Kate Richardson

5 December 2019

Final Project Proposal

Professor Schwartz, IDN 526

Speaker's Toolkit: A Language-learning Application for L2 Pronunciation

An Instructional Design Proposal

Abstract

This proposal is to develop an application called *Speaker's Toolkit*. The purpose of this instructional tool will be to equip the language learner with the knowledge and tools needed to effectively practice pronunciation in the target language.

There will be three major components, organized as three discrete sections of the app, which will reflect the instructional objectives. The first component will be pedagogical in nature, with the goal of familiarizing learners with the International Phonetic Alphabet (IPA). It will consist of modules that contain videos, diagrams, and written explanations to explain the science behind articulatory phonetics. This is not designed to be a linguistics course; the purpose will be to provide the learner with just enough conceptual understanding about sound production to have a useful framework for identifying and practicing target sounds. The second component will be the actual pronunciation tool, which enables effective pronunciation practice by displaying visual feedback that compares the learner's attempted utterance to the target utterance of a native speaker. Learners will be able to rehearse sounds, words, and phrases, which can be helpful for learning how to produce difficult phonemes (sounds), stress patterns, and prosody or tones, respectively. Furthermore, it will collect data that allows the student (and professor, if LMS integration has been achieved) to track their progress. Finally, the third component will simply

be an interactive IPA to consult for reference. Altogether, this software application will help the language learner by providing the knowledge and tools needed to improve pronunciation in the target language.

Problem Analysis

Problem Description

Pronunciation is an essential skill for any speaker wishing to become conversant in a foreign language. Recently, there has been a swell of interest in explicit pronunciation practice (Olson 2014). Pronouncing a language well is important because good pronunciation increases intelligibility, allowing the speaker to be well understood by native interlocutors, as well as increasing positive attitudes of native speakers towards the learner. These factors lead to better success in immersive environments for language learners.

Explicit pronunciation practice has been found to improve speakers' pronunciation skills. For example, Sturm (2013) found that university students in French immersion courses who were also enrolled in a phonetics course that trained them explicitly in pronunciation showed significantly greater improvement than a control group of French learners not concurrently enrolled in a phonetics course. Effective pronunciation rehearsal typically involves visual feedback. Studies in Computer Assisted Language Learning (CALL) have shown that visual feedback improves pronunciation (e.g., Suemitsu et al. 2016). Saito et al (2010) found that form-focused instruction — training learners about speech production — combined with corrective feedback, resulted in more accurate product of the English 'R' by Japanese learners of English than a control group and a group that received only form-focused instruction. Offerman and Olson (2016) found that language learners who received instruction with visual pronunciation feedback improved significantly compared to a control group. They found that these

improvements were generalizable to spontaneous speech, suggesting strong transferability in skills acquired during pronunciation training. This research shows that visual feedback is most effective when combined with immediate corrective feedback about pronunciation. This technology is useful for learning the prosodic and pitch features of a language as well as the suprasegmental features, such as working on the production of individual vowels and consonants (Olson 2014). Therefore, visual feedback combined with repeated rehearsal of different pronunciation-related challenges can improve overall speaker performance.

Thus, recent research suggests that computer-assisted, explicit pronunciation instruction with visual feedback is effective. However, actual classroom practice is slow to follow paradigm shifts in the research. Despite the increasing prioritization and desirability of pronunciation today, students rarely receive explicit pronunciation instruction in the classroom. On a recent survey by Sturm (2016), language instructors on a survey rated the importance of pronunciation for successful communication as 3.64 on a five-point scale, with 3 as slightly disagree and 4 as slightly agree. Educators report that evidence from empirical research is impractical or inaccessible. The applied research requires technical tools and knowledge of phonology that instructors do not possess or feel they have the time to implement. Teachers also reported lack of pronunciation instruction in TESOL methods courses, which may influence priorities of uses of classroom time because instructors are more likely to focus on what they feel more equipped to teach (Baker 2014).

Research shows that pronunciation is often neglected in classroom settings, despite the fact that learners generally consider pronunciation an important skill that they are willing to practice. An online tool that provides explicit pronunciation instruction with a mechanism for effective rehearsal would meet the desires of learners who wish to gain oral fluency, as well as

their instructors who feel that it is important but do not have the time to implement it themselves. Speaker's Toolkit will ideally be integrated into a course's curriculum as a resource that students may use on a regular basis outside of the classroom, so that it is not competing for valuable class time. Furthermore, the pronunciation tool can gather data on student performance, so that the instructor may gather empirical data to decide which sounds, words, or phrases warrant the use of class time for rehearsal and further explanation.

Review of Resources

Many web-based resources for exploring sound production already exist.

One good example of a resource that is helpful for learning the IPA is an application called *IPA Phonetics* by Coey et al. which is available in the iTunes store. This app showcases an interactive, elaborate IPA chart. Clicking on any one IPA symbol allows the user to hear the sound and see a video of the relevant articulators in action. This resource provided inspiration for the idea to include videos that play on the button press of certain IPA characters in the *Interactive IPA* section of the application. Similarly in the *Learn* section, videos and diagrams of the vocal tract will be shown alongside the relevant parts of the IPA. While my application and this one have similarities, it is important to note that *IPA Phonetics* is not explicitly instructional. There are definitions but it does not walk the user through learning the IPA.

Some IPA charts are modified to reflect only the relevant sounds for the target language. A user of *Speaker's Toolkit* will reach a point where they compare two IPA charts that have been modified only to reflect the inventory of their native and target languages side-by-side. An example of a resource that uses comparison of IPA charts as an instructional method for second language acquisition can be found at http://www.keytap.com/french/Compare/IPAChartCompareJPG.aspx. However, this format

makes direct comparison difficult because the two charts are never shown on the same screen side-by-side. Additionally, the charts are displayed but no explanation is provided to highlight the differences and explain how the learner might apply this information to pronunciation practice.

There are a few tools that provide a platform for explicit pronunciation practice, of which *SpeechAce* is the most robust. With this API, users may listen to native speaker pronunciation, record their attempt, and then receive a computer-generated grade in the form of a percentage point about the accuracy of pronunciation. The software transcribes the audio input, and after sufficient data is collected, identifies problem phonemes to be the focus of more rehearsal. The audio is also parsed such that learners may listen to their own phoneme-by-phoneme production. This way, the user is directed to specific and individualized practice. This works very similarly to how the *Practice* section of *Speaker's Toolkit* will function. However, the application assumes that the native user is able to read the IPA without first providing training. Furthermore, there is no visual feedback (other than progress bars) that allow the student to interpret their own errors in articulation. A further limitation of this application is that it is for learning English only.

The widely-known application (at least among phoneticians) for generating visual feedback from audio input is PRAAT. This open-source software analyzes sounds and generates a variety of visual feedback. It will certainly be integral to the visual feedback element of the pronunciation tool in the *Speaker's Toolkit.*

A good example of a website and application that provides comprehensive, explicit instruction in pronunciation is the *Learning English* section of the BBC. A large inventory of instructional videos provides conceptual knowledge for the specific sounds of English. Notably, each of these videos uses IPA transcriptions alongside videos of the articulations. Users learn

about a variety of important concepts in phonetics and are encouraged to practice their application while viewing the video. This instruction is entirely video-based, so there is no corrective, visual feedback based on the speaker's own language production. Learners who complete this course will have gained hours of dedicated study of pronunciation as well as a conceptual understanding about articulatory phonetics. The goals of this *Learning English* by the *BBC* therefore closely align with those of *Speaker's Toolkit*, although it uses a different method of instruction that does not provide direct feedback to user input.

To the author's knowledge, there exists no tool that is freely accessible on the web that provides both a brief conceptual foundation of articulatory phonetics alongside an interactive tool that provides corrective visual feedback. *Speaker's Toolkit* will attempt to provide both in an application that is general enough to be useful to learners of all languages (although the application itself will be in English) along with a tool for rehearsal that takes advantage of speech recognition technology.

Design Considerations

Learner Characteristics

The primary user base of *Speaker's Toolkit* will be undergraduate university students enrolled in introductory and intermediate language courses. Collectively, the majority of undergraduates in the United States today belong to Generation Z. These learners have matured in particular environmental and societal conditions that have given rise to characteristic preferences and attitudes. Namely, this generation of "digital natives" has matured in an era of ubiquitous technology and immediate access to information, a poorly funded school system largely driven by standardized testing, and a difficult job market. In their book *Generation Z*

Goes to College, Corey and Grace (2016) describe some of the consequences of these factors on their educational experiences and learning preferences, which I summarize below.

Members of generation Z have always had the internet at their disposal. They know how to quickly search and retrieve information from the internet, and are generally comfortable switching between tasks and working with multiple screens. They can quickly find and access informational videos to aid them in their assignments, and for many students this has become a routine process of their workflow. However, all the stimulation technology provides may have a serious impact on the average attention span; one study by Vidyarthyi (2011) found that the average attention span in this population has decreased by half. Therefore, they will likely value the app more highly if the technological aspects are highlighted, and if their overall comfort with navigating technology is acknowledged.

Another major trend that Seemiller and Casey (2018) found is that Generation-Z students prefer to learn independently and at their own pace, perhaps as a result of the ubiquity and personalization of internet-based resources. However, despite their preference for independent learning, they generally enjoy collaboration. Corey and Seemiller (2018) point out that despite the increasingly obsolete need for hard copies of books and periodicals, libraries remain central to the academic life of campuses as study and collaborative spaces. With this balance for autonomy and social learning in mind, I have chosen to make this tool a desktop application in order to encourage the students to work in a computer laboratory, which is a library-like space where they are likely to work independently but alongside their peers. Additionally, working in a computer lab will encourage focus and consistency in sound quality since they will be using identical headsets during each session.

A final generalization is that the Gen-Z cohort is highly pragmatic and career-driven (Seemiller and Casey 2018). The authors suggest that this preference is linked to coming of age during the recession, which has likely impacted many of their families. In general, they see their university experience as building skills that they will use in the "real world." They gravitate towards courses and learning experiences that offer employable skills that they believe will be helpful in their future professions. They describe themselves as responsible, determined, and thoughtful, and enjoy learning that is self-reliant. Therefore, *Speaker's Toolkit* should emphasize the pragmatic, job-related benefits of fluency in order to motivate learners.

Altogether, these findings indicate that an application designed for Gen-Z learners should take into account their desire for autonomy, potentially shorter attention spans, pragmatism, and technological fluency.

Finally, it is important to note that although university students in introductory language courses will be the target audience, this tool will be able to help any mature, motivated learner wishing to gain oral fluency in a language or dialect. Independent learners attempting a new language, actors and performers practicing a new dialect, or anybody else who wishes to learn about the basics of articulatory phonetics in the context of language learning would benefit from *Speaker's Toolkit*.

Goals and Objectives

The overarching goal of *Speaker's Toolkit* is to improve the learner's pronunciation in the target language. This application proposes using two parallel methods to achieve this goal. The first method is to develop the learner's conceptual knowledge about phonetics and phonology, so that they may become a self-aware learner in their continued acquisition of both the target

language and any future languages they wish to acquire. The second method is to improve pronunciation by imitating native speaker recordings and interpreting visual feedback.

1. Learn how to navigate the International Phonetic Alphabet.

The International Phonetic Alphabet (IPA) is organized a conceptually logical way such that that if a user knows how to read and consult it, then they already possess a sufficient understanding about the production of sounds. For this reason, and also because it is a tool that is frequently used in text and reference sources about linguistic materials, instruction in the principles of articulatory phonetics will be presented as a series of modules for learning the IPA.

The goal will not be for the user to memorize the IPA. Instead, the user should be able to recognize the names of the natural classes (row and column names in the IPA) and be able to give an example of a phoneme associated with each class along with a brief but accurate description. Additionally, they should be able to fill in a blank diagram of the vocal tract.

2. Learn how to properly pronounce sounds relevant to their target language.

This objective more directly addresses the overall learning goal of improving pronunciation performance. ‘Sounds’ covers an entire range of utterances, from individual sounds to intonation of an entire sentence. It is impossible to specify the specific sets of sounds the learner will need to acquire, because these will be personalized based on the parameters of the speaker’s native language(s) and the target language. However, whatever the target language, there will certainly be specific learning objectives that address each of the following categories:

1. Novel sounds: The user will learn how to recognize and produce sounds that are not used in their native language.

2. Modifications to known sounds: The learner will make slight adjustments to sound more native-like. This is less of a priority than learning the novel sounds.

3. Other factors: The learner will learn how to adjust stress patterns, tone, intonation, and rhythm to sound more like a speaker of the native language.

3. Identify areas of difficulty.

Another learning goal is for the learner to be self-aware about which sounds they struggle to either perceive or produce. If they are aware of sources of potential communication issues, they can more efficiently mitigate them by focusing on these aspects during practice.

The user will be able to identify areas of feedback by reading immediate, automatically-generated feedback each time they attempt an utterance. They will also be able to view a summary score at any time that automatically lists the phonetic features that give the user the most trouble. This summary will clearly indicate the aspects of pronunciation that the user struggles with. The app will also aid the user in developing simple goals in order to promote motivation and focus.

Content Analysis

Some of the content of the analysis will be generated based on the user's native language(s) and target language, but there is also content that will be universal to all users. The content in the introductory modules will teach the principles behind the International Phonetic Alphabet, which are listed below.

Since the International Phonetic Alphabet is already organized by the natural classes (groups of sounds sharing some common feature) that the learner will need to know, instruction will be centered around those concepts and vocabulary.

Below is a list of concepts that the user should recognize:

- Places of articulation: bilabial, labiodental, dental, alveolar, postalveolar, retroflex, palatal, velar, uvular, pharyngeal, glottal

- Manners of articulation: plosive, nasal, trill, flap, fricative, approximant, lateral approximate

- Anatomy of the vocal tract: pharynx, larynx, uvula, velum, tongue root, blade, body, epiglottis

- Vowel: frontness, height, roundedness, schwa

- Other concepts and symbols from the IPA: labial-velar approximant, primary and secondary stress, long vowel, syllable breaks, voicelessness, aspiration, dental, apical, no audible release, affricate

- Other concepts important to phonology: phonemic, phonetic, phones, homophony, allophonic variation, dialects, assimilation, gemination, obstruents and sonorants, liquids

- Imaging technology; oscillation, spectrogram

- Reading a spectrogram: formants

Project Description

The homepage of the application is a login screen, where the user may log in as a private user, as a student in a class, or sign up for the first time. Once logged in, all users follow the same basic template for simplicity and ease of navigation. On the left-hand side of the screen is a navigation bar that has three options: Learn, Practice, and Interactive IPA. Below, I describe each of these sections of the application in more detail.

In terms of development of the application, the prototype will be designed only for native English speakers learning French. However, I use language-neutral expressions when possible to emphasize that the eventual intention is to support a large number of combinations of native and target languages.

Learn

The *Learn* section will consist of a collection of learning modules. At top of the screen a suggested next unit will be prominently displayed to encourage a logical instructional order. It is either automatically generated, based on performance collected from the *practice* sections and from the last unit covered. The professor always has the option to overrule the next suggested module and choose a specific one to be completed in a prespecified time frame. Some modules are more general and conceptual and conceptual in nature, whereas others are more language-specific and cover production or perception of specific sounds. The user will always have access to all modules in case they wish to practice something specific.

The longest and most robust module will be the first set, which is dedicated to learning the IPA. This set of lessons will teach the learner concepts that are important to understanding the mechanics of sound production. The first part of the lesson will use stories and humor to show the users why it is important to learn pronunciation. Then, the module will explain

everything that the user will expect to learn and what they might expect to do with this information as part of their language learning journey and beyond.

After this introduction, the user first takes a “tour” of the consonant chart. As each different section of the chart is explained, the relevant components are highlighted and the rest of the chart is visible, but faded to make it clear which section is the current focus. The user controls when to proceed on to the next section by clicking on the ‘next’ button when ready. On many screens, the user is instructed to practice producing specific related sounds. These introductory modules will be the most terminologically and conceptually loaded parts of instruction, demanding a high cognitive load from the learner. Therefore, the instruction will exercise best practices for multimedia instruction, particularly the signaling principle (relevant portions are indicated by visual indicators such as highlighting) the segmenting principle (instruction is user-paced), and the coherence principle (only the information needed to understand the chart is shown). These lessons will be kept short to maximize focus and minimize risk of burnout.

Upon completion of the consonants module, the user is encouraged to take a break for a day or two and return later to review and then begin the vowels, which are somewhat more complex. The tutorial of the vowel chart will follow the same principles as the consonants. Both modules will be relatively brief and focus on learning how to read the chart to locate the appropriate symbol instead of the details of production for each sound (this will come later, in the language-specific instruction). The user will also learn to group different types of sounds in meaningful ways by learning about the technical categories of sounds. After the vowels component, the user will be encouraged to practice material covered so far in the P*ractice* section.

Next time the user returns to the *Learn* section of the application, they will transition into the language-specific material. At this point, the instruction is much more highly detailed and less conceptual. First, the IPA chart of the native language (i.e. its phonemic inventory) will be shown directly next to that of the target language. The user will be encouraged to note any differences before proceeding to the instruction.

The first instruction will identify the commonalities between the two inventories—saying something like *Keep all these sounds, you're going to need them!* Then, the instruction will go through all the phonemes that are similar but are expressed slightly differently— (i.e. *This 't' looks the same, but all 't's in French are actually labiodental— they are pronounced with the tongue against the teeth!*) With each new phoneme, the user will practice enunciating different monosyllabic words in the target language, using minimal pairs and cognates when possible. However, the purpose is not to perfect pronunciation at this point; the goal will be to familiarize the learner with the specific sounds they will need to practice later. Once all the shared phonemes have been covered, new phonemes will be introduced. These will receive the most attention in future mini-modules, but in this introductory module each one will be introduced and the user will have only a brief opportunity to practice.

To conclude this introductory module, the user will be shown videos of users speaking with a strong accent *in the native language of the user*. That is, if the user is an English speaker learning French, they will watch a video of a heavily accented French person speaking English. The user will then explain why the French person has trouble pronouncing the *th* sound for example (i.e. because it doesn't exist in French).

At this point, the user has likely covered several new sounds and learned some surprising facts about pronunciation. As before, the user will be encouraged to revisit after a day or two

before moving on to comparing the vowels, which in general are harder for language learners to master than new consonants.

The next module of the *learn* section will focus on the new vowels. Again, a side-by-side comparison will be shown to compare the native language inventory to that of the target language. The user will again first identify vowels that are the exact same, and then the completely new vowels. They then practice the new vowels as they did before with the consonants, and answer a few comprehension questions that are designed to prompt their thinking about why it can be difficult to acquire the ability to produce new sounds.

At this point, the user should be comfortable reading and identifying IPA symbols relevant to both their native and their target language. They will need the to use the IPA symbols in order to use the pronunciation tool most effectively.

After the learner has completed these mandatory prerequisite introductory modules, they will have a large inventory of instructional modules to choose from that focus on individual sounds and phonological characteristics. For example, one module might focus on the intonation of questions, another might focus on the nasalization of medial vowels. These units will be relatively short (probably about five minutes to complete the short reading, video, and practice), and are intended to be visited one at a time throughout the duration of a semester-long course, so that the user is continually improving pronunciation.

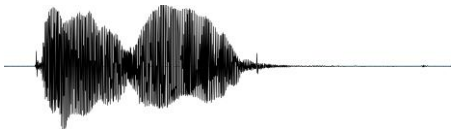
The *Learn* component will likely be important for beginning users of the application, but will be visited only very briefly (likely only as the instructor assigns units) after the introductory modules. The development of this section of the application will rely heavily on input from phoneticians, phonologists, language instructors (all of whom will be the SME's) and lots of user

testing for comprehension. Furthermore, the application be able to easily integrate new modules and support for new languages as they are developed.

Practice

The *Practice* section of the application will be where users go to test and improve their pronunciation skills. For users who are assigned to practice as a part of the course, the set of utterances to record may be hand-picked by the professor. Otherwise, the set will be automatically generated based on the user's current level, the units they have already completed in the *Learn* section, and the sounds that need to be improved based on data from previous practice sessions.

As with the application as a whole, the setup of the page will be minimal. Each page will only have one utterance for the user to immediately repeat. The phrase will be spelled out in the target language, as well phonetically (in the IPA). If the user scrolls over the word in the native orthography, they will see the translation in the native language (because memorization of vocabulary is not a concern of this resource). If the user scrolls over any individual phoneme in the IPA transcription, they will see the official name of that IPA character to refresh their memory in case they've forgotten. (It's unlikely this feature will be used often, but it is unintrusive and may be helpful to some users).



Directly below the word and IPA transcription will be a play button to hear the native speaker's utterance, followed by a waveform. The waveform, generated by the software Praat, will look something like the image shown at right. It will always correspond with the recording played upon pressing of the playback button that is directly next to it.

Directly below the native speaker's waveform will be the record button. When the user presses the record button, they are prompted to speak into the microphone to record their attempt. When done, they press the record button again, and then are asked "save recording?" if they would like to see the results of the attempt. If they say *no*, they are prompted to record again. If they say *yes*, the recording is processed, and it can not be deleted. This is to ensure the statistics about user improvement are valid. Only after the user chooses *yes* will they be able see their score.

Audio processing behind the scenes is an involved process that the user will not be aware of. The details of the audio processing are not important from a design perspective; however, it is important to note that the audio processing tool will normalize the pitch and speed of the speaker's utterance, so users should be assured that they don't need to match the absolute pitch of the native speaker or repeat the utterance at the same speed. Therefore, neither gender of the speaker nor speed of speech should have a consequence on the speaker's performance. Additionally, multiple native speaker recordings who have slightly different accents within the target dialect will be 'averaged' together and compared against the attempted utterance, in order to account for an acceptable range of variation with a given target dialect. All of these measures will be taken to ensure that the pronunciation score is as valid as possible.

After the user submits their pronunciation attempt, the visual representation of their attempt immediately appears. The oscillogram of the attempted utterance will appear directly below that of the native speaker utterance in order to maximize ease of comparison.
The default representation is the bare oscillogram with a pitch contour line overlaid (which provides visual feedback about the duration of each phoneme, aspiration, volume, and the pitch) but the user may choose to also show a spectrogram, which provides highly detailed feedback

about acoustic information for advanced users. (A module for how to read a spectrogram will be available in the *learn* section). Lines through the oscillogram will denote the approximate division between phonemes. An IPA transcription of the attempted utterance will also be generated if the utterances are single words. Phonemes that do not sufficiently match the native speaker's representation will be marked with a red X. The user will also see a score for each attempt, which is calculated based on how well they matched the target utterance.

Once the user has completed all assigned utterance attempts, they see a summary page which lists the new vocabulary, any progress on old utterances, and tips about what to practice for next time.

Interactive IPA

The application will also feature an interactive IPA, which is to be consulted for reference or for practice learning the IPA. If the user hovers their cursor over a symbol, its official name will appear. If they click on the symbol, a video of the articulators in action appears (i.e. a video of the relevant part of the vocal tract creating the sound). There will be a *quiz me* option, which asks the user to pick the relevant symbol based on a sound that is played. This helps them practice discrimination of sounds as well as mastery of the IPA table.

User Experience

John is a freshman in an undergraduate French course. Since he studied French for three years in high school, he was placed into French 102. He is a beginner but not an absolute beginner. Like many of his classmates, he is familiar with basic French grammar and has a respectable vocabulary, but knows that his pronunciation needs work. To his surprise, his professor said that a part of his daily French homework will consist of learning about, and rehearsing, French pronunciation using a tool called *Speaker's Toolkit*. He is not thrilled about

what feels like “extra” work. However, he genuinely wants to be conversational in French one day and knows that in order to get there, his pronunciation will have to improve. Therefore, although using the application sounds annoying (especially because he was told he has to be in a computer lab equipped with headsets to do it), it will hopefully be worth the effort.

The first time he opens the application, he has to enter a secret code in order to join the class. This way, he can see assignments that his own professor has posted to the application Speaker’s Toolkit. (Or, rather, the professor posted the assignments to Moodle but they are somehow linked to the *Speaker’s Toolkit*). Now, the application automatically knows that he is an English speaker learning French.

The webpage is minimalist— there are three choices on the left of the screen: *Practice, Learn,* and *Interactive IPA* (whatever that is). The screen which takes up most of the page has a brief welcome message informing him that he is part of the class, that he should return at least every other day. The message tells him to click on the *Learn* button, so he does.

The *Learn* button leads him to a screen that has a bit more text and a *next* arrow. It grabs his attention because it is humorous, reads more casually than his textbook, and promises that it doesn’t take a lot of time to become a better speaker. His homework assignment was to look around the application and do the first module. So far, this isn’t too bad.

The module starts to feel a little less like a game and more like a lecture as it explains why it is important that he learn pronunciation and shows him around the application, but it’s tolerable because some of the stories it tells and videos it shows about the consequences of mispronunciation are relatable and kind of funny. He is surprised to find out that there is a whole science behind the art of pronunciation, and even though it hasn’t mentioned French at all yet

and is quite jargony, it seems useful and there's a lot to click on. He decides to proceed to the next unit to learn a little bit more about consonants.

When he comes back a couple days later (to complete his homework for class the next day), he spends more time learning the consonants because that was his homework and the questions are difficult enough that he has to actually pay attention. Even so, it takes him about 10 minutes to get through the entire consonants module. That was all that was assigned for today, but he decides to click on *practice* to see what it's all about.

Immediately he sees a word— "Bonjour" — that he has to try to pronounce. It's fairly intuitive, so he has no trouble clicking and recording his attempt. When he saves it, a funny looking chart that must represent the sounds he has produced shows up. He now sees why the professor requires them to do the recordings using the headset in the language learning lab— the high quality sound reception allows for a very detailed analysis! It picks up every little sound. He can read his scores, which are 70-80% for the most part. A few words he messes around with to see if he can achieve a higher score, and he actually manages to score 100% for a few (after many attempts).

For class the next week, the professor has actually assigned a few phrases as well as the vowels unit. He learns new things about French vowels, such as the fact that French has many more vowels than English and that they are generally much shorter in duration. This actually makes sense, because he could see it in the picture that that computer made when he recorded his voice (in a lab, this time). His vowels were longer (which he could tell because the computer wrote were the "o" sound was) than the native speaker's. He kept trying until the vowels were the shortest length. It felt funny to say, like he wasn't pronouncing the word properly, but the computer told him it was more accurate! When he played back his recording to the native

speaker's, he was surprised to find that it did in fact sound better when he focused on shortening his vowels!

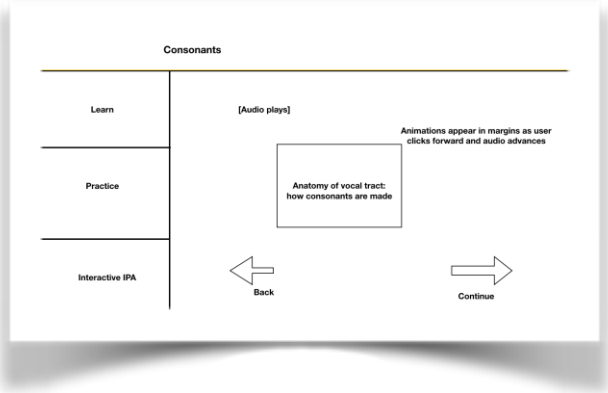
Most words he breezed through without a problem, getting 90% on the first attempt, but a few words (particularly those with main stress on the vowel) gave him more trouble. The visual feedback was actually pleasing to look at. It was kind of like a game, adjusting his voice to change the picture to look like the native speaker's. He even felt a little obsessive about a few words, trying again and again until all target sounds matched up. From the immediate feedback, he knew that he struggled with vowel duration and a couple specific sounds like the vowel /y/ and the consonant /r/, which he now knew were called phonemes.

Since the professor assigned practice on this application (for which he received a participation grade in his French course), he continuously visited this application throughout the semester. Although it was kind of annoying because it felt like extra homework, it wasn't too bad because it didn't take a long time, felt like a game, and he could see his progress over time. It was satisfying to see progress.

Evaluation

Evaluation in Speaker's Toolkit is straightforward: if the speaker has improved their pronunciation, then the resource was successful! Pronunciation improvement is measured every time the speaker records an utterance. Their score is recorded in the backend, and the professor can see a summary of the average class performance for each utterance. The user can also view a progress report at any time and judge their own progress for themselves.

If speakers improve in subsequent sessions from the pretest, then the tool was successful. It is also successful if they articulate why certain sounds may still challenge them despite continuous practice; one of the learning objectives is to have a scientific understanding about



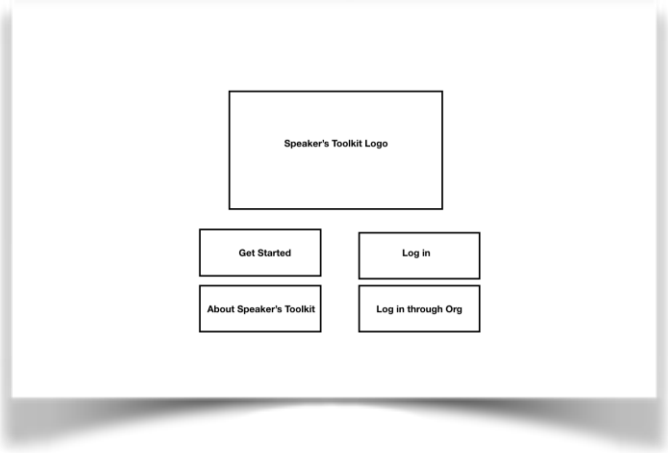
sound perception and production. If they are aware and able to articulate the reasoning behind continuing challenges, then they may effectively direct their efforts to pay attention to these specific challenges in future experiences with the language. The application does not assess conceptual understanding directly, but if speakers voluntarily consult the interactive IPA chart, this would be a sign that they wish to master the sounds so that they can expand their knowledge about language.

Prototype Materials

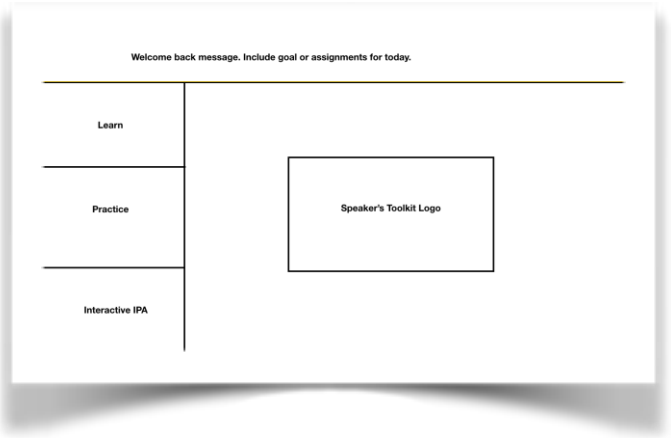
The wireframes below provide an outline of the proposed structure of the application:

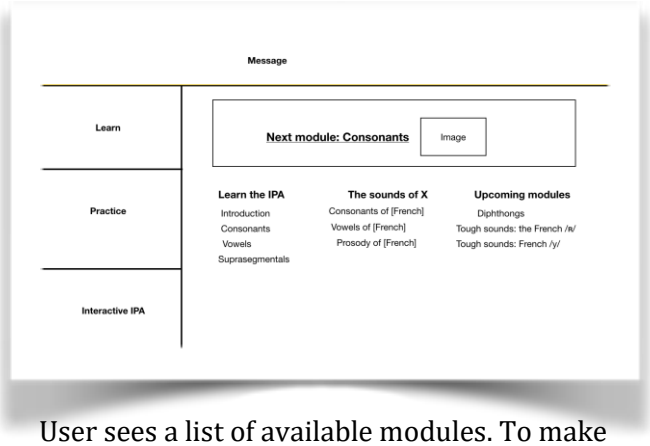
The log in screen features four options: two log in options- one for independent users and one for users as part of a class- one button for getting started if they don't yet have an account, and one to learn more about the application.

The home screen shows the template that is consistent on all other screens that the user may access. The welcome message flashes

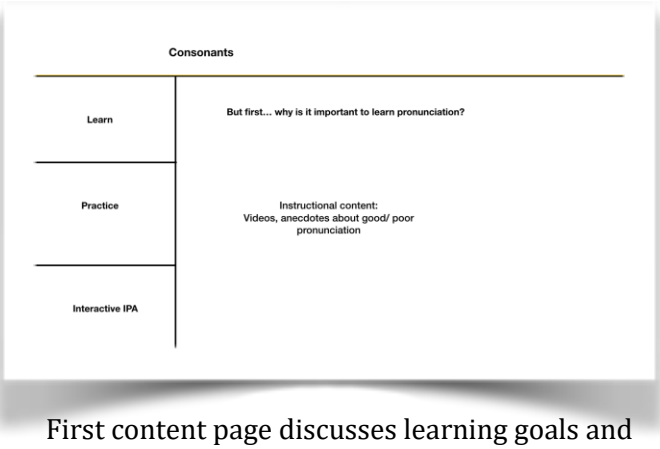


once and then remains on the screen. It includes a motivational message, generates a goal, and/or lists homework assignments. Below is a storyboard for an early module, where the user learns about consonants while walking through the IPA. This is probably the most-content heavy module.

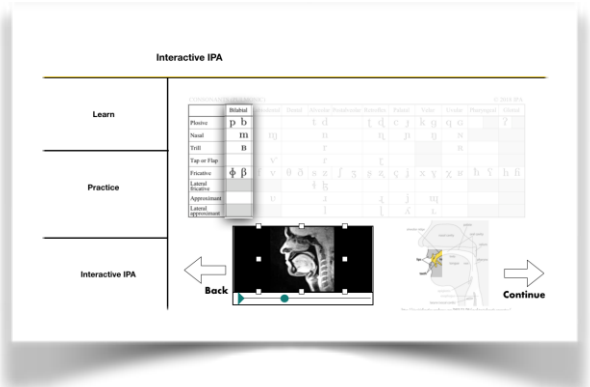




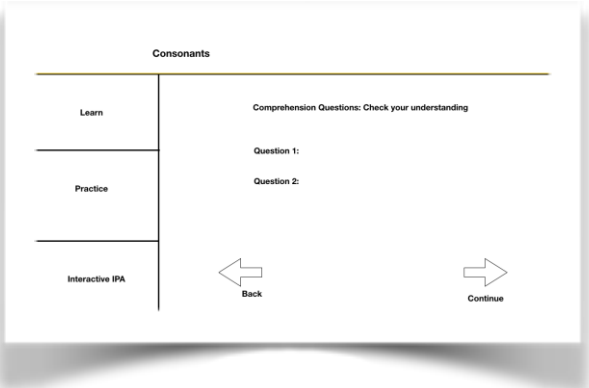
User sees a list of available modules. To make choice easiest, the recommended next module is clearly highlighted at the top of the screen.



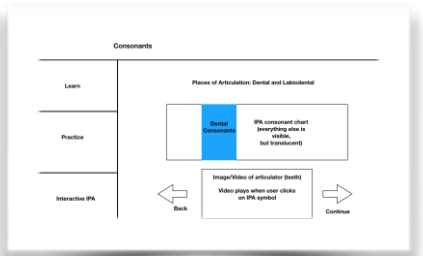
First content page discusses learning goals and why they are important.

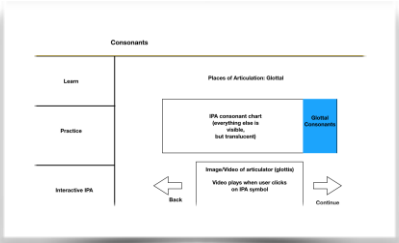


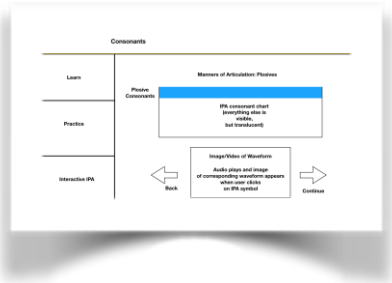
As the relevant row or column is spotlighted, an accompanying video illustrates the concept while narration plays. If the user clicks on one of the spotlighted phonemes, the video shows footage of this sound being articulated.

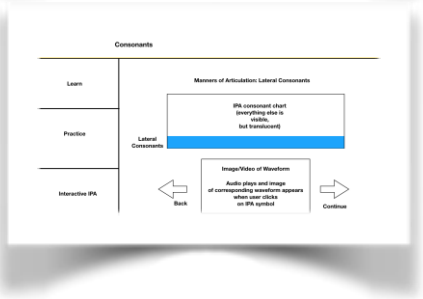


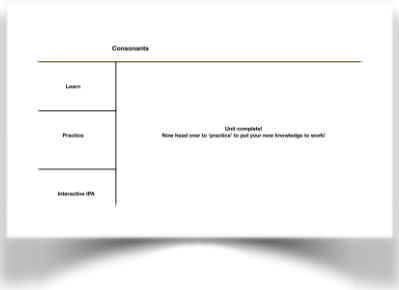
Periodically, concept checks prompt the user to assess their comprehension.



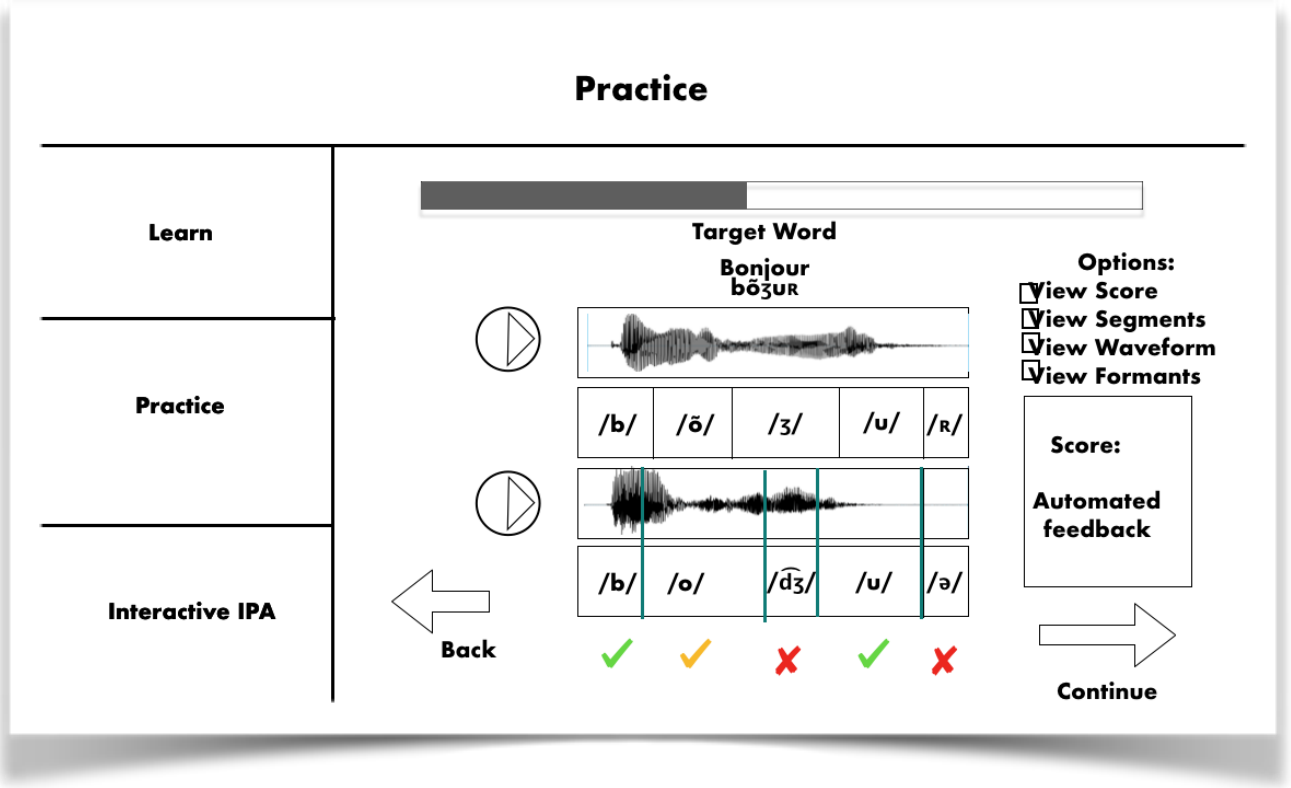




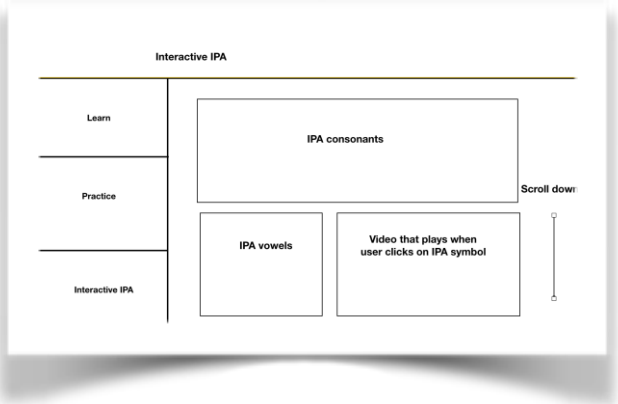


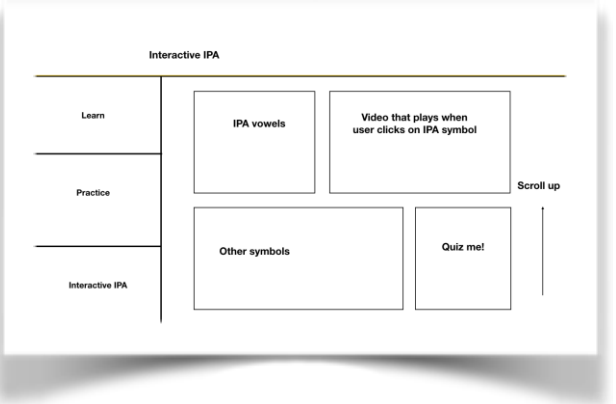


With each new page, the user hears some audio giving a definition. After a brief audio clip, the user may click on the phonemes in the interactive IPA chart until they are ready to move on to the next row or column.



Above is a wireframe for the 'practice' section. This is the layout for the screen *after* the user has listened to the native speaker and recorded their own attempt. On top is the native speaker recording and waveform. Below is the user's attempt. Both oscillograms are juxtaposed with the IPA transcription that also indicate information about length. They are stacked vertically for the most easiest comparison possible.





This shows the layout of the Interactive IPA section. On this page, the user can scroll up and down to navigate the chart.

Theoretical Considerations

The design of *Speaker's Toolkit* adheres to multimedia principles of instructional design in order to maximize the learning outcomes and minimize the time and effort the user must exert to achieve these outcomes. Three principles that are especially relevant to this design proposal are the signaling principle, the segmenting principle, and the personalization principle. In this section, I will briefly describe how each principle informs the design of *Speaker's Toolkit.*

The signaling principle states that highlighting general concepts helps learners gain new knowledge. By providing cues that signal the most essential information, users' attention may be guided to the most critical information, which reduces a bit of cognitive effort to prevent overload (Mayer and Fiorella 2014, 280). I plan on using the signaling principle in both the *Learn* section, which imparts declarative knowledge about many basic phonetic principles by spotlighting the relevant parts of IPA chart, as well as in the *Practice* section, which will use several subtle techniques to direct the learner's attention to the most pertinent aspects of their practice.

In the *Learn* section, users will walk through each row and column of the International Phonetic Alphabet as they in order to learn about the different natural classes (phoneme categories) and distinctive features. I plan to have the entire chart visible to the learner on every screen. Although this seems to be in violation of the coherence principle —which states that only relevant information should be included and displayed in instruction (Mayer & Fiorella 2014, 280) — since I will technically be displaying irrelevant information, I choose to display the entire chart because the position of the natural class within the chart provides useful information, as the chart is organized spatially by row (from the front of the vocal chart to the back) and somewhat qualitatively by column (more similar sounds are generally closer on the vertical axis).

The vowel chart is organized entirely by tongue position in the oral tract, so the vowel chart will be especially useful to show all at once. In order to help the learner focus on current natural class that they are learning about, I will highlight the entire row or column and make the rest of the chart more translucent.

There are many more ways to apply the signaling principle in other l*earn* units. Once the learner has completed the general module, they will observe the various phonemic inventories of the their native and target language. The phonemic inventory, by definition, already spotlights the relevant phonemes because only those phonemes used in the language are shown. When comparing two phonemic inventories, instead of introducing it all at once, first those phonemes that are the same will be highlighted (by bolding them and making the rest more translucent, in order to maximize visual contrast). Then those phonemes that bare semblance but are different in some minimally contrastive way (for example, the /t/ and /d/ being pronounced with the teeth instead of the alveolar ridge) will be highlighted. In the third section, the completely novel phonemes will be highlighted. This should help the learner distinguish between the phonemes that will not be used at all, those that they will need to learn and use only in the target language, those that they already use but need to articulate differently, and those that are the same. This way, when must articulate a new word, they can immediately identify the difficult sounds that they should pay most attention to for best results. Hopefully, these sounds should become automatic to both perceive and produce fairly quickly.

The signaling principle will also be put to use in the *practice* section through the use of feedback. The feedback will automatically generate a suggestion based on the the sound that is most dissimilar to the target phoneme. (Although a checkmark or ‘X’ will be assigned to each attempted phoneme, only a single written suggestion will be offered so that the user has one

thing to focus on in order to minimize essential cognitive load. Later attempts may generate other suggestions as the user becomes a better speaker). Thus, the feedback draws attention to the most important discrepancy and the user's attention is prioritized.

The signaling principle also applies in the way the visual feedback is represented; it will consist of an oscillogram that automatically divides the speech into phonemes. If the 'perceived' phoneme matches the target phoneme, a checkmark appears directly below the relevant section of the oscillogram. Otherwise, an 'X' will appear, along with the automated transcription of the actual utterance. Dividing the oscillogram into sections based on approximate phoneme boundaries, along with the juxtaposed icon that provides feedback, will help the user pay attention only to those sounds which are the most relevant.

Many of the design decisions that take advantage of the signaling principle also apply to the segmenting principle, which states that learners learn best from instruction that is user-paced, where the learner must focus on one thing at a time. Restricting how much information the user sees at any one time helps to manage the essential processing, or mitigates the challenges presented by what might otherwise be too much information for the user to effectively acquire (Mayer & Pilegard 2014, 317). The segmenting principle supports the decision for the general flow of *Speaker's Toolkit*, which involves a user clicking through many screens that contain a little bit of information, instead of a scrolling through a large page that contains all the information. There is a large number of concepts that Speaker's Toolkit expects the learner to be able to recognize once they have completed all the modules. Furthermore, since articulatory phonetics is somewhat orthogonal to actual language learning, it is expected that the learners will not have encountered many of these concepts before. Therefore, the intrinsic load is relatively high. This poses a challenge because *Speaker's Toolkit* is assumed to be used as a companion to

an actual language course, which means that these students already have a high intrinsic load from other aspects of their language learning endeavor. Therefore, reducing extraneous cognitive load is of utmost important to this project.

Another reason for the design to adhere so strictly to the demands of the segmenting principle is that it helps to gamify the experience in the *Practice* section. If the user is focusing on one segmented 'challenge' at a time, it may have more of the effect of playing a game to meet the challenge, beat their previous score, and 'level up' to a new, more difficult challenge. For example, once the user has mastered all the new sounds, they can 'level up' to focus on intonational phrasing.

The segmenting principle will apply in both the *Learn* and the *Practice* section in different ways. In the *Learn* section, the segmenting principle applies because each screen will (whenever possible) cover only one concept or natural class which has a pace that the user has control over. The user may choose when to proceed to the next page (or previous page) via forward and backward arrows. Additionally, the learner has some control over how much material they will cover. It will be easy on the site to pick up where the user left off since the last visit. If they wish to continue to cover more modules, they have the option of continuing. However, they can not do everything at once; if a professor gives an assignment, the student must complete it in that timeframe, in order to prevent users doing too much in one session.

In the *Practice* section, users rehearse their pronunciation by imitating utterances of native speakers and receiving immediate visual (and a bit of textual) feedback. This will be segmented because each screen will feature the information for only one utterance. Each screen in the practice section will have the target utterance along with an oscillogram (and a spectrogram, should a user choose to see this advanced visual feedback). Exercising the

segmenting principle therefore also allows for easy adherence to the spatial and temporal continuity principles, since only data relevant to the target utterance will be shown. Once a user records their attempt and confirms the attempt, the audio will be processed and the oscillogram (and/or spectrogram) of the speaker's attempt will be shown directly below that of the native speaker for easiest comparison. The user may add as many attempts (and therefore oscillograms) to the screen as they choose, but this still follows the segmenting principle because the user is practicing this one utterance. Furthermore, the textual feedback will only offer advice for one aspect of the speaker's target utterance, which encourages the user to focus on one thing at a time. These design choices allow for limited demands to the speaker's cognitive capacity, as they are in control moving from one utterance or attempt to the next.

Another principle that influences the design of *Speaker's Toolkit* is the personalization principle, which states that people learn more deeply when words are in a conversational style instead of a formal style (Mayer 2014, 346). The personalization principle is one of several principles that attempts to activate social responses from the learners in order to promote learning outcomes. After all, since language learning is at heart a social endeavor, it is only logical that softwares catered to language learners should adhere to principles of social learning. Additionally, since *Speaker's Toolkit* will attempt to gamify the experience of pronunciation practice as much as possible, an informal tone will be necessary to prevent the game from feeling too pedantic.

The *Learn* section of the *Speaker's Toolkit* will inject some informal prose into the lessons. Informality will come from using colloquial language and second-person narration. Learning how to pronounce different phonemes is often a humorous experience, as the learner manipulates their vocal tract to create sounds and sound patterns that may seem unfamiliar,

bizarre, and (at times) ridiculous. The tone of instruction should take this into account and encourage the learners to have fun with it. For example, if the page is explaining how to pronounce a velar fricative, which would be a novel phoneme for an English speaker, the speaker may be encouraged to think about hissing like a cat instead of more precise anatomical instructions (i.e. avoid verbose phrases like 'keep your tongue root loosely pressed against the hard palate' when a less formal suggestion is available).

Similarly, anecdotes that recount actual or theoretical experiences based on the mispronunciation of certain sounds will add some humor while illustrating the utility of practicing good pronunciation. These anecdotes should appear sporadically, but will be most relevant in the language specific section of learning relevant phonemes.

Of course, the narration will be presented in polite wording, using a conversational tone instead of a directive one, when providing feedback and telling users to complete a task.

The *Practice* section is more difficult to make informal or humorous because of the paucity of text, other than the target utterances and their descriptions. However, it may feel more informal by using some slightly gamified elements. The user may be assigned a score and given 'challenges' that fall outside the assignment. For example, for a 'bonus' fun round, they may try a tongue twister in the target language. The goal of this section will be to feel slightly gamified without using rewards as extrinsic motivation.

Speaker's Toolkit uses the empirically-supported signaling, segmenting, and personalization principles in order to reduce intrinsic cognitive load, increase germane cognitive load, and cue social learning. Therefore, the decisions to signal important information, chunk the information into user-paced segments, and use an informal tone are deliberate and contribute to will contribute towards more efficient achievement of the learning goals.

Conclusion

This proposal has provided justification for the development of an instructional resource called *Speaker's Toolkit.* There is a demonstrated need for a pronunciation tool and instructional resource in undergraduate language courses. The goals of *Speaker's Toolkit* are to provide the conceptual background and technological resources to efficiently acquire pronunciation skills in a foreign language. The instruction will be presented using theoretically-justified methods that maximize effectiveness and cater to the learning preferences of the tech-savvy, practical and hard-working Generation-Z students who will be using it.

References

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly, 48*(1), 136-163. doi:10.1002/tesq.99

Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.43

Fugett, Bob (2013). http://www.keytap.com/french/Compare/IPAChartCompareJPG.aspx

Language Learning & Technology, 18(3), 173–192. http://dx.doi.org/10125/44389

Levis, John (2005). *Changing Contexts and Shifting Paradigms in Pronunciation Teaching.* Tesol Quarterly 39(3), 369-377.

Interactive Phonemic Chart. Retrieved from *https://www.englishclub.com/pronunciation/phonemic-chart-ia.htm*

Mayer, R. (2014). Principles Based on Social cues in Multimedia Learning: Personalization, Voice, Image, and Embodiment Principles. In Mayer, R. (Ed.). *The Cambridge Handbook of Multimedia Learning* (Cambridge Handbooks in Psychology). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139547369

Mayer, R. and Fiorella, L. (2014). Principles for Reducing Extraneous Processing in Multimedia Learning: Coherence, Signaling, Redundancy, Spatial Contiguity, and Temporal Contiguity Principles. In In Mayer, R. (Ed.). *The Cambridge Handbook of Multimedia Learning* (Cambridge Handbooks in Psychology). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139547369

Mayer, R. and Pilegard, C. (2014). Principles for Managing Essential Processing in Multimedia Learning: Segmenting, Pre-training, and Modality Principles. In In Mayer, R.

(Ed.). *The Cambridge Handbook of Multimedia Learning* (Cambridge Handbooks in Psychology). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139547369

Nagle, C, Sachs, R., and Zárate-Sández G (2018). *Exploring the Intersection Between Teachers' Beliefs and Research Findings in Pronunciation Instruction.* The Modern Language Journal 102(3)

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. Language Learning & Technology, 18(3), 173–192. http://dx.doi.org/10125/44389

Saito, K. and Lyster, R. (2012). *Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /r/ by Japanese Learners of English.* Language Learning 62(2), 595-633

Sturm, Jessica L. (2013). *Explicit Phonetics instruction in L2 French: A Global Analysis of Improvement.* System 41(13), 654-662.

Suetmitsu, A. and Dang, J. (2015). A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning. Acoustical Society of America, 138(4), 382-387.

Interactive Phonemic Chart. Retrieved from *https://www.englishclub.com/pronunciation/phonemic-chart-ia.htm*

Reetz and Jongman. Phonetics: Transcription, Production, Acoustics and Perception. 2009. John Wiley & Sons. *Chapter 2: Articulatory Phonetics* & *Chapter 3: Phonetic Transcription* pp. 14-48

Saito, K. and Lyster, R. (2012). *Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /r/ by Japanese Learners of English.* Language Learning 62(2), 595-633

Seemiller, Corey and Meghan Grace. *Generation Z Goes to College.* John wiley & sons, Inc. Chapter 9: Maximizing Learning pp. 227-264

Speechace [computer software] (2019). Retrieved from https://www.speechace.com

Sturm, Jessica L. (2013). *Explicit Phonetics instruction in L2 French: A Global Analysis of Improvement.* System 41(13), 654-662.

http://www.bbc.co.uk/learningenglish/

Suetmitsu, A. and Dang, J. (2015). A real-time articulatory visual feedback approach with target presentation for second language pronunciation learning. Acoustical Society of America, 138(4), 382-387.

Vidyarthyi, N. (2011). *Attention spans have dropped from 12 to 5 minutes: How social media is ruining our minds.*