposium, 165–178.

Castelo, A. 2022. "FunEasyLearn: An App for Learning Pronunciation?" In *Perspectives and Trends in Education and Technology*, 395–405. Singapore: Springer.

Chaudron, C. 1988. *Second Language Classrooms*. New York: Cambridge University Press.

Chen, H. H. J. 2011. "Developing and Evaluating an Oral Skills Training Website Supported by Automatic Speech Recognition Technology." *ReCALL* 23 (1): 59–78. https://doi.org/10.1017/S0958344010000285.

Chiu, T. L., H. C. Liou, and Y. Yeh. 2007. "A Study of Web-Based Oral Activities Enhanced by Automatic Speech Recognition for EFL College Learning." *Computer Assisted Language Learning* 20 (3): 209–233. https://doi.org/10.1080/09588220701489374.

Colpaert, J. 2018a. "Exploration of Affordances of Open Data for Language Learning and Teaching." *Journal of Technology and Chinese Language Teaching* 9 (1): 1–14.

Colpaert, J. 2018b. "Transdisciplinarity Revisited." *Computer Assisted Language Learning* 31: 483–489. https://doi.org/10.1080/09588221.2018.1437111.

Couper, G. 2017. "Teacher Cognition of Pronunciation Teaching: Teachers' Concerns and Issues." *TESOL Quarterly* 51 (4): 820–843. https://doi.org/10.1002/tesq.354.

Crowther, D., P. Trofimovich, and T. Isaacs. 2016. "Linguistic Dimensions of Second Language Accent and Comprehensibility." *Journal of Second Language Pronunciation* 2 (2): 160–182. https://doi.org/10.1075/jslp.2.2.02cro.

Cucchiarini, C., A. Neri, and H. Strik. 2009. "Oral Proficiency Training in Dutch L2: The Contribution of ASR-Based Corrective Feedback." *Speech Communication* 51 (10): 853–863. https://doi.org/10.1016/j.specom.2009.03.003.

Cucchiarini, C., and H. Strik. 2017. "Automatic Speech Recognition for Second Language Pronunciation Training." In *The Routledge Handbook of Contemporary English Pronunciation*, 556–569. Routledge.

Dai, Y., and Z. Wu. 2021. "Mobile-assisted Pronunciation Learning with Feedback from Peers and/or Automatic Speech Recognition: A Mixed-Methods Study." *Computer Assisted Language Learning*, 1–24.

Daniels, P., and K. Iwago. 2017. "The Suitability of Cloud-Based Speech Recognition Engines for Language Learning." *The JALT CALL Journal* 13 (3): 211–221. https://doi.org/10.29140/jaltcall.v13n3.220.

DeKeyser, R. 2007. *Practice in a Second Language, Chapter Introduction: Situating the Concept of Practice*. New York: Cambridge University Press.

Education First. 2022. *The Report of English Proficiency Index* 2022. Retrieved from https://www.ef.com/wwen/epi/.

Ehsani, F., and E. Knodt. 1998. "Speech Technology in Computer-Aided Language Learning: Strengths and Limitations of a new CALL Paradigm." *Language Learning & Technology* 2: 45–60.

Elimat, A. K., and A. F. AbuSeileek. 2014. "Automatic Speech Recognition Technology as an Effective Means for Teaching Pronunciation." *The JALT CALL Journal* 10 (1): 21–47. https://doi.org/10.29140/jaltcall.v10n1.166.

Eskenazi, M. 1999. "Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some Issues and a Prototype." *Language Learning & Technology* 2: 62–76.

Evers, K., and S. Chen. 2020. "Effects of an Automatic Speech Recognition System with Peer Feedback on Pronunciation Instruction for Adults." *Computer Assisted Language Learning*, 1–21.

Foote, J. A., A. K. Holtby, and T. M. Derwing. 2011. "Survey of the Teaching of Pronunciation in Adult ESL Programs in Canada, 2010." *TESL Canada Journal*, 1–22.

Golonka, E. M., A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik. 2014. "Technologies for Foreign Language Learning: A Review of Technology Types and Their Effectiveness." *Computer Assisted Language Learning* 27 (1): 70–105. https://doi.org/10.1080/09588221.2012.700315.

Hahn, L. D. 2004. "Primary Stress and Intelligibility: Research to Motivate the Teaching of Suprasegmentals." *TESOL Quarterly* 38 (2): 201–223. https://doi.org/10.2307/3588378.

Han, Z. 2002. "A Study of the Impact of Recasts on Tense Consistency in L2 Output." *TESOL Quarterly* 36: 542–572.

Havranek, G. 2002. "When is Corrective Feedback Most Likely to Succeed?" *International Journal of Educational Research* 37: 255–270. https://doi.org/10.1016/S0883-0355(03)00004-1.

Hincks, R. 2003. "Speech Technologies for Pronunciation Feedback and Evaluation." *ReCALL* 15 (1): 3–20. https://doi.org/10.1017/S0958344003000211.

Horwitz, E. K., M. B. Horwitz, and J. Cope. 1986. "Foreign Language Classroom Anxiety." *The Modern Language Journal* 70 (2): 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x.

Iwashita, N. 2003. "Negative Feedback and Positive Evidence in Task-Based Interaction: Differential Effects on L2 Development." *Studies in Second Language Acquisition* 25 (1): 1–36. https://doi.org/10.1017/S0272263103000019.

Jenkins, J. 2007. *English as a Lingua Franca: Attitude and Identity*. Oxford: Oxford University Press.

Khusniyah, F. Ismiatun, and F. A. Sholihah. 2021. "Javanese Interference on Students' English Speaking Skill in the EFL Context." *Jurnal Penelitian, Pendidikan, dan Pembelajaran* 16 (18): 1–14.

LaRocca, S. A., J. J. Morgan, and S. M. Bellinger. 1999. "On the Path to 2X Learning: Exploring the Possibilities of Advanced Speech Recognition." *CALICO Journal* 16: 295–310. https://doi.org/10.1558/cj.v16i3.295-310.

Levis, J., and R. Suvorov. 2013. "Automatic Speech Recognition." In *The Encyclopedia of Applied Linguistics*, edited by C. Chapelle, 1–8. Hoboken, NJ: Blackwell Publishing.

Liakin, D., W. Cardoso, and N. Liakina. 2014. "Learning L2 Pronunciation with a Mobile Speech Recognizer: French /y/." *CALICO Journal* 32 (1): 1–25. https://doi.org/10.1558/cj.v32i1.25962.

Lyster, R. 1998. "Negotiation of Form, Recasts, and Explicit Correction in Relation to Error Types and Learner Repair in Immersion Classrooms." *Language Learning* 48: 183–218. https://doi.org/10.1111/1467-9922.00039.

McCrocklin, S. M. 2016. "Pronunciation Learner Autonomy: The Potential of Automatic Speech Recognition." *System* 57: 25–42. https://doi.org/10.1016/j.system.2015.12.013.

Mroz, A. P. 2018a. "Noticing Gaps in Intelligibility Through Automatic Speech Recognition (ASR): Impact on Accuracy and Proficiency." Paper presented at 2018 Computer-Assisted Language Instruction Consortium (CALICO) Conference, Urbana, IL, United States.

Mroz, A. 2018b. "Seeing how People Hear you: French Learners Experiencing Intelligibility Through Automatic Speech Recognition." *Foreign Language Annals* 51 (3): 617–637. https://doi.org/10.1111/flan.12348.

Munro, M. J., and T. M. Derwing. 1995. "Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners." *Language Learning* 45 (1): 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x.

Munro, M. J., and T. M. Derwing. 2001. "Modelling Perceptions of the Comprehensibility and Accentedness of L2 Speech: The Role of Speaking Rate." *Studies in Second Language Acquisition* 23 (4): 451–468. https://doi.org/10.1017/S0272263101004016.

Munro, M. J., and T. M. Derwing. 2006. "The Functional Load Principle in ESL Pronunciation Instruction: An Exploratory Study." *System* 34 (4): 520–531. https://doi.org/10.1016/j.system.2006.09.004.

Munro, M. J., and T. M. Derwing. 2020. "25 Years of Intelligibility, Comprehensibility and Accentedness." *Journal of Second Language Pronunciation* 6 (3): 283–309. https://doi.org/10.1075/jslp.20038.mun.

Nakazawa, K. 2012. "The Effectiveness of Focused Attention on Pronunciation and Intonation Training in Tertiary Japanese Language Education on Learners' Confidence: Preliminary Report on Training Workshops and a Supplementary Computer Program." *The International Journal of Learning* 18 (4): 181–192.

Neri, A., O. Mich, M. Gerosa, and D. Giuliani. 2008. "The Effectiveness of Computer Assisted Pronunciation Training for Foreign Language Learning by Children." *Computer Assisted Language Learning* 21 (5): 393–408. https://doi.org/10.1080/09588220802447651.

Norris, J. M., and L. Ortega. 2000. "Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis." *Language Learning* 50: 417–528. https://doi.org/10.1111/0023-8333.00136.

NovoLearning. 2019. *Improving English Language Proficiency using Novo's Mobile Learning Solution: A Pilot Project.* Retrieved from https://www.novo-learning.com/assets/pdf/research-report.pdf.

Perwitasari, A. 2013. "Slips of the Ears: Study on Vowel Perception in Indonesian Learners of English." *Humaniora* 25 (1): 103–110.

Purwaningsih, R., and D. Nurdiawati. 2020. "The Influence of Javanese Accent Toward the Students' English Consonant Pronunciation at English Education Study Program of Universitas Peradaban." *Jurnal Dialektika Program Studi Pendidikan Bahasa Inggris* 8 (1): 55–68.

Saito, K. 2011. "Examining the Role of Explicit Phonetic Instruction in Native-Like and Comprehensible Pronunciation Development: An Instructed SLA Approach to L2 Phonology." *Language Awareness* 20 (1): 45–59. https://doi.org/10.1080/09658416.2010.540326.

Saito, K., P. Trofimovich, and T. Isaacs. 2016. "Second Language Speech Production: Investigating Linguistic Correlates of Comprehensibility and Accentedness for Learners at Different Ability Levels." *Applied Psycholinguistics* 37 (2): 217–240. https://doi.org/10.1017/S0142716414000502.

Schmidt, R. 1990. "The Role of Consciousness in Second Language Learning1." *Applied Linguistics* 11: 129–158. https://doi.org/10.1093/applin/11.2.129.

Swan, M., and B. Smith. 2001. *Learner English. A Teacher's Guide to Interference and Other Problems*. 2nd ed. Cambridge, UK: Cambridge University Press.

Teimouri, Y., J. Goetze, and L. Plonsky. 2019. "Second Language Anxiety and Achievement: A Meta-Analysis." *Studies in Second Language Acquisition* 41 (2): 363–387. https://doi.org/10.1017/S0272263118000311.

Tejedor-García, C., D. Escudero-Mancebo, E. Cámara-Arenas, C. González-Ferreras, and V. Cardeñoso-Payo. 2020. "Assessing Pronunciation Improvement in Students of English Using a Controlled Computer-Assisted Pronunciation Tool." *IEEE Transactions on Learning Technologies* 13 (2): 269–282. https://doi.org/10.1109/TLT.2020.2980261.

Van Zanten, E., and V. J. Van Heuven. 2004. "Word Stress in Indonesian: Fixed or Free." *NUSA Linguistic Studies of Indonesian and Other Languages in Indonesia* 53: 1–20.

Zentz, L. 2015. "'Is English Also the Place Where I Belong?': Linguistic Biographies and Expanding Communicative Repertoires in Central Java." *International Journal of Multilingualism* 12 (1): 68–92. https://doi.org/10.1080/14790718.2014.943233.

Appendix A

Below are the target words and phrases/sentences used in the pronunciation pre- and post-test.

Table A1. Target words and phrases/sentences.

No	Unit	Word	Phrase/Sentence
1	1	Birthday	My birthday is in May.
2	1	Cool	That's a cool name.
3	1	Year	What year are you in?
4	1	Lost	You look lost.
5	2	Examination	Revise everything for the examination.
6	2	Repeat	Can you repeat that?
7	2	Mathematics	I have not done my mathematics homework.
8	2	Write	Write an email to your friend.
9	3	Headache	I have got a headache.
10	3	Wrong	What's wrong?
11	3	Cake	What about some chocolate cake?
12	3	Cheese	There is cheese and tomato.
13	4	Fridge	Over there in the fridge.
14	4	Change	There is one pound sixty change.
15	4	Magazine	How much is this magazine?
16	4	Bottle	Do you have a bottle of water?
17	5	Size	Do you have size eleven?
18	5	Color	What color would you like?
19	5	Black	But not in black.
20	5	Trainer	I really like the trainers.
21	6	Celebration	We are going to have a celebration.
22	6	Idea	That's a great idea.
23	6	Listen	Listen, it's his birthday next week.
24	6	Cinema	How about going to the cinema?
25	7	Twice	I have been to visit him twice
26	7	Enough	I think it is enough.
27	7	Originally	She is from Mexico originally.
28	7	Version	Is that the new version?

Appendix B

Table B1. The raters' demographic information.

Rater's Code	TOEFL ITP / IELTS score	Current occupation	Education level	Teaching years	Teaching speaking experience
M1	563 / 6.5	Freelance English tutor; Master's student	Bachelor	0.5 (6 months)	Yes Target: elementary and junior high school students
M2	573	Master's student	Bachelor	0	No
M3	527	English teacher	Master	14	Yes Target: senior and vocational high school students
F1	527	English teacher; translator	Master	6	Yes Target: cadets and senior high school students
F2	533	English teacher	Bachelor	0.17 (2 months)	Yes Senior high school students
F3	540	Website translator	Master	0	No
F4	540	English tutor	Master	6	Yes Target: elementary, junior and senior high school students
F6	6.5	English teacher	Master	7	Yes Target: junior high school students
F7	540	Doctoral student; English teacher	Master	10	Yes Target: senior high school students

Appendix C

Table C1. Major themes, subthemes, and excerpts selected from the raters' written comments on sentence ratings.

Themes	Subthemes	Examples
Local accent	*Medhok* or Javanese accent	'In my opinion, the accent is more Javanese. Rarely (are they) not *medhok*. Almost all (of them are) *medhok*. (F2)
Segmental errors	Consonant	'The errors that occur are mostly in the pronunciation of vocabulary that has more than one consonant letter.' (M2)
	Vocal	'..Can also be in (the pronunciation of) vocabulary that has more than one vocal letter.' (M2)
	Spelling-related mispronunciation	'..it seems that they don't really know how to pronounce each word. Some even actually pronounce it according to their writing.' (F2)
Suprasegmental errors	Intonation and stress	'There is no stress, and flat intonation in every sentence.' (F6) 'Most speakers use monotonous intonation. This intonation also makes their local accent very thick/clear. (F1)
Rating challenges	Audio clarity	'Some of the students' voices are not clearly heard. Some are too fast, so it takes more effort to assess them.' (M3)
	Unfinished words	'There are also (some target) sentences that are not pronounced until finished (by the students).' (F4)
	High workload	'Because perhaps a lot of recordings are listened to, sometimes it makes (me) unfocused and there is a bias when rating.' (F6)
Recommendation for learners	Exposure	'Increase the level of exposure with English.' (F7)
	Practice	'..often listen to the teacher when pronouncing and practice (yourself) at home.' (F3)
	Native speaker-oriented	'Look back at how each key word is pronounced in standard American or British English.' (F4)

Table C2. A brief overview from the expert who helped make the 7,168 phonetic transcriptions.

Aspect	Explanation
Most frequent pronunciation errors	The majority (if not all) of the speakers seem to start learning English from reading instead of listening so that they pronounced the words the way they would if the texts were written in Indonesian. For example, they pronounced the word "version" with an "f" instead of a "v". This case was very much prevalent that an Indonesian listener would directly be able to tell that these speakers were Indonesian just from the way they pronounced English words the Indonesian way. Some speakers did not take time to read the words in their head first before pronouncing but jumped right into reading them out loud instead. This resulted in some major mispronunciation and they ended up pronouncing completely different words.
Transcription challenges	Making sure that I transcribe exactly how they pronounced it because some audio were in low quality and when the speakers were in doubt they would sometimes skip certain syllables. There is this one "t" sound that is not available in English, the kind that Javanese people pronounce when they say "*bathuk* (forehead)", there was this one speaker who used this "t" to pronounce the word "bottle". I was struggling with what to use to represent this sound and settled with keeping the "t" because it was the closest sound representation I could find.
Suggestions for speakers	Watch movies with English subtitle so that you get accustomed to seeing the pronunciation of English words and notice how they are usually written differently than they sound. Listen to English songs with lyrics, repeat some of your favorite lines, compare it with the clip, make sure you pronounce the words correctly. Practice listening with transcript text. Read a book that has an audio-book companion.