

## A System for Analyzing and Evaluating Computer-Assisted Second-Language Pronunciation-Teaching Websites and Mobile Apps

Lynn E. Henrichsen  
Linguistics Department, Brigham Young University, Provo, Utah  
United States of America  
Lynn\_Henrichsen@byu.edu

**Abstract:** The potential benefits of computer-assisted pronunciation teaching (CAPT) include a private, stress-free learning environment; virtually unlimited input; practice at the student's own pace; individualized, instantaneous feedback through Automatic Speech Recognition (ASR); and visual acoustic and articulatory displays. Regrettably, however, CAPT software does not always measure up to its potential. Furthermore, many L2 teachers and learners are not familiar with the full range of CAPT possibilities and may not be aware of what features to look for in an instructional product. This paper shares a remedy to this problem—a comprehensive set of criteria for analyzing and evaluating CAPT software, websites, and mobile apps. Utilizing an easy-to-use checklist format, as well as Likert-scale and open-response items, this system is designed to guide teachers and learners in evaluating CAPT programs.

### The Need to Evaluate CAPT Software

Today, increasing numbers of language learners and teachers are relying on websites and mobile apps to help them improve their second or foreign language skills (Rosell-Aguilar, 2017, p. 243). A large number of sites and apps have been developed with the goal of helping language learners (and their teachers) with various aspects of the target language—including pronunciation. Regrettably, not all these websites and apps are equally helpful or effective. For this reason, rather than being “seduced” by publishers’ marketing, language teachers and learners “need to analyze these programs from a critical perspective using pedagogically coherent and technically elaborated criteria” (Martins, Levis, & Borges, 2016, p. 142).

O’Brien and Levis warn: “Many of the commercially available products are often neither pedagogically sound nor informed by research” (2017, p. 1). Neri, Cucchiari, Strik, and Boves (2002, p. 441) laud the “wealth of CAPT systems” available but caution...

When examined carefully...the display of products may not look entirely satisfactory. Many authors describe commercially available programs as fancy-looking systems that may at first impress student and teacher alike, but eventually fail to meet sound pedagogical requirements....These systems, which do not fully exploit the potentialities of CAPT, look more like the result of a technology push, rather than of a demand pull. (p. 442)

Kaiser (2017, slide 45) notes, “Instruction is often at the service of what is easier to program or what will ‘sell’ the app than what is best pedagogy.”

### Characteristics of Ideal CAPT Software

Considering not only the pedagogical but also the technological aspects of CAPT software, Neri, Cucchiari, Strik, and Boves came up with “some basic recommendations for the ideal design of effective pronunciation teaching and learning” software:

Learning must take place in a stress-free environment in which students can be exposed to considerable and meaningful input, are stimulated to actively practice oral skills and can receive immediate feedback on individual errors. Input should pertain to real-world language situations, it should include multiple-speaker models and it should allow the learner to get a sense of the articulatory movements involved in the production of L2 speech. Oral production should be elicited with realistic material and exercises catering

for different learning styles, and should include pronunciation of full sentences. Pertinent and comprehensible feedback should be provided individually and with minimum delay and should focus on those segmental and suprasegmental aspects that affect intelligibility most. (2002, p. 449)

Chun (2013, p. 9) pointed out that “given the technological capabilities that exist, an ideal CAPT program would combine auditory and visualization features, automatic speech recognition (ASR), and appropriate and accurate feedback.” Fouz-González (2015, p. 324) further advocated, “An ideal ASR system would recognize everything the user says, point out those areas that are most problematic (depending on the user’s priorities...) and then offer explicit feedback indicating how to improve.”-

## **Current CAPT Reality**

In the real CAPT world today, pronunciation-focused websites and mobile apps rarely measure up to all these ideal criteria and expectations. Eleven years ago, Levis complained, “pedagogically, a significant gap often exists between CAPT applications and goals advocated by current pronunciation theory and pedagogy, such that CAPT applications can look suspiciously like traditional, drill-oriented pedagogy in new clothing” (2007, p. 185). More recently, Kaiser (2017) analyzed 30 L2 pronunciation teaching/learning apps and found the same problem. Of the 30 apps, 22 (73.3%) relied heavily on a simple listen-and-repeat instructional approach and provided no feedback to learners regarding the accuracy of their production. A few apps provided visual feedback in the form of spectrograms.

What about the eagerly anticipated, long-promised benefits of ASR? Kaiser’s (2017) analysis determined that the few mobile apps he examined that employed automatic-speech-recognition provided simplistic, dichotomous “correct” or “incorrect” feedback that was not necessarily accurate. It appears that in spite of the “great potential for the provision of automatic feedback on learners’ pronunciation” (Fouz-González, 2015, p. 328) that it holds, ASR “needs to improve substantially before learners can use these systems autonomously and rely entirely on their judgments. The effectiveness of these systems decreases significantly when dealing with non-native speech... and ASR ratings do not always correlate with those by human raters....In spite of advances in the field, an acceptable level of reliability is only guaranteed when the tasks are simple and utterances are kept to a restricted set from which students select a response...something that limits the usability of this technology for spontaneous practice.” L2-accented speech produces “numerous false alarms and low rates of correct detection....The experience may be quite frustrating for users if mistakes are not detected or are detected incorrectly...as the machine is supposed to be an ‘expert’ they can rely on. Once learners suspect the system is not reliable, they will lose confidence in it” (Fouz-González, 2015, p. 328). In sum, using inadequate ASR software can “lead to frustrating and counter-productive experiences if learners waste time trying to match a model when their pronunciation is already acceptable” (Fouz-González, 2015, p. 327). Apparently, ASR software has still not reached the point where it provides reliable feedback to L2 learners. In this area and others, ASR “has a way to go before meeting [the above noted] goals” (Chun, 2013, p. 9). On the bright side, in recent years, as the application programming interfaces (APIs) —such as IBM Watson or Google Speech (T. D. Kehoe, personal communication, September 7, 2018)— underlying them have improved, the accuracy of computer-based ASR dictation systems has increased (McCrocklin, Humaidan, & Edalatishams, 2018), and the evaluation validity and reliability of ASR in pronunciation apps will surely do likewise.

For all these pedagogical and technological reasons, L2 teachers and learners must exercise caution when selecting CAPT software. They must consider a variety of criteria. For instance, “there might be pronunciation software programs whose interface may look attractive but fail at reflecting solid grounded principles for teaching pronunciation” (Martins, Levis, & Borges, 2016, p. 143). Some L2 pronunciation websites and apps may provide articulatory explanations but no practice. Others might require an accompanying teacher or textbook since they provide practice but minimal explanation or guidance for learners.

Many other differences can be found in the online resources that exist for L2 pronunciation teaching and learning. In terms of monetary cost, some are free, while others require users to pay a fee. Some sites focus only on segmentals, others on suprasegmentals, and a rare few provide instruction and practice with both. In terms of visual design, some websites or apps provide helpful graphics, some contain only text, and a few provide video clips to

help learners both see and hear how to pronounce English sounds correctly. Some programs allow for a more flexible, individualized approach, in which different learners may choose different learning paths, while others expect everyone to follow the same curricular path. In sum, the number of characteristics to consider when analyzing, evaluating, or selecting CAPT software is extensive! Potential users run the risk of focusing on some features or criteria while ignoring others—at their peril. Choosing incorrect or inadequate software can lead to diminished or unfavorable learning experiences.

To help language teachers and learners avoid such problems, the purpose of this paper is to provide a comprehensive set of criteria for analyzing and evaluating computer-assisted pronunciation teaching (and learning) software, websites, and mobile apps.

## **The System for Analyzing and Evaluating CAPT Software**

Derwing and Munro (2015, p. 124) urge teachers evaluating, selecting, or recommending CAPT software to “read reviews and recommendations from authoritative sources and then to screen apps carefully before recommending them to students.” The system I share in this paper is a tool for conducting such screening. It includes information that will guide teachers and learners of L2 pronunciation in selecting the most appropriate and helpful online resources for their learning/teaching needs. This tool was developed over many years. It started with Persichitte’s (1995) “Basic Criteria for Selecting and Evaluating Instructional Software.” It was expanded with elements from Epstein and Ormiston’s “Criteria for Developing and Evaluating Materials” (2007, pp. 9-10). Then, various criteria developed by pronunciation and CALL experts (Derwing & Rossiter, 2002; Martins, Levis, & Borges, 2016; Munro & Derwing, 2006; Neri, Cucchiarini, Strik, & Boves, 2002; Rosell-Aguilar, 2017) were added to produce a two-page listing of characteristics or criteria that potential users of CAPT software, websites, or mobile apps should look for and evaluate before deciding on a particular instructional product. For years, I have used different, pilot versions of this system to evaluate language-teaching software. In the last few years, I have refined and focused it, and my graduate students and I have successfully used these criteria for evaluating CAPT software (Henrichsen, 2019; Henrichsen, et al., 2018).

Figure 1 shows what the front of the two-page evaluative system looks like. Figure 2 shows the reverse side. The criteria are organized in five sections: (1) general descriptive information, (2) instructional purpose(s) and activities, (3) functionality and usability, (4) instructional factors, and (5) presentation—plus a summary section. The first page consists of “fill in the blank” or “check” items for (A) general descriptive information and (B) instructional purposes and activities. For instance, the last item under item 4 in this section asks specific questions about automatic speech recognition (ASR)—if it is used in the software being evaluated—such as “How often (\_\_\_ %) does it reject correct/acceptable pronunciation as incorrect?” On the second page, users evaluate descriptive statements using a five-point scale with additional “does not apply” and “cannot tell” options to rate (C) functionality and usability, (D) instructional factors (such as D-7. “Provides various speech models [i.e., multiple speakers’ voices],” which refers to the benefits of high variability phonetic training [Bradlow, 2018; Wang & Munro, 2004]), and (E) presentation. The wide-open items in the “Summary” section allow evaluators total freedom in describing what they see as the strengths and weaknesses of the software. Unfortunately, space limitations do not allow me to describe and defend the inclusion of each criterion factor in detail.

## **Conclusion**

I have found the guidance this system provides to be very helpful in focusing attention on the multitude of characteristics that need to be considered when evaluating or adopting CAPT software, and my students have also benefited from it. It often draws attention to factors that evaluators or potential adopters might otherwise ignore. The result is a more thorough, professional analysis of the software product being evaluated. I offer it to you for the same reasons. I hope you will find it useful when you need to select or evaluate CAPT software.

**Criteria for Analyzing and Evaluating Computer-Assisted Second-Language Pronunciation-Teaching and Learning Software, Websites, and Mobile Apps**

© 2019, Lynn Henrichsen

**A. General descriptive information**

1. Name of evaluator: \_\_\_\_\_ 2. Date of evaluation: \_\_\_\_\_
3. Software title: \_\_\_\_\_
4. Copyright (or last update) date of software: \_\_\_\_\_
5. Author, sponsor, or publisher's name (and qualifications): \_\_\_\_\_
6. Platform:  iOS,  Android,  Macintosh,  Windows,  Linux,  other \_\_\_\_\_
7. Target language(s): \_\_\_\_\_ (If English:  British,  North American,  other \_\_\_\_\_)
8. Language(s) in which instructions are provided: \_\_\_\_\_
9. Cost: \$\_\_\_\_\_,  free,  ads,  subscription,  other \_\_\_\_\_
- 10 Target audience age(s):  children,  teenagers,  adults,  other \_\_\_\_\_
11. Target audience language level(s):  novice,  intermediate,  advanced,  other \_\_\_\_\_

**B. Instructional purpose(s) and activities**

1. Primary objectives specified by program (e.g., focuses on developing intelligible [not native-speaker] pronunciation): \_\_\_\_\_
2. Other (secondary, peripheral) objectives: \_\_\_\_\_
3. Aspects of pronunciation addressed (check all that apply):
  - Listening perception:  segmentals,  suprasegmentals, other \_\_\_\_\_
  - Segmentals:  vowels,  consonants,  consonant clusters, other \_\_\_\_\_
  - Suprasegmentals:  intonation,  word stress,  sentence stress,  rhythm,  pausing/juncture,  blending,  reduction,  other \_\_\_\_\_
  - Fluency:  pausing appropriately,  delivery speed,  other \_\_\_\_\_
4. Type(s) of learning activities provided (check all that apply):
  - Analysis or diagnosis of learner's difficulties
  - Listening perception or discrimination
  - Listen to a model and imitate/repeat
  - Listen, record, replay, listen, and compare to model
  - Minimal pairs
  - Variable input (using speech models of different genders, regional dialects, registers, etc.)
  - Contrasts between the learners' L1 sounds and corresponding L2 sounds
  - Loaded sentences or "tongue twisters"
  - Phonetic alphabet symbols and/or charts
  - Articulatory explanations (text or video)
  - Articulatory displays (sagittal section diagrams)
  - Animated "mouth movement" models (video)
  - Visual pitch-contour displays
  - Spectrograms, waveforms, or formant data (of model and/or learner's pronunciation)
  - Flash cards
  - Instructional game
  - Pair-work or group-collaboration activities
  - Automatic speech recognition (ASR) How accurate is the ASR? How often (\_\_\_%) does it reject correct/acceptable pronunciation as incorrect? How often (\_\_\_%) does it accept incorrect pronunciation as correct? How does it do with different dialects?
5. Feedback and record keeping:
  - Provides users with immediate feedback on the correctness of their responses
  - Provides pronunciation feedback that is easily understandable by L2 learners
  - Provides users with helpful, comprehensible guidance on how to correct their mistakes
  - Tracks the number of right and wrong responses for an individual user
  - Keeps track of various individuals in one class and reports scores to a teacher

**Figure 1.** First page of system for analyzing and evaluating CAPT software

Rate each of the items in sections C, D, and E according to the following scale:					
-2	-1	0	1	2	
strongly disagree	disagree	no opinion	agree	strongly agree	
NA=Does not apply    CT=Cannot tell (insufficient data)					
<i>Write comments anywhere they fit or on a separate sheet (please refer to item numbers).</i>					

**C. Functionality and usability**

- |   |    |    |   |   |   |    |    |
|---|----|----|---|---|---|----|----|
| 1. Runs properly (i.e., no bugs, crashes, long delays, etc.) .....  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 2. Guides the user well (i.e., intuitive interface, provides clear directions for starting, navigating, and stopping) ..... | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 3. Uses consistent commands and directions throughout .....   | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 4. Provides operational “Help” for users .....  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 5. Allows users to provide feedback or ask questions to the creators.....   | -2 | -1 | 0 | 1 | 2 | NA | CT |

**D. Instructional factors**

- |   |    |    |   |   |   |    |    |
|---|----|----|---|---|---|----|----|
| 1. Presents information well (i.e., clearly, concisely, interestingly etc.) .....           | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 2. Provides adequate, thorough, and effective practice .....                                | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 3. Provides helpful feedback.....   | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 4. Focuses on priority aspects of pronunciation (e.g., functional load).....                | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 5. Content is authentic, up to date, and accurate.....                                      | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 6. Presents speech in contexts (not just unrelated, individual words).....                  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 7. Provides various speech models (i.e., multiple speakers’ voices) .....                   | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 8. Delivers instruction at a level appropriate for the target audience .....                | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 9. Maintains a constant (or gradually increasing) level of difficulty.....                  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 10. Presents teaching/learning activities in a good sequence .....                          | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 11. Provides helpful interaction with the user(s) .....                                     | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 12. Allows for learner autonomy and independence .....                                      | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 13. Allows users to repeat activities they have difficulty with.....                        | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 14. Allows individualization (learners choose which pronunciation features to work on)..... | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 15. Provides meaningful practice (using words learners know) .....                          | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 16. Provides communicative practice (bridging an information gap) .....                     | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 17. Provides variety in practice activities .....   | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 18. Promotes metacognitive activity regarding pronunciation .....                           | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 19. Encourages learners to take responsibility for their improvement .....                  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 20. Encourages learner strategy development .....   | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 21. Supports a variety of learning styles (e.g., visual, auditory, etc.).....               | -2 | -1 | 0 | 1 | 2 | NA | CT |

**E. Presentation (User interface)**

- |  |    |    |   |   |   |    |    |
|--|----|----|---|---|---|----|----|
| 1. Uses appropriate, readable text (size, style, variety, and continuity).....   | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 2. Avoids distracting elements (unnecessary sounds, animations, ads).....        | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 3. Is not too busy or confusing (e.g., employs “white space” appropriately)..... | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 4. Utilizes an attractive, appropriate color scheme.....                         | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 5. Is aesthetically pleasing in general and looks professional .....             | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 6. Audio clarity level is high.....  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 7. Audio volume is adequate and adjustable.....                                  | -2 | -1 | 0 | 1 | 2 | NA | CT |
| 8. Audio can be played at different speeds (fast, slow).....                     | -2 | -1 | 0 | 1 | 2 | NA | CT |

**F. Summary**

1. Strong points?
2. Weak points?
3. Other comments?

Figure 2. Second page of system for analyzing and evaluating CAPT software

## References

- Bradlow, A. (2018, September). High variability training in the lab and in the language classroom. Plenary presentation at the 10<sup>th</sup> annual PSLLT (Pronunciation in Second Language Learning and Teaching) Conference, Ames, IA.
- Chun, D. M. (2013). Computer-assisted pronunciation teaching. In C. A. Chapelle (Ed.), *Encyclopedia of applied linguistics* (pp. 823-834). Oxford: Wiley-Blackwell. doi: 10.1002/9781405198431.wbeal0172
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, The Netherlands; Benjamins.
- Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, 30, 1515-166.
- Epstein, R., & Ormiston, M. (2007). *Tools and tips for using ELT materials: A guide for teachers*. Ann Arbor: University of Michigan.
- Fouz-González, J. (2015). Trends and directions in computer-assisted pronunciation training. In Mompean & Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314-342). Basingstoke & New York: Palgrave Macmillan.
- Henrichsen, L. (2019, February). 21 online pronunciation resources for teaching and learning. *TESOL Connections*. Retrieved from <http://newsmanager.commpartners.com/tesolc/issues/2019-02-01/2.html>
- Henrichsen, L., Devenport Blanco, K., Carreño, S., Carter, S., Decker, L., Fry, L... Zhao, K. (2018). Online resources for learners and teachers of English language pronunciation. *TESOL Reporter*, 51(1), 23-89. Retrieved from <https://tesol.byuh.edu/sites/tesol.byuh.edu/files/Henrichsen%20et%20al%20HOT%20pdf.pdf>
- Kaiser, D. J. (2017). iPronounce: Understanding pronunciation apps. Online Webinar, 8 June 2017. Recording and slides available at <http://ollren.org/events/past-events#s-lg-box-wrapper-17792711>
- Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184-202.
- Martins, C. G. de F. M., Levis, J. M., & Borges, V. M. C. (2016). The design of an instrument to evaluate software for EFL/ESL pronunciation teaching. *Ilha do Desterro*, 69(1), 141-160. doi: <http://dx.doi.org/10.5007/2175-8026.2016v69n1p141>
- McCrocklin, S., Humaidan, A., & Edalatishams, I. (2018, September). ASR dictation programs accuracy: Have current programs improved? Paper presented at the 10<sup>th</sup> annual PSLLT (Pronunciation in Second Language Learning and Teaching) Conference, Ames, IA.
- Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, 34, 520-531.
- Neri, A., Cucchiari, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5), 441-467.
- O'Brien, M. G. & Levis, J. M. (2017). Pronunciation and technology. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8<sup>th</sup> Pronunciation in Second Language Learning and Teaching Conference* (pp. 1-9). Ames, IA: Iowa State University. [https://apling.engl.iastate.edu/alt-content/uploads/2017/05/PSLLT\\_2016\\_Proceedings\\_finalB.pdf](https://apling.engl.iastate.edu/alt-content/uploads/2017/05/PSLLT_2016_Proceedings_finalB.pdf)
- Persichitte, K. (1995). Basic criteria for selecting and evaluating instructional software. In D. A. Willis, B. Robin, & J. Willis (Eds.), *Technology and teacher education annual, 1995 (Proceedings of SITE 95—Sixth International Conference of the Society for Information Technology and Teacher Education (SITE))* (pp. 379-382). Charlottesville, VA: Association for the Advancement of Computing in Education.
- Rosell-Aguilar, F. (2017). State of the app: A taxonomy and framework for evaluating language learning mobile applications. *CALICO Journal*, 34.2, 243-258. doi: <https://doi.org/10.1558/cj.27623>
- Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32, 539-552. doi: 10.1016/j.system.2004.09.011