[Software Review]

Pronunciation Coach 3D

Simon Moxon (simon@sitamoxon.com)
Walailak University, Nakhon Si Thammarat, Thailand

Application	Pronunciation Coach 3D
Publisher	icSpeech (A division of Rose Medical Solutions Ltd)
Product Type	Microsoft Windows Application Software
Language(s)	English (US and UK)
Level	Any
Media Format	EXE
Operating Systems	Microsoft Windows 64-bit 7, 8 and 10 (requires US English language pack)
Hardware Requirements	x64 architecture, Minimum 4 GB of RAM, 126 MB of free hard disk space, Speakers, Microphone (headset recommended)
Supplementary Software	None
Price	Free (7-day Trial Version) $70.00 (Home User) $350 (Commercial License) * * Quoted amount included 30% Discount for ten or more licenses, 50% discount for one hundred or more licenses Prices are a one-off payment for a lifelong license and include free upgrades and technical support

Abstract

Pronunciation is a defining characteristic of speech, yet, for EFL students, it is often afforded the least time in the classroom. In large class sizes, teachers can rarely dedicate sufficient time to evaluate every student's pronunciation ability in detail. Consistency of evaluation and quality of informative feedback is also compromised in such teaching and learning environments. Factors such as these have created the need for more autonomous forms of mastering pronunciation. Software solutions can help alleviate the teacher's workload while furnishing students with an interactive means of independent learning that provides accurate and reliable feedback. This article reviews a Microsoft Windows application called Pronunciation Coach 3D. Although intended for use by professionals such as speech therapists, Pronunciation Coach 3D offers excellent potential as an autonomous learning resource for practicing English pronunciation. While the application

has its strengths, this article highlights several weaknesses that should be considered before applying the application to the teaching and learning environment.

Keywords: pronunciation coach; speech evaluation; ASR, software, EFL

Introduction

Pronunciation is a fundamental attribute of second language (L2) proficiency and is often assessed as a critical descriptor of speaking ability. Listeners often interpret errors in pronunciation as dysfluency (Derwing & Munro, 2005). However, mispronunciation has been attributed to mother tongue influence (Jahandar et al., 2012; Lai et al., 2009), which becomes problematic when L2 sounds are absent or pronounced differently in the mother tongue dialect (Ambalegin & Suryani, 2018). Providing students with corrective feedback at the time of error is crucial for effective correction (Huang & Jia, 2016). Yet, this skill is often allocated the least time in the L2 classroom (Derwing et al., 2012; Derwing & Munro, 2005). Technological solutions, such as Automatic Speech Recognition (ASR), can offer advantages over or in support of human evaluation, namely in the form of instant feedback (Haggag, 2018) and visual representation of sounds, which have been attributed to improved intonation (Hincks, 2003; Portmann & Leemann, 2018). ASR can also facilitate pronunciation development more than human teachers (Golonka et al., 2014). According to Pennington & Rogerson-Revell (2019), pronunciation technology increases student exposure to the target language while maximizing practice opportunities through autonomous learning and automated feedback. However, the authors state that existing solutions have yet to use accents and 3D modeling to their full potential. In terms of correcting mispronunciation, automated feedback from ASR systems is only beneficial if it is reliable and accurate (Pennington & Rogerson-Revell, 2019) and can be correctly interpreted and utilized (Moxon, 2021). Improving students' knowledge and awareness of English pronunciation has been shown to improve their listening comprehension skills (Dao et al., 2020).

The focus of this article is to describe and evaluate the primary features of Pronunciation Coach 3D (PC3D). Although intended for medical purposes, this application provides instruction and evaluation on the pronunciation of English language sounds, making it a valuable tool to facilitate teaching and learning in the L2 classroom. The application combines state-of-the-art 3D pronunciation modeling, with TTS, ASR, and pronunciation evaluation.

Description

Developed by icSpeech, PC3D is an English language pronunciation application capable of demonstrating and evaluating the pronunciation of sounds in the English language. Although intended for professional use, the software has a user-friendly interface and offers intuitive and reliable features for practicing English pronunciation. First released in November 2020, at the time of writing, the application is on its third minor release, version 1.3.0.

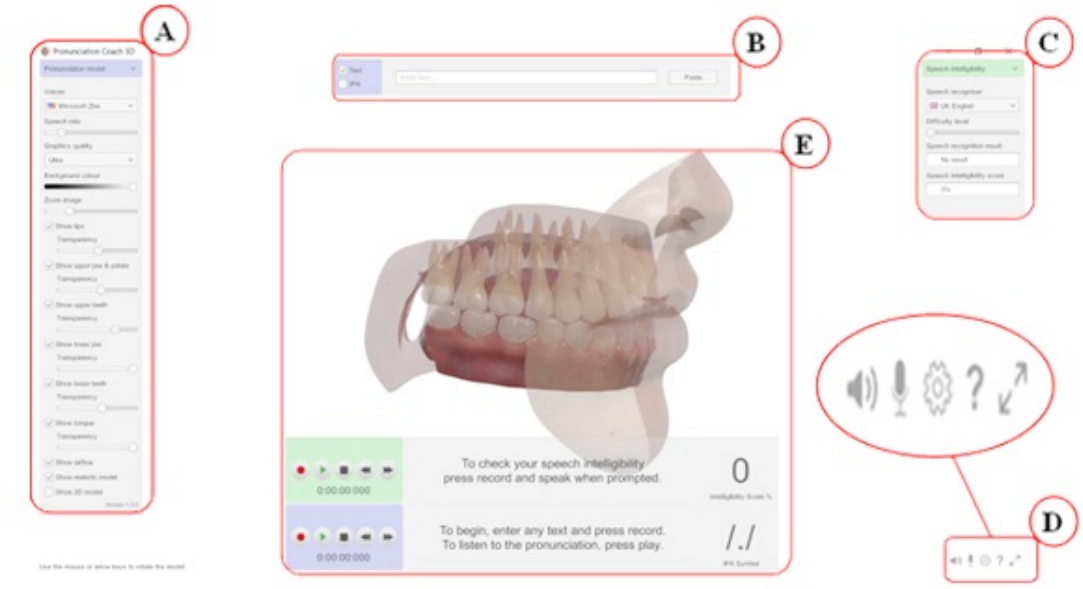
PC3D is an enhanced 3D version of its 2D predecessor and presents a 3D pronunciation model (PM) of the mouth, enabling one to focus on specific areas from any angle. Version 1.2.0 included the earlier 2D PM, allowing users to exchange between 2D and 3D renderings. The user interface comprises three principal areas: pronunciation tutorial, pronunciation evaluation, and audio configuration. The pronunciation tutorial features include Text-to-Speech (TTS), waveform images, and 3D computer modeling and animation to demonstrate the production of phonemes, words, and phrases. The pronunciation evaluation aspect of the software offers a voice recording interface and uses a speech waveform and ASR to provide instant visual and numerical feedback on pronunciation accuracy. Finally, the audio configuration enables one to optimize voice capture hardware for accurate pronunciation evaluation.

Overview

PC3D comprises a single window divided into five areas, as shown in Figure 1. The five areas consist of:

A. Pronunciation model configuration
B. TTS and International Phonetic Alphabet (IPA) interface
C. Pronunciation accuracy evaluation configuration
D. Hardware configuration and help options
E. Pronunciation model interface (2D and 3D versions)

Figure 1
Main User Interface

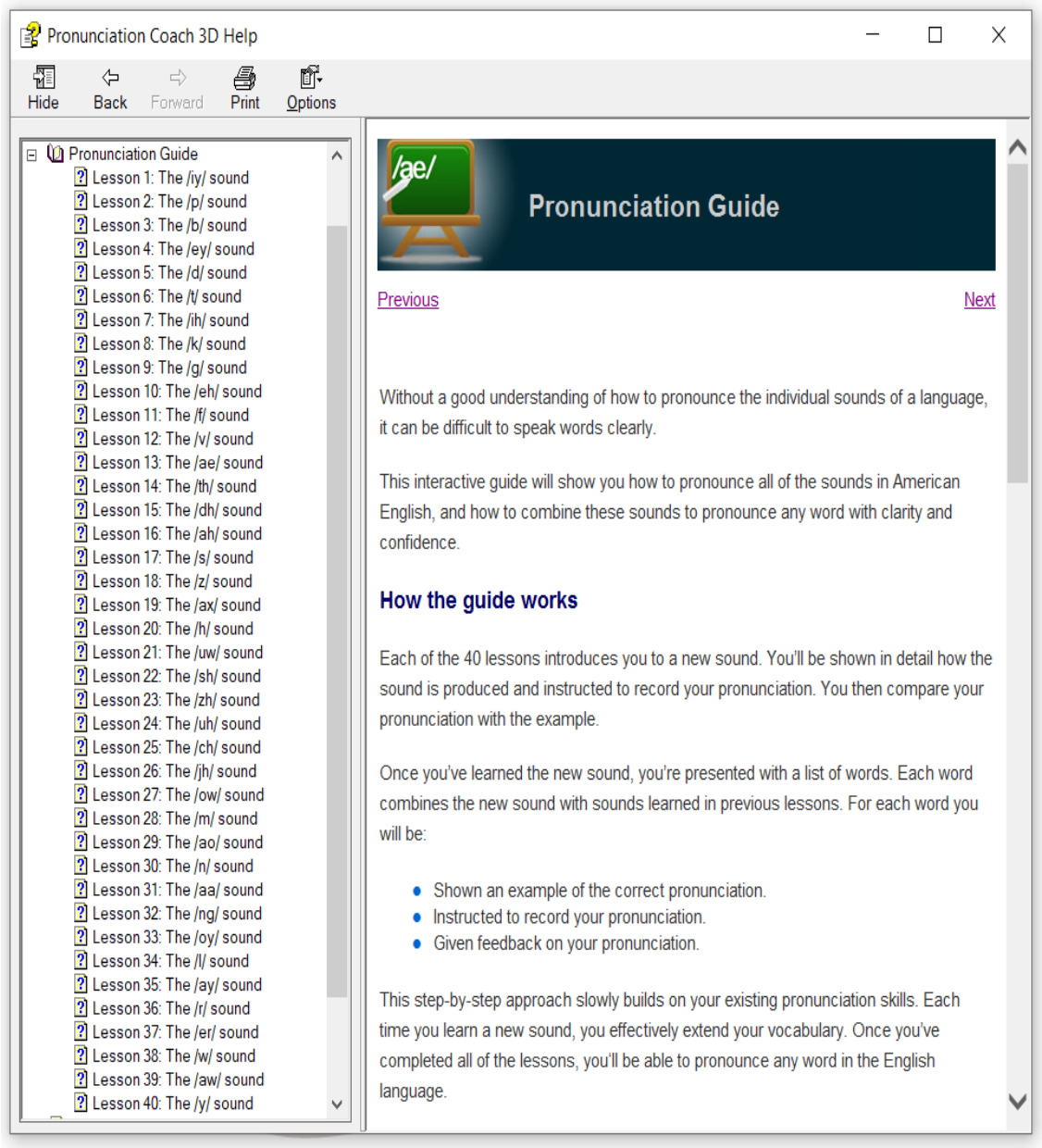


Each waveform is color coded to match its associated configuration panel, blue for the PM and green for ASR pronunciation accuracy.

PC3D includes a 40-lesson pronunciation guide, as shown in Figure 2, which offers an informative breakdown of the production of each sound in the IPA library. The end of each lesson provides self-assessment exercises and a list of example words. The

guide is included as part of the “Help” interface, which is accessed from the hardware configuration and help area of the screen (see Figure 1, item D).

Figure 2
Pronunciation Guide



Pronunciation Models

PC3D enables one to create 2D and 3D models for any phoneme, word, or phrase. The required model can be created by typing the phoneme, word, or phrase into the TTS interface or clicking the required IPA symbol (see Figures 3 and 4).

Figure 3
Text To Speech (TTS) Interface

Figure 4
International Phonetic Alphabet (IPA) interface

The configuration options for the PM allow for control over the audio properties, namely the TTS voice provided by Windows (see Table 1) and the rate of speech. They also control the visibility and transparency of individual parts of the model, namely the teeth, lips, tongue, and jaws. Two additional options provide the ability to toggle between 2D and 3D animated renderings and control airflow visibility. Most visual configuration settings apply only to the 3D PM (see Figure 5).

Table 1
TTS Voices Based on Microsoft Windows Version

Windows Version	Language	Gender	Name
7	US English	Female	Microsoft Anna
8	US English	Male	Microsoft David
8	US English	Female	Microsoft Sierra
10	US English	Male	Microsoft David
10	US English	Female	Microsoft Zira

Figure 5
Pronunciation Model Configuration Options

TTS voice.

TTS rate of speech.

Graphics quality: Adjustable to suit video performance of the user's device.

Background colour: Varying shades from white to black.

Zoom in or out.

Visibility and colour intensity of the lips.

Visibility and colour intensity of the upper jaw and palate.

Visibility and colour intensity of the upper teeth.

Visibility and colour intensity of the lower jaw.

Visibility and colour intensity of the lower teeth.

Visibility and colour intensity of the tongue.

Visibility of airflow (voiced, voiceless, nasal).
Realistic or artificial appearance.
Display 2D or 3D renderings.

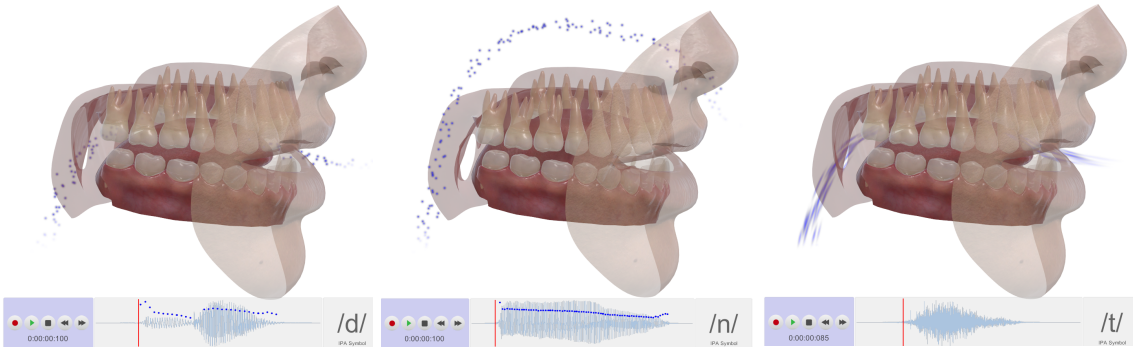
The PM interface can alternate between 2D and 3D renderings using the "Show 2D model" configuration setting (see Figures 6 and 7).

Figure 6
2D Pronunciation Model

Figure 7
3D Pronunciation Model

Using the 3D version, one can also differentiate between voiced, voiceless, and nasal phonetic sounds via a visual representation of the airflow associated with each. In both versions, voiced and voiceless sounds are distinguished separately in the speech waveform via frequency markers, which are displayed as blue dots along the speech waveform. In the 3D version, voiced and nasal sounds can be identified by observing a dotted airflow in the PM and frequency markers on the speech waveform. In contrast, voiceless sounds can be identified by observing flat lines for the airflow in the PM and the absence of frequency markers in the speech waveform (see Figure 8).

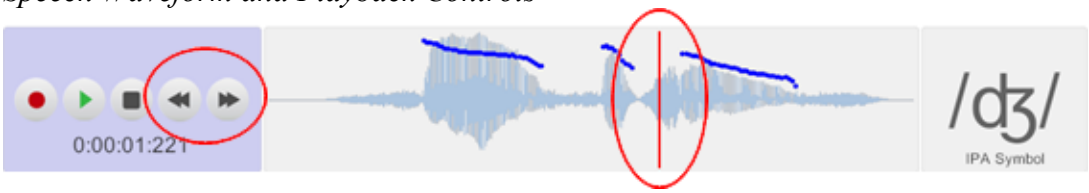
Figure 8
(Left to right) Voiced, Nasal, and Voiceless Phonetic Models



For added clarity, one can use the mouse and mouse wheel to dynamically zoom in and rotate the 3D PM through 360º, making it possible to observe the tongue, teeth, palate, jaw and lip positions from any viewing position. The keyboard's cursor keys can also be used to rotate the model. The key combination of SHIFT + Cursor Key will rotate the model through 90º in the relevant direction.

One can navigate through individual phonetic sounds on the speech waveform by using the step forward and step back buttons or using the mouse to position the red place marker at the required location, as shown in Figure 9.

Figure 9
Speech Waveform and Playback Controls



In cases where a letter is a consonant cluster sound of two or more individual phonetic sounds, such as the letter "X" (/ɛ/ /k/ /s/), the speech waveform controls enable one to step through the individual phonetical sounds and practice each in isolation. The same is also true when practicing whole words or phrases. In Figure 9, the speech waveform is illustrating the phonetic sounds of the word "sausages" (/s/ /ɔ/ /s/ /ɪ/ /dʒ/ /ɪ/ /z/).

Although seemingly undocumented in the user guide, additional features can be observed by pressing the function keys F1 to F4. F1 and F2 cause the model to revolve continuously, even during speech. The behavior is slightly different for F2 in that the model's facial features fade out, one by one, with each rotation. F3 cycles through the pronunciation of each IPA symbol, pronouncing one sound with each revolution of the model. F4 pronounces numbers, starting from one and incrementing with each model revolution. Manipulating the PM configuration options is possible without interrupting the animation of these features.

Pronunciation Evaluation

Entitled "Speech Intelligibility," the pronunciation evaluation features of PC3D are simplistic and intuitive in their design. The Speech Intelligibility (SI) panel, shown in Figure 10, consists of two configuration settings, speech recognition language and difficulty level, plus two textual evaluation feedback fields, ASR result and SI score.

Figure 10
Speech Intelligibility Panel

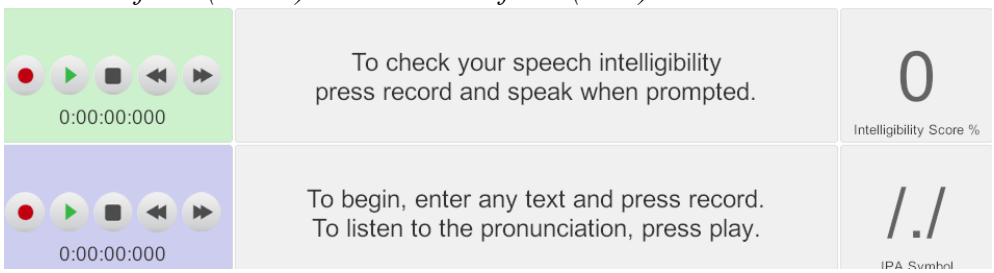
The speech recognizer supports American and British varieties of English, which require the installation of Windows language packs prior to use. Three difficulty levels are available, which can be adjusted using the slider control; the lowest difficulty level is

selected by default. While the evaluation criteria between the difficulty levels is not documented, the scores appear to correlate more with the clarity of one's pronunciation rather than the accurate reproduction of the target waveform in terms of speech rate and intonation. In the classroom, students found the transition from the lowest difficulty level a significant challenge. For example, students who scored 90% at the lowest difficulty level struggled to achieve 30% and 3% at the intermediate and highest levels, respectively. On completion of the recording of one's speech, a speech recognition result and intelligibility score are displayed. The recognition result displays a green tick and the recognized word if the ASR has recognized the pronunciation clearly enough for evaluation; otherwise, a red cross is displayed.

On opening the SI panel, an additional speech waveform panel, the ASR waveform (colored green), is displayed directly above the TTS waveform (colored blue), as seen in Figure 11. An ASR waveform and intelligibility score are generated for any speech submitted for evaluation.

Figure 11

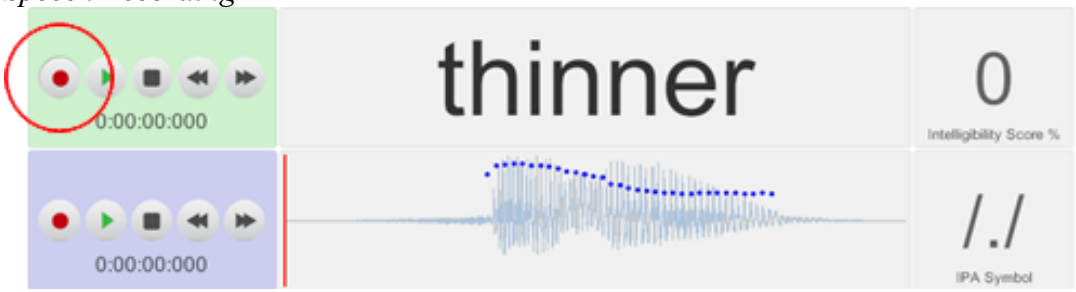
ASR Waveform (Green) and TTS Waveform (Blue)



Once the microphone has been correctly set up (see Configuration), and a PM created, a speech recording can be created by clicking the red record button. PC3D prompts for speech to begin by displaying the target sound or word in the ASR waveform panel (see Figure 12). It automatically stops recording when speech is detected to have ceased.

Figure 12

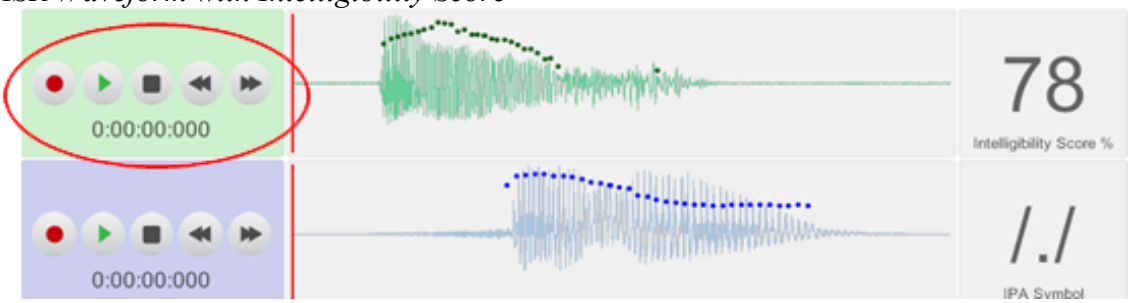
Speech Recording



Once the ASR has recognized the pronunciation, an intelligibility score is calculated and displayed. While this numeric feedback is helpful, one can achieve a more informative evaluation of pronunciation accuracy by comparing ASR and TTS waveforms shown in Figure 13. As with the TTS waveform, one can review each step of the submitted pronunciation using the playback and navigation buttons.

Figure 13

ASR Waveform with Intelligibility Score



Pronunciation intelligibility scores range from 0 to 100% and correlate to Unintelligible, Poor, Fair, Good, or Excellent pronunciation ratings, as shown in Table 2.

Table 2

Pronunciation Evaluation Scores

Score	Description
0%	Unintelligible
10% - 30%	Poor
30% - 50%	Fair
50% - 75%	Good
75% - 100%	Excellent

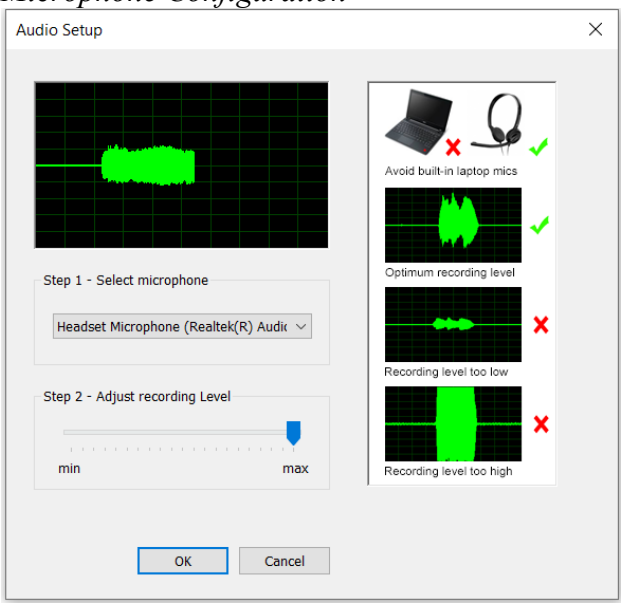
Configuration

The configuration interface is minimalistic and simplistic. There are three options for configuration: 1) Speaker configuration, which uses the Windows volume mixer dialogue. 2) Microphone configuration, consisting of device selection and volume level setting (see Figure 14). 3) Restore default settings, which reverts the PM settings to their default values.

Upon opening, the microphone configuration dialogue automatically commences voice capture and displays the volume level as a waveform. Correct volume adjustment at this stage is vital for accurate pronunciation evaluation. Recommended volume levels are shown on the right-hand side of the window, as shown in Figure 14.

Figure 14

Microphone Configuration



Evaluation

This in-depth review of PC3D originated from a need to find a platform capable of reliably evaluating pronunciation accuracy, identifying pronunciation errors, and providing corrective feedback for Thai undergraduate students studying in their first year of EFL (n = 107, age 18 years, 91 females, 16 males). Alternative platforms, such as Rosetta Stone, PRAAT, BoldVoice, Say It: English Pronunciation, and ELSA were considered before selecting PC3D. Table 3 lists the strengths and weaknesses for each of these applications. PC3D was selected because it offered an animated 3D model for demonstrating pronunciation, along with TTS, ASR, and pronunciation evaluation capabilities on a platform that would be best suited for classroom instruction. For educational purposes, a commercial license cost $350, however students who wished to use the software for personal use could purchase a home license for $70 or use the free 7-day trial license.

The review considered four evaluation benchmarks: user interface, tutorial and instructional content, consistency of pronunciation evaluation, and methods of corrective feedback.

Table 3

Comparison of Alternate Platforms

Name	Strengths	Weaknesses	Cost ($)
Rosetta Stone	• General tutorial content. • ASR. • Waveform. • Pronunciation evaluation via TruAccent® technology. • Progress recording.	• Content it broad, not focused on pronunciation. • No TTS. • Expensive for students. • Large program data storage requirements.	11.99/ month (1 language) 299 for life (25 languages)

	• Suitable for large screen demonstration. • Tutorial content. • Multi-platform compatible.	• No pronunciation models. • British and American English forms are separate subscriptions.	
PRAAT	• Suitable for large screen demonstration. • Powerful analysis features. • Waveforms. • Some ASR ability.	• Technically complicated. • No tutorial content. • No pronunciation models. • Not designed for teaching pronunciation, only audio analysis and manipulation.	Free
BoldVoice	• Pronunciation evaluation. • Suitable for mobile devices. • ASR. • Pronunciation evaluation. • Accent evaluation. • Tutorial content.	• Not suitable for large screen demonstration. • No pronunciation models. • No waveform. • No TTS. • American English only.	Free** 7.50/month
Say It English Pronunciation	• British and American English. • Suitable for mobile devices. • Waveform. • In-App features purchase separately. • Personal study list. • Pronunciation evaluation. • Score history. • Tutorial content.	• Requires iOS 9.0 or later. • Not suitable for large screen demonstration. • No pronunciation tuition. • No TTS. • No pronunciation models. • Self-assessed scoring.	Free** 2.99 - 106
ELSA	• Suitable for mobile devices. • Progress tracking. • Tutorial content. • Test and activities. • Pronunciation evaluation. • Listening and speaking content.	• Not suitable for large screen demonstration. • No waveform. • No TTS. • No pronunciation models. • American English only.	Free** 11.99/month 74.99/year
Pronunciation Coach 3D	• British and American English. • ASR. • TTS. • 3D pronunciation model. • Waveform. • Pronunciation evaluation. • Suitable for large screen demonstration.	• Windows 64bit only. • No score history. • Limited tuition. • Initial cost.	Free** 70 350 CUL

Note. Unless otherwise stated, prices are for non-commercial use only.
** Free for 7 days, CUL = Commercial User License.

User Interface

Compared to other ASR applications, such as PRAAT, which is considered too complicated and technical for students (Brett, 2004; Nushi & Sadeghi, 2021; Setter & Jenkins, 2005), and well-established, but broader focused applications, such as Rosetta Stone, PC3D affords a simplistic user interface that focuses entirely on pronunciation. The screen layout is clear and unimposing, making it inviting for the user. While observing user interaction, students and teachers appeared comfortable and relaxed.

However, a few areas of the screen design were noted as causing confusion or hindrance during use.

Firstly, the placement and choice of icons used for audio configuration and help (see Figure 1, item D). Here, a gear wheel, the icon used to represent the restoration of default settings, caused confusion due to its common use in other applications (Google Chrome, MS Windows) to mean setting rather than resetting. Moreover, although positioned next to the audio configuration, it does not affect audio settings but instead resets the PM's parameters. Arranging this icon with the other PM configuration settings would seem more intuitive to the user.

Secondly, from a visual perspective, the screen layout has been oversimplified in that unused features cannot be repositioned or hidden. For example, if zoomed to fit the extent of the screen, as may be necessary during classroom demonstrations, the TTS and waveform panels obscure the view of the 3D model. Related to this is the visibility of the TTS and ASR waveform panels. Once the PM and SI preferences are set, their configuration panels serve no other purpose. However, as their visibility dictates the visibility of their associated waveform, it is impossible to view the waveform panels if their associated configuration panel is minimized (hidden). A more intuitive interface might employ a simple menu bar that offers options to turn each panel on or off as required.

Finally, the text size in the application's main window does not align with the Windows display settings (text scaling). Although this may be insignificant on a personal use device, during testing on a projector screen in the classroom, despite increasing the text scaling in Windows, the text in PC3D remained too small to see clearly from six meters or more. The text in PC3D scales in proportion to the window dimensions, which may present a problem on smaller Windows devices.

Overall, the user interface appears simple enough for EFL students to operate after minimal tuition. However, there are notable limitations in the design that could be improved for use in an EFL learning environment.

Tutorial and Instructional Content

While the 2D and 3D PMs are beneficial for demonstrating sound production, PC3D offers no structured lesson content and assessment that the student can follow. Beyond the basic sounds provided in the help window, there is no learning content to stimulate continued use, such as consonant clusters, minimal pairs, vowel combinations, and short quiz exercises. As a tool for practicing pronunciation, the 3D PMs are excellent. However, without sufficient guided instructional content, autonomous learners may soon lose motivation to use the software despite its benefits.

Another notable limitation in this area is the range of TTS and ASR voices and varieties of English. Windows 10, for example, currently offers male and female voices for six varieties of English, including British, American, Australian, Irish, Canadian, and Indian. However, although the TTS voices are realistic, PC3D offers only one male and one female US English voice for TTS (see Table 1). Likewise, while the ASR offers American and UK varieties of English, other English varieties installed into Windows do not appear to be recognized. As English is now a globalized language, this is considered a disadvantage as the variety of English should match the needs and preferences of the

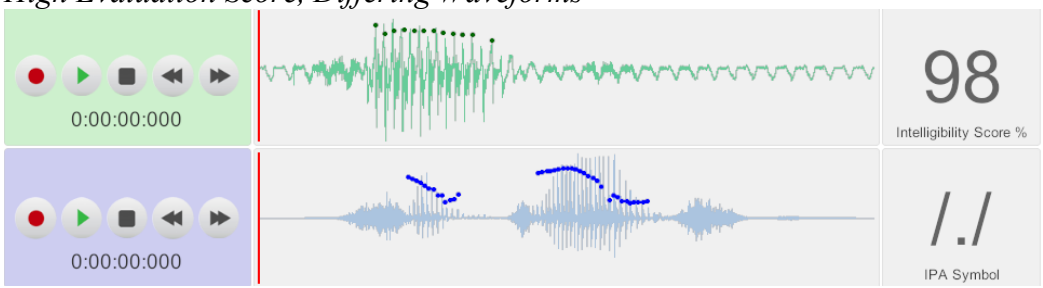
user. Practicing with American pronunciation TTS while evaluating against the UK English variety of ASR, for example, seems counterproductive.

Pronunciation Evaluation Consistency

Overall, PC3D appeared quite robust in this area. At the default "low" setting, ASR evaluation was consistent during repeated attempts by the same student. However, during software evaluation, difficulties in maintaining the correct microphone configuration led to irregularities in ASR accuracy. In some cases, a high accuracy score was achieved, yet the ASR and TTS waveforms bore little similarity, as shown in Figure 15. In contrast, instances were encountered when no evaluation score was displayed despite the presence of remarkably similar ASR and TTS waveforms and clear, accurate pronunciation. These irregularities were either confusing or demoralizing for the student.

Figure 15

High Evaluation Score, Differing Waveforms



In contrast, higher accuracy scores were sometimes obtained for inaccurate pronunciation when spoken at a low volume. Although this had the effect of boosting student confidence, it was seen as positive reinforcement of mispronunciation.

Regarding difficulty levels, while three levels are available, the pronunciation accuracy required at the medium and high levels seems disproportionately higher than at the low level. For progressive EFL students, the difficulty scale should cover a broader range to allow students to progress in smaller increments.

Overall, the pronunciation evaluation appeared reliable. However, during the software evaluation, the ASR evaluation scores seemed less consistent at the medium and high difficulty levels. Maintaining the correct microphone configuration was essential for obtaining reliable evaluation and could pose a problem during independent study.

Corrective Feedback

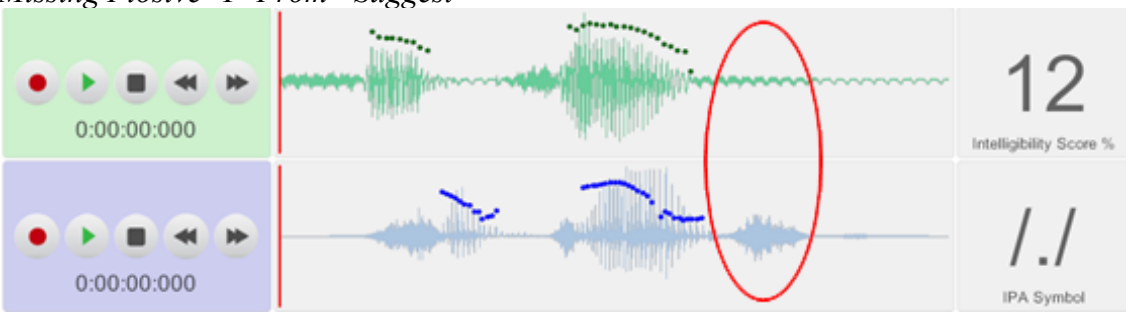
ASR pronunciation evaluation is a common feature in language learning applications. SpeechACE, for example, can evaluate the pronunciation of words and sentences to phoneme level. However, it provides only numerical feedback. Feedback void of visual and instructional information may be insufficient for autonomous learners to identify and correct mistakes (Moxon, 2021).

Although PC3D does not give evaluation scores at the phonetic level for words and phrases, it offers an ASR waveform for visual comparison against the TTS waveform. While lacking in corrective instruction, it allows the user to adopt a trial-and-error method of self-correction. Comparing the two waveforms, students can quickly identify and self-

correct pronunciation errors, such as inadequately sounding plosives or word endings, as the segment is visibly absent in their ASR waveform (see Figure 16).

Figure 16

Missing Plosive 'T' From "Suggest"



For the autonomous learner, visual feedback, such as waveforms, has been shown to aid the perfecting of pronunciation (Hincks, 2003; Olson, 2014). While PC3D performs well in this area, it offers no means for the user to compare one attempt with another, which can be problematic for the user if, despite multiple attempts, they cannot improve their accuracy score. It was felt that a recorded history of pronunciation attempts would be beneficial for students and teachers to monitor progress.

Conclusion

It should be emphasized that PC3D was not designed primarily as a learning tool. While its lack of tutorial content may be considered a disadvantage, as a tool to aid accurate pronunciation, it was felt that the advantages of PC3D outweigh the disadvantages. costcPC3D has much to offer to support EFL students in their English pronunciation, particularly in the areas of autonomous learning, instant feedback, and a 3D modeling interface that clearly illustrates how to produce English language sounds.

For students, the interface is simple and intuitive. The manipulation of the 3D model makes it easy for students to isolate different attributes of the mouth and visualization of the airflow, which gives them concise visual instruction on mouth positioning and a clear distinction between voiced and voiceless sounds.

The pronunciation evaluation functionality provides students with an autonomous means to practice and assess their pronunciation accuracy. While the limited intervals between difficulty levels create a challenge for students, the feedback is precise and reliable. The waveforms, score, and textual representation of the recognized words provide clear feedback for the student to identify problematic words and sounds. Some initial training is required for the students to understand how to compare the waveforms and identify voiced and voiceless sounds, but it was felt that this did not represent a technical challenge.

Overall, PC3D would be a valuable addition to the teacher's toolbelt. While the initial cost of commercial licenses may be high, it is a one-off payment that would prove cost-effective in the long run. In a multi-terminal classroom environment, a Learning Management System (LMS) based application may be a more appropriate solution as PC3D offers no means for the teacher to monitor user progress and performance remotely.

In contrast, as a platform for autonomous learning, on a standalone laptop or desktop PC application using a home user license, PC3D represents good value for money. However, where mobile devices are concerned, screen size, operating system compatibility, and the suitability of audio hardware on the device pose significant issues. Finally, PC3D is only available for MS Windows, which presents a barrier for students who may rely on Android and iOS devices.

References

Ambalegin, A., & Suryani, M. S. (2018). Mother tongue affecting the English vowel pronunciation of Batak Toba adults. *The 1st Annual International Conference on Language and Literature* (pp. 78-86). KnE Social Sciences. https://doi.org/10.18502/kss.v3i4.1920

Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL, 16*(1), 102-113. https://doi.org/10.1017/S0958344004000813

Dao, P., Nguyen, M. X., & Nguyen, N. B. (2021). Effect of pronunciation instruction on L2 learners' listening comprehension. *Journal of Second language Pronunciation, 7*(1), 10-37. https://doi.org/10.1075/jslp.19012.dao

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379-397. https://doi.org/10.2307/3588486

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal, 30*(1), 22-44. https://doi.org/10.18806/tesl.v30i1.1124

Fraser, H. (2000). *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*. DETYA (ANTA Innovative Project).

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (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

Haggag, H. M. (2018). Teaching phonetics using a mobile-based application in an EFL context. *European Scientific Journal, 14*(4), 189-204. https://doi.org/10.19044/esj.2018.v14n14p189

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL, 15*(1), 3-20. https://doi.org/10.1017/S0958344003000211

Huang, X., & Jia, X. (2016). Corrective feedback on pronunciation: Students' and teachers' perceptions. *International Journal of English Linguistics, 6*(6), 245-254. https://doi.org/10.5539/ijel.v6n6p245

Jahandar, S., Khodabandehlou, M., Seyedi, G., & Abadi, R. M. (2012). A gender-based approach to pronunciation accuracy of advanced EFL learners. *International Journal of Scientific & Engineering Research, 3*(6), 1-9. https://www.ijser.org/researchpaper/A-Gender-based-Approach-to-Pronunciation-Accuracy-of-Advanced-EFL-Learners.pdf

Lai, Y.-S., Tsai, H.-H., & Yu, P.-T. (2009). A multimedia English learning system using HMMs to improve phonemic awareness for English learning. *Journal of*

Educational Technology & Society, 12(3), 266-281. https://doi.org/10.1109/MUE.2007.30

Moxon, S. (2021). Exploring the Effects of Automated Pronunciation Evaluation on L2 Students in Thailand. *IAFOR Journal of Education: Language Learning in Education, 9*(3), 41-56. https://doi.org/10.22492/ije.9.3.03

Nushi, M., & Sadeghi, M. (2021). A Critical Review of ELSA: A Pronunciation App. *Computer Assisted Language Learning Electronic Journal, 22*(3), 287-302. http://callej.org/journal/22-3/Nushi-Sadeghi2021.pdf

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania, 97*(1), 47-68. https://doi.org/10.1353/hpn.2014.0030

Pennington, M. C., & Rogerson-Revell, P. (2019). Using technology for pronunciation teaching, learning, and assessment. In M. C. Pennington, & P. Rogerson-Revell (Eds.), *English Pronunciation Teaching and Research. Research and Practice in Applied Linguistics* (pp. 235-286). London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-47677-7_5

Portmann, L., & A. Leemann. (2018). Visual input trumps auditory input: Teaching English intonation. *Proceedings of the Conference on Phonetics & Phonology in German-speaking countries (P&P 13)*, 149–152. https://doi.org/10.18452/18805

Setter, J., & Jenkins, J. (2005). State-of-the-art review article: Pronunciation. *Language Teaching, 38*(1), 1-17. https://doi.org/10.1017/S026144480500251X