Vowel Shapes: An open-source, interactive tool to assist singers with learning vowels

Cyndi Ryan, Tait Madsen, Nathan Buckley, Veronika Alex, and Josh Bronstein

ABSTRACT

The mastery of vowel production is considered a basic building block of vocal technique. In vocal music, performers are required to replicate vowels specific to the language and stylistic requirements of each piece, regardless of genre. Currently, vocal instructors employ a number of strategies to instill accurate and healthy vowel production, such as providing verbal description, modeling correct and incorrect vowel sounds, and assigning exercises to encourage proper reflexes in their students. With few exceptions, this instruction is accomplished solely by means of auditory cues. In this paper, we present an open source interactive system, Vowel Shapes, to automatically capture and visualize vowel sounds to assist singers in learning and producing their correct pronunciation. Our system allows the user to listen to the vowel, see its correct vowel shape, and practice replicating that vowel shape in real time. The design of our system was informed by iterative evaluation sessions with a vocal professor and students of vocal music. The final evaluation of the system included eleven vocal students and their professor, of which 70% of the participants learned vowels faster and more effectively using our system compared to traditional methods.

Author Keywords

Application Domains; Technological Innovation; Experience; E-Learning and Education; Auditory I/O and Sound in the UI; Visualization; Multidisciplinary Design/Interdisciplinary Design; Singing; visual; vowels; diction; vocals; music; interactive.

ACM Classification Keywords

H5.2. [User Interfaces]: Input devices and strategies, Interaction styles, User-centered design

Submitted for consideration to
DIS 2014, June 21–25, 2014, Vancouver, BC, Canada

INTRODUCTION

Let's consider Maria, a student majoring in vocal performance at a music school. Maria takes individual lessons, usually one hour in length, with an applied instructor once a week. In addition she has several other opportunities to get the instructor's feedback in class or departmental performances throughout each semester. Otherwise spends most of her individual singing time practicing alone. Practice time is set aside in order to apply the information and experience from her lesson to vocal exercises and assigned vocal literature in various styles and languages. The practice methodology currently implemented in the music school has the following limitation–Maria does not hear what an outside listener hears when she sings. This is due to the internal locations of the respiratory, phonatory and resonating structures that produce human speech and singing. In addition, the feedback Maria receives from her instructor is subjective, non quantifiable, and usually only provided during her lessons.

Thus, when Maria practices alone, she cannot verify that she is singing correctly. Aspects of singing that Maria is concerned about–and would love to get more feedback on–include tempo, pitch accuracy, vowel quality, tone quality, amplitude, nuance, and efficient, healthy emission. In her situation, an automated interactive system would allow her to determine whether she is singing with the correct technique. Is possible to design and validate an interactive system that could assist musician like Maria?

In this paper, as part of an initial exploration, we propose a dynamic interactive practice tool for vocal students: Vowel Shapes. Vowel Shapes was conceptualized as an interactive system that would automatically translate a vowel sound to a visual shape. For example, when students practice a vowel they are continuously making minute adjustments to their vocal tract and articulators to match an internalized vowel sound. They are currently limited to their memory of the target vowel and their aural, kinesthetic, and subjective abilities to decide when the vowel has been matched. In our work, we hypothesize that a real time visual vowel shape could significantly improve the students' ability to master a vowel more rapidly and accurately. The visual vowel shape would need to be interactive to continually provide students feedback as they make subtle vocal changes to the vowel

(Figure 1). Ideally, students should be able to understand how the vowel shape changes when they adjust their breath pressure, vocal tract conformation and articulators as they

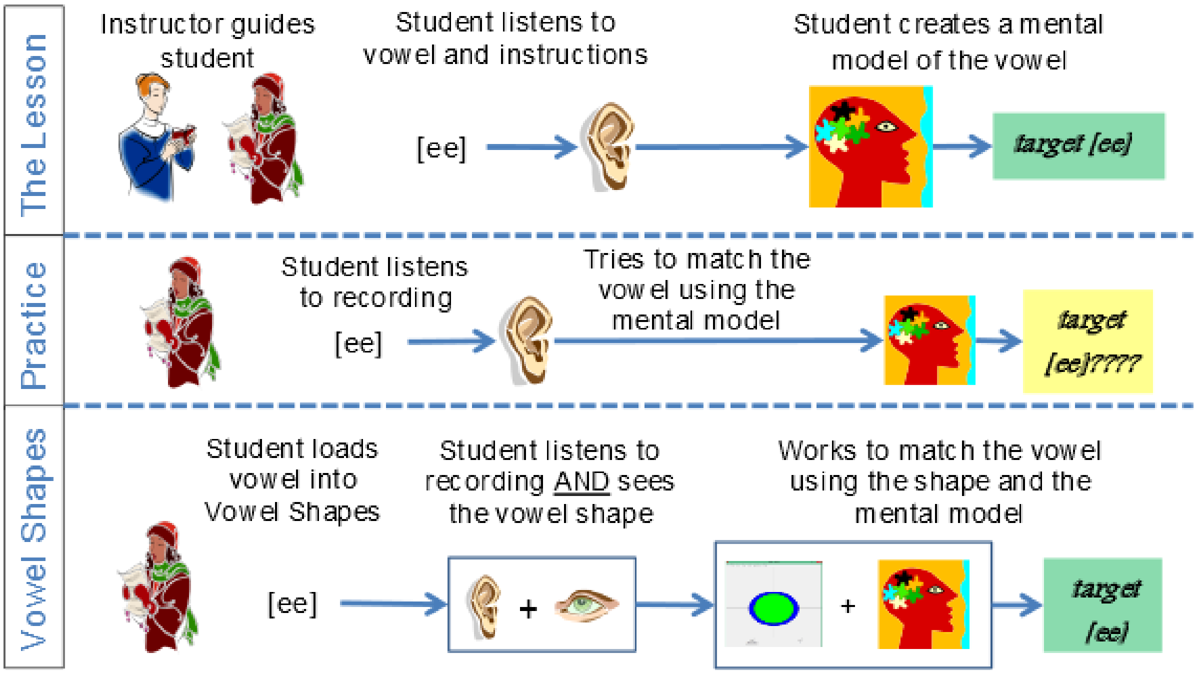

Figure 1: Learning a vowel – current method and using Vowel Shapes

vocalize the vowel. For example, the music student Maria should be able to say to herself "When I move my tongue upwards the vowel sound is made brighter and the shape becomes flatter", and express similar statements. The shape should consistently reflect these vocal production analogies across all vowels.

Vowel Shapes is an automated, interactive system that can capture a live audio feed from a microphone and visualize it based on known values for vowel sounds. To define the shapes of the vowels, we use formants [8].

Formants are found in speech as well as singing and can quantitatively distinguish each vowel sound. They are the amplitude peaks in a frequency spectrum of that sound within the human range of hearing. In a frequency spectrum, the amplitude peak with the lowest frequency, above the fundamental pitch f0, is called the first formant, or f1. The amplitude peak with the second lowest frequency is f2, the third is f3, and so on. A vowel can be distinguished by f1, f2, and f3. Specific vowel formants are known and well defined [8].

In order to design and validate Vowel Shapes, we teamed up with a major music school and completed a number of design iterations in collaboration with vocal students and their respective instructors. Each iteration provided an opportunity for improved ease of use and technical accuracy. Based on the findings of the iterative design, we proposed a customized algorithm and a normalization scheme to account for the individual voice differences of each singer and their gender differences. By involving the users in our design, we were able to adjust the tolerance for how close the user's vowel is to the target vowel as opposed to opting for a one-size-fits-all approach. Based on users input, we also implemented a smoothing algorithm to ensure that the interactive nature of the shape, while being useful, does not disrupt the focus of the student.

The rest of the paper is organized as follows: in the next section, we provide a discussion of current vowel practice methodologies. The next section describes a process for analyzing audio to characterize a specific vowel. We then provide a detailed description of Vowel Shapes and the design process used for the initial prototype. Next, the initial prototype evaluation process is described and the results are presented. The final section discusses the limitations of the current prototype as well as future possibilities for the system.

BACKGROUND RESEARCH

The current methods to teach vowels to students include instructor or vocal coach-based learning in secondary and university settings. For practice the student may have access to handwritten notes and/or recordings from the lesson. If individual instruction is not available a singer may practice listening to a recording of a master performer and imitating the perceived vowel sounds. This often happens when a student is learning a piece in a foreign language What the singer hears in these instances is influenced by the quality of the recording, the existing knowledge of the desired vowel sound from prior language study and the specific musical context of the vowel in question. In all cases, the student is influenced not only by the vibrations carried to the ear externally, but also by the vibrations carried to the ear internally through bone conduction and other perceived pressure sensations ranging from the chest upward to the neck and head.

Further methods for practicing vowels may include watching laboratory videos of the vocal tract or reviewing written instructions available through scientific publications, pedagogy books or web sites [2, 6, 18]. A few interactive websites exist that allow the user to select characteristics of the sound to be produced and a visualization of the vocal tract is provided [5]. These methods do not address the individual needs of each student and lack any sense of interactivity. Though the practice material is static, generalized, and useful as a way to train the student in the general formation of vowels by the human vocal tract, this method does not address the specifics of the student's own vocal tract.

Diction for Singers.com, a subsidiary of Celumbra, has developed an on-line classroom for instructors [3]. The instructor creates a class from one or more lessons from the book "Diction for Singers". For each vowel in a lesson the instructor creates a recording of a correct vowel. Students register for the class and may practice the vowels

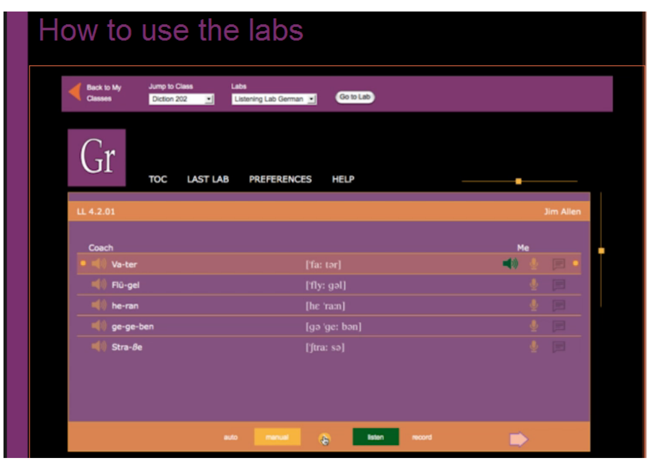

Figure 2: Screen shot from the Learners Lab demo site.

for each lesson by playing the instructor-recorded vowels. The student may record their results for the instructor to review and provide feedback (Figure 2). A benefit of this method is that students may focus on one specific vowel at a time in a practice setting. But this method does not provide any real time feedback to assist the student with the mastery of the vowel. Students must wait for the instructor's assessment.

Software tools that provide spectrograms or other visualizations of the audio have been developed [7]. One of the more recently developed applications is "Sing & See" [1]. This application provides a real-time visual feedback to the student with a variety of screens. A pitch screen charts the pitch at which the student is singing. A spectrogram screen shows the harmonic structure of the vocal tone and is used to monitor vowel quality and other desired qualities (Figure 3). The spectrogram is only available in the professional version available for $99 (USD); it is not an open source application.

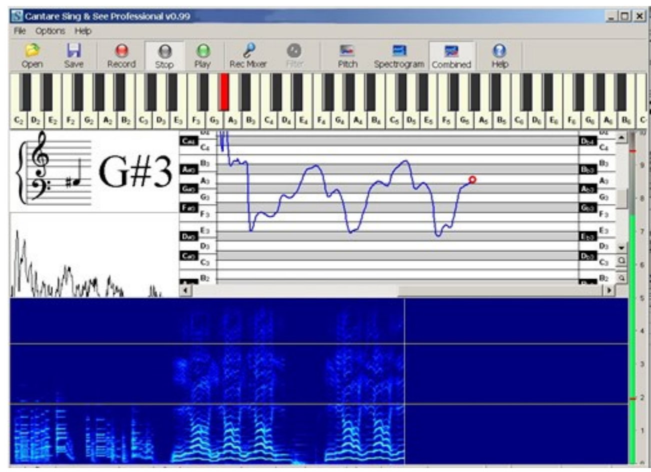

Figure 3: Screen shot from Sing & See [1] Professional

"Sing & See" can be used as part of the lesson with the instructor and the for individual student practice. In the lesson the instructor advises the student on how to correctly sing a vowel. The instructor will provide comments on how to improve the vowel production and the student views the results of the bodily or vocal tract changes on the display. When used during practice the student relies on the information provided by the display and the memory of the instructor's advice. There is no mode for comparing a correct visualization of the vowel to the current vowel production in the practice room. Students' ability to view the pitch screen requires purchase of the professional version.

DESIGN OF VOWEL SHAPES

Throughout our design and prototyping process, the design of our system was informed by interviews with five students of vocal music, nine responders to a survey, found here: http://goo.gl/1foDmf, and constant contact with a vocal professor with over 20 years of experience in professional opera singing. Our review of the current practice methodologies for vowels concluded that they primarily compare the student's auditory components to that of the teacher's, or provide written guides for vowel production. Some practice software uses spectrograms as a visual aid to the students for vowel production. To get a sense of how familiar the students are with techniques such as spectrograms, we designed the previously mentioned survey, and sent it the students of vocal music at our the aforementioned partner school, and received nine responses. Our survey results indicate that out of nine singing students surveyed, none were familiar with spectrograms. To the best of our knowledge, no such tool exists that allows the students to visually compare their vowel sound during practice with a target-vowel visualization. Mastery of a vowel in current tools consistently lies with the instructor and with the students' memory of the instructors' subjective recommendations.

Therefore, in Vowel Shapes, we aim to achieve the following objectives to address current limitations:

- Real time visual feedback during the production of a spoken or sung vowel
- Interactive visual representation of a target vowel
- Intuitive and generalizable visualization that maps the vowels sounds as part of real time feedback.
- An assessment on the efficacy of our tool with students and professor from a music school.

Our design introduced several challenges and design considerations. First, the visual needs to be based on a common definition of a vowel—a specific articulatory and vocal tract shape. A characterization of the vowel needs to accommodate differences in gender and pitch. The use of a vowel shape that is easily understood by students is critical to the tool. The vowel shape must be generated from the real time capture and analysis of the students' singing. To assist the student with determining mastery, the instructor's singing must be captured, analyzed, and the vowel characterized and saved. The instructor's vowel characterization may then be loaded into the tool and the student can practice the vowel, striving to match the instructor's vowel shape, to gain mastery.

Vowel Characterization

Vowels, whether uttered during speaking or singing, may be characterized by using frequencies, or formants, filtered from an audio input. Vowels have well defined formant definitions. Vowel formants may be normalized using Barks Formant Normalization [11, 13] to permit vowel comparison across physiological or anatomical differences [15] and gender [10]. By analyzing the audio a vowel can be characterized with normalized formants.

When viewing a two-dimensional space, such as a computer screen, individuals are likely to more quickly comprehend a two-dimensional visualization than any higher dimension visualization [16]. Calculating a modified Bark Difference metric [11, 13] from the normalized formants results in a two-parameter vowel characterization. The two parameters are the height of the vowel, Y, and the extent of the vowel, X. The Bark Difference metric also permits the comparison of vowels independent of pitch. A two dimensional shape that the student understands is possible with the {X, Y} vowel characterization (Figure 5).

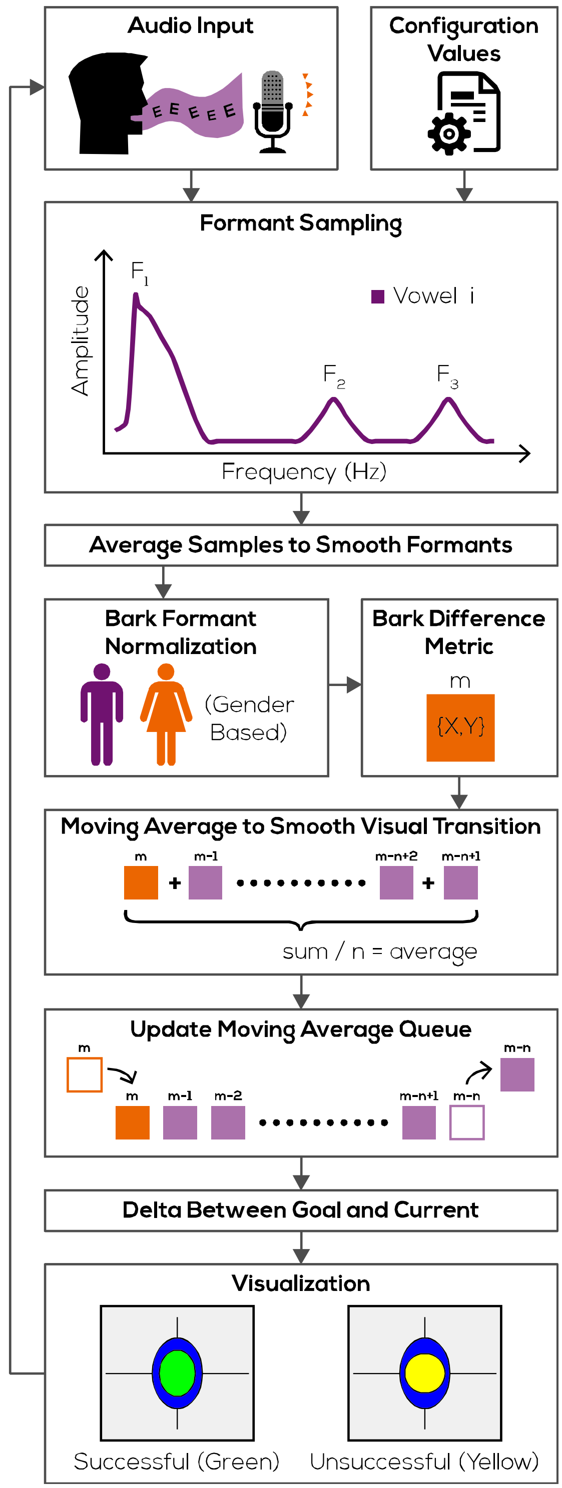

Figure 4: The design of Vowel Shapes

Bark Formant Normalization; Z_i can be expressed by

$Z_i = 26.81/(1 + 1960/F_i) – 0.53$ Male

$Z_i = 26.81/(1 + 1960/F_i) – 1.53$ Female

Bark Difference Metric

$X = Z_3 – Z_2$

$Y = Z_3 – Z_1$

where F_i is the value for a given formant *i*.

Design of a vowel shape

Given an {X, Y} pair of coordinates a set of vowel shapes may be considered in two-dimensional space. We informed our design through discussions with a linguist, a speech prosody expert and a singing professor. In addition, five singing students were interviewed and asked to consider a set of vowel shapes. Based on these interviews three initial vowel shapes were selected: an ellipse, a triangle and a {X, Y} grid.

The triangle is based at the origin of the vowel shape coordinate system (see Figure 5). The length of the first leg off shooting from the origin into the first quadrant is set by the {x} parameter. The {y} parameter is used to calculate the angle between the triangle base leg and the {x} leg that extends from the origin. The base leg is held constant for comparison purposes.

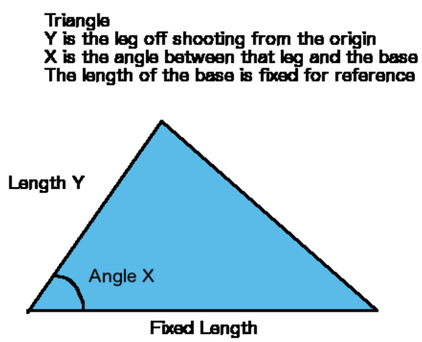

Figure 5: An example of an instructor triangle.

The ellipse vowel shape uses the {x, y} vowel parameters as major and minor axes length system (see Figure 6). The ellipse is centered on the origin of the vowel shape coordinate.

The {X, Y} grid plots the {x, y} parameters. The vowel shape is shown in the negative x, negative y quadrant, as per normal with linguists. This is consistent with the plotting of normalized formants (see Figure 7).

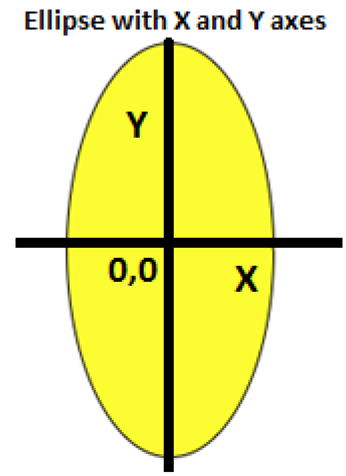

Figure 6: An example of an instructor ellipse. Vowel parameters are major and minor axes lengths

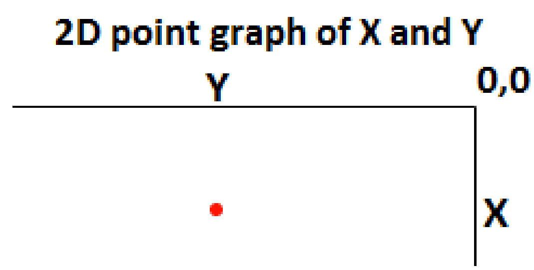

Figure 7: An example of an instructor {X, Y} grid. Parameters are plotted in the negative/negative quadrant.

Audio Analysis

In order to develop our prototype, we leverage an existing sound toolkit called Snack [14]. Snack is an extension of the Tcl/Tk language [15] that includes a Python wrapper. Snack allows for the recording, saving, playing and analysis of audio for rapid development of sound tools. In addition, Snack also provides algorithms for formant analysis of audio.

Formant analysis and normalization

The Snack formant method analyzes an audio stream and provides a list of formant values. The analysis of the formants may be characterized by adjusting input parameters. Types of parameters available to the analysis are:

- Nominal first formant frequency – this value is used to make adjustments to all other formants. This parameter is tied to an assumed vocal tract length and speed of sound.
- Audio sample selection – parameters are available for the start frame, stop frame and sampling rate of the audio.

- Analysis parameters – a linear predictive coding (LPC) algorithm is used to find the formants. Parameters are available to control the specifics of the algorithm i.e. windowing type, windowing size, LPC analysis order, conditioning of the signal prior to analysis, etc.

Vowel Shapes utilizes most of the formant default analysis settings. The length of the samples audio was originally set to the default used by the examples. Experimentation at two and four times this length indicated that the length should be increased by four times. The sampling rate of the audio was originally set to the default value of 16000 Hertz. While this is a good sampling rate for speech, the suggested rate for music is 44100 hertz [4]. This is 2.75 times higher than the default. This may indicate that an increased audio length of three times would be acceptable, potentially reducing analysis time.

The formants are normalized and the Barks Difference Metric is calculated. Updating the vowel shape on each resampling of the formants caused the vowel shape to jump or jitter. The lack of a smooth visual transition of the shape was considered distracting by the singers. To smooth the shape transition a moving average of the Barks Difference Metric is calculated to update the display (Figure 8). Part of our future work will determine the optimal values for other parameters of the audio analysis.

Matching a Vowel

Initial design implementation of Vowel Shapes included a green target vowel and a yellow student practice vowel. The student audio modified the practice vowel in real time as the student adjusted the vocal tract to match the target vowel.

Initial feedback from the singers indicated that the singer was unsure when the practice vowel adequately matched the instructor vowel (Figure 8). A user suggested that the color of the practice vowel change when the instructor vowel was matched. A second suggestion was made to allow for a tolerance of when a vowel was matched. The tolerance could be adjusted based on the experience of the singer or the mastery that was desired (Figure 9).

Both suggestions were implemented. The instructor vowel color was changed to blue. The color of the student practice vowel changes to green when there is an acceptable match. Our implementation for an acceptable match is based on an application start-up parameter that defines a tolerance for the vowel shape-matching algorithm.

Other Design Considerations

In the initial design of the system, the student's current vowel shape was to be displayed alongside the target vowel shape. Interviews with students of singing and with a linguist indicated a preference for the student's current vowel shape and the target vowel shape to be overlaid. Therefore, this design was adopted in all prototypes following the interviews.

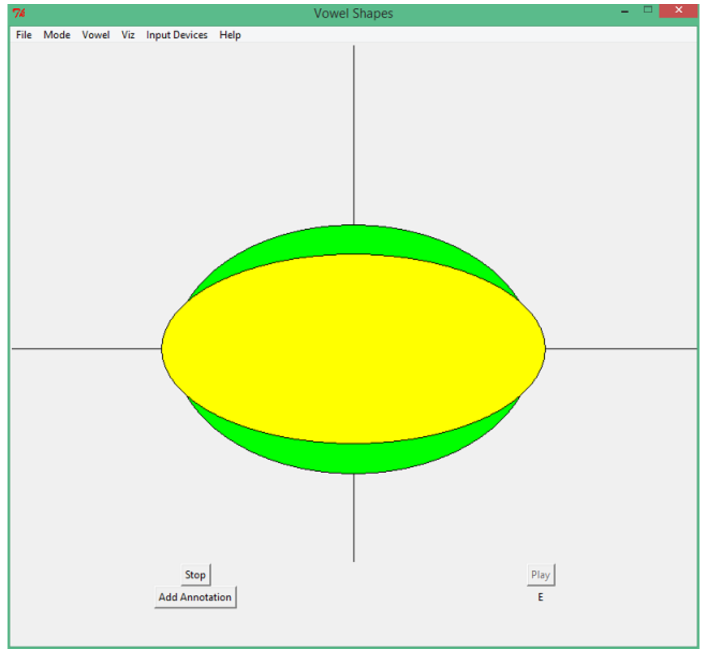

Figure 8: The original representation

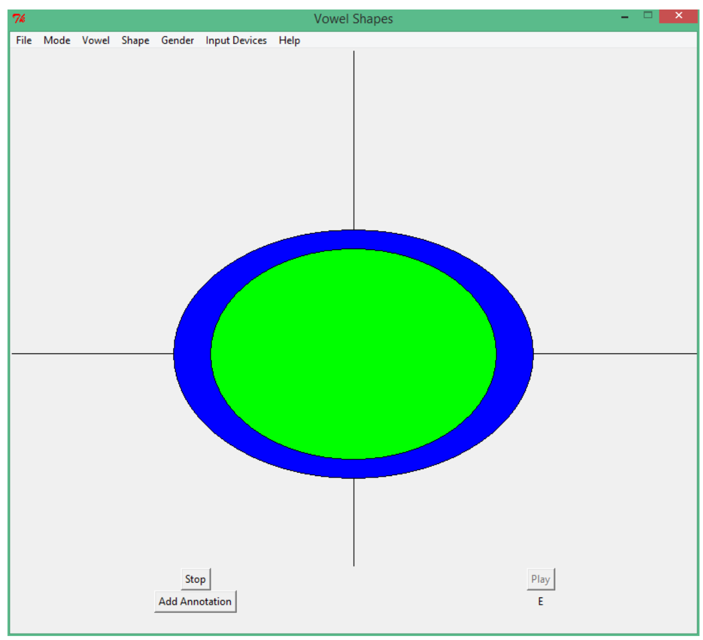

Figure 9: A matching vowel based on tolerances changes color

In addition to the ellipse, triangle, and graph methods of visualizing the vowel, presenting the vowel on a spectrogram was also considered. Interviews revealed that vocal students did not know how to read a spectrogram, so this feature was omitted.

Interviews with a linguist and with a speech prosody expert suggested using a Z-score normalization, a vowel-extrinsic[1] algorithm, for comparing the vowels of different singers. However, a Z-score normalization would require a set of vowels, which makes it impractical for our goal of analyzing, normalizing, and displaying audio data in real time. Therefore we used the Bark Formant Normalization, a vowel-intrinsic[2] algorithm, to normalize individual vowels.

Design iterations.

The initial prototype included vowel shapes with seven vowel formant definitions. No analysis of singing was completed; only the cycling of the vowel shapes through the seven vowels as defined by their formants. This allowed us to demonstrate the vowel shapes to vocal students and the professor. Responses indicated that the students understood why the shapes were changing based on the vowel definition; additionally, the professor and the students felt that this would be a good practice tool.

The second iteration integrated the recording and analysis of singing. Vocal students and the professor were asked to use the tool to record their singing and watch the changing vowel shape. This iteration only provided an interactive vowel shape based on the real time singing; no target vowel shape provided. These sessions allowed the singers to observe how the vowel shape would change based on how they changed their production of the vowel. Students often asked questions during these sessions and the professor would provide insight as to how the vocal tract could be adjusted. There was positive feedback from both the professor and the students that this would be a good practice tool.

The third iteration added saving a vowel definition and loading of the vowel definition. A vowel definition consists of recording information, formant definition for the recording, annotation and gender of the singer. Now the tool could be provided to an instructor, the instructor could save a vowel definition for the student and the student could load the vowel for a practice session. Student and instructor interaction with this iteration provided for some final design changes discussed earlier in the paper. The fourth design iteration provided the first prototype that was used for the full evaluation of the tool.

IN-LESSON EVALUATION DESIGN

The prototype was evaluated during meetings between vocal students and their professor.

Participants

The evaluation group consisted of eleven[3] university vocal students and their professor. The group consisted of five freshman students majoring in vocal performance, one sophomore, one junior, two masters students, and one PhD student. None of our evaluation participants were among the five students interviewed. The professor was present at all evaluations.

Evaluation Parameters

The evaluation compared the time it took a student to match the professor-demonstrated vowel with and without the prototype. The evaluation time without the prototype is thus an approximation of a student practicing with a lesson recording and notes. The presence of the professor allowed him/her to discern when the student had successfully matched the target vowel.

The evaluation consisted of three steps. The first step measured how long it took the student to match a phonetic [i] (ee) at a moderately high pitch, either C5 for females or C4 for males. The second step measured how long it took the student to match the same vowel at a lower pitch, C4 for females and C3 for males. The third step measured how long it took the student to match a diphthong[4] of the professors' choice (Figure 10).

The professor noted that the students were already matching the [i] vowels satisfactorily. Therefore, the diphthongs were used to compare the time it took to match the target vowel with and without the prototype.

The evaluation followed a within participant design. All participants completed a session with and without the prototype. It was also an in-between design, half of the participants completed the session with the prototype first, half of the group with the prototype second. This addressed the issue of ordering effect.

When matching a vowel without the prototype, the professor demonstrated the vowel. When matching a vowel with the prototype, the recording of the vowel was played and the vowel shape was displayed for the student. The student's time to correctly produce the vowel was measured. The professor determined when a correct vowel was obtained in all cases. A follow up questionnaire (available at http://goo.gl/lXcijr) was provided after each session.

---

[1] Vowel-extrinsic algorithms compare formant values over a set of vowels for an individual.

[2] Vowel-intrinsic algorithms all formant normalization parameters are found within a single vowel.

[3] There was an error when timing the Professor matched portion of the session for one student. Only ten students are reported for this result.

[4] A diphthong is a sound formed by the combination of two vowels in a single syllable, in which the sound begins as one vowel and moves toward another (as in coin, loud, and boy).

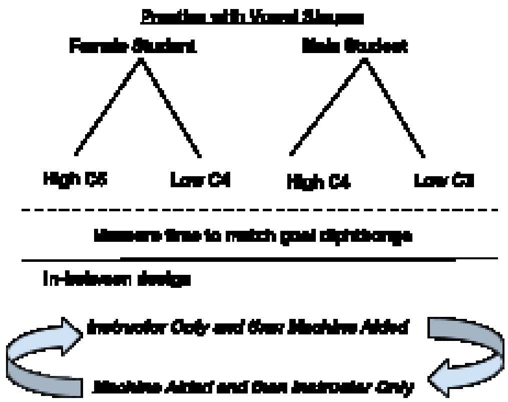

Figure 10: Three steps of the evaluation – two vowels, diphthong, alternate between prototype and instructor.

Results

Using the prototype during a lesson reduced the time required to match the target vowel for seven out of ten students[5]. The following tables report how long it took each participant to match a target vowel with and without Vowel Shapes. Table 1 demonstrates the average time for the seven participants who did improve using the Vowel Shapes tool. Table 2 provides the average time details on the participants who didn't improve.

Based on our results, Vowel Shapes does not favor one experience level over the other (see figure 11). Figure 11 illustrates the average difference per experience level as reported in Table 1 and Table 2.

Only 64% of the participants said they could see the movement in the diphthong from the first vowel to the second vowel.

Student Class Standing	With Vowel Shapes (seconds)	Professor Only (seconds)	Difference (seconds)
Freshman	50.00	60.02	10.02
Freshman	3.00	10.00	7.00
Freshman	8.00	23.00	15.00
Freshman	10.66	28.00	17.34
Sophomore	7.52	27.77	20.25
Junior	12.10	25.43	13.33
PhD	28.00	72.00	44.00
Average improvement with Vowel Shapes			**18.13**

Table 1: Students results that improved with Vowel Shapes

[5] There was an error when timing the professor-matched portion of the session for one student. Only ten students are reported for this result.

Student Class Standing	Professor Only (seconds)	With Vowel Shapes (seconds)	Difference (seconds)
Freshman	22.70	28.00	5.33
Masters Student year 1	15.00	20.00	5.00
Masters Student year 2	9.53	13.55	4.02
Average time better with professor			4.77

Table 2: Student results which were better with the professor

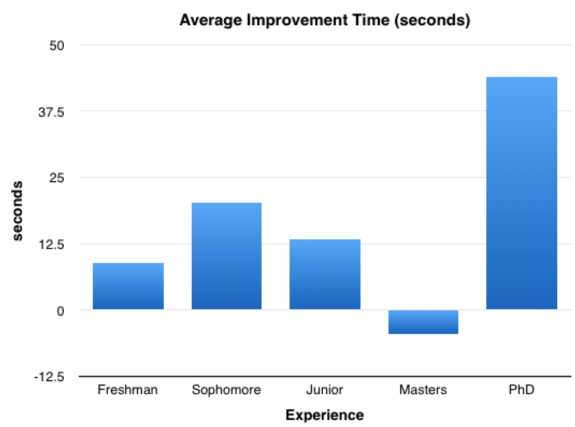

Figure 11: The average time improvement using Vowel Shapes

Results from questionnaire after evaluation

The follow-up questionnaire provided many insights into future work and ways to enhance the learning application. A notable statistic from the follow-up questionnaire was that nine out of eleven students said they would use this tool in their studies. One of the users mentioned,

> *"... it could be useful for beginners to narrow down their spectrum ..."*

When asked if there was a preference for vowel representation 91% of the students preferred the ellipse and 9% preferred the graph. The masters student who preferred the graph said that finding the point was easier than matching the ellipse. He found the ellipse too abstract and distracting. Both graduate students were also more critical of the ellipse than others. However, the undergraduates, specifically the freshman, expressed their satisfaction by the ellipse and its helpfulness.

DISCUSSIONS

Improving the tool for diphthongs may require a more robust analysis of the audio. The first stage would require distinguishing the first vowel of the diphthong. The second stage of the analysis would determine the transition point between the two vowels. The third stage would continue to analyze the audio to determine the second vowel. With a method of analysis that provides both vowel definitions, a diphthong could be represented by two shapes with different colors.

Participants expressed desire for an additional feature that would allow for more than one target vowel to be loaded at once, for instance, a beginning vowel and an end vowel.

Some students stated that there could be starker differences between correct and incorrect vowels. Often the vowel visualization would rapidly alternate between correct (green) and incorrect (yellow), yet there would be no apparent change in the vowel produced. This suggests that the tolerance algorithm requires finer tuning. This would likely be done by working with instructors and students of singing, measuring the tolerances accepted by the instructor.

As it currently stands, the tool is implemented in a way that the vowel shape may completely obscure the target vowel shape. In order to avoid this, the display of the overlaid student shape may be displayed with a transparency factor. Or, when two vowels overlap, a third color could be introduced.

In addition to providing feedback through shapes, it may also be possible supplement the feedback with real time text to the user. However, interviews with students of singing caution that this must be very precise and very accurate else it would render the tool less helpful than if advice was not given.

Another alternative could be to allow the instructors to include additional annotation in the vowel definition. A more long-term solution may be text that suggests some adjustment of the vocal tract to achieve the target vowel. This would require additional training of the tool for each user from their recordings or other analysis.

While our initial prototype works in the Windows environment, it is possible to make it platform independent and even make it work via webpage and mobile phones applications.

Acceptance of this tool could generate a library of vowel definitions. An online community could be built where experienced singers may provide their vowel definitions to be matched by others.

LIMITATIONS

The current prototype has first-generation algorithms for formant analysis, formant normalization and tolerance checking of student vowels to target vowels. Work in all of these areas would improve the potential of Vowel Shapes.

The formant analysis could add parameter inputs that would fine-tune the algorithm to a particular student. The formant analysis LPC algorithm parameters need to be tested for singing audio. As was the case with audio frequency sampling there may be some differences involved with the optimal formant analysis of the singing voice versus the speaking voice.

The formant normalization selected, Barks, is one of many vowel-intrinsic algorithms. More than one normalization algorithm might be implemented and further evaluation completed to determine the optimal method. Evaluation may even determine that algorithms fall into groups – better for Italian vowels, better for females – and this option could be provided for the student.

The tolerance algorithm requires the most work and evaluation. The current algorithms' use of area, X/Y ratios and distance is sufficient for the determining the prototype viability. It will need to be considerably improved and tested so that a student may have confidence in the color change of the vowel shape and the possibility of sustaining a correct vowel shape over a series of pitches.

The initial pilot study allowed for the initial eleven student evaluation sessions. With the encourage results in the pilot, our current goal is to develop the second-generation prototype with the continual involvement with other students and instructors.

CONCLUSION

In this paper, we demonstrate an open source automated and interactive tool that allows singers to practice vowel sounds. The evaluation of our system with 11 participants indicates that more than 70% of the participants could more effectively improve their vowel sounds compared to the traditional methods. The implications of our findings is that students are now able to master more vowels in less time, increasing their overall productivity and rate of mastery. With this interactive technology, students can continue to fine-tune their vowels outside of lessons.

Our future goal with this technology is to port it a web based application making it ubiquitously available via any computer or mobile platforms. The implementation of Vowel Shapes and ensuing evaluation show that aural performance can be learned more rapidly with the aid of an interactive visual system. We believe that our findings provide further implications and insights on the role of interactive systems in other aural disciplines, such as learning a new language, or aid for those with difficulty speaking.

Finally, this application invites interactive technological solutions to other areas of singing, other areas of music, or even other areas of the performing arts.

ACKNOWLEDGMENTS

We would like to thank all the singers and instructors that assisted us in the design and evaluation of Vowel Shapes. Vowel Shapes is freely available for Windows at http://goo.gl/HfSWrv. The open source repository can be viewed at{omitted to maintain anonymity}.

REFERENCES

1. CantOvation. (2012). Singing Software | Voice Training | Vocal Training | Sing & See. Retrieved Jan. 7, 2014 from http://www.singandsee.com/[singandsee.com]
2. Celumbra. (n.d.-a). Listening Labs DictionForSingers. Retrieved from http://www.dictionforsingers.com/listening-labs.html
3. Celumbra. (n.d.-b). DictionForSingers. Retrieved from http://www.dictionforsingers.com/literature.html
4. Colletti, J. (2013). The Science of Sample Rates (When Higher Is Better—And When It Isn't). Trust Me I'm A Scientist. Retrieved from http://trustmeimascientist.com/2013/02/04/the-science-of-sample-rates-when-higher-is-better-and-when-it-isnt/
5. Hall, D. C. (n.d.). Interactive Sagittal Section. Retrieved from http://homes.chass.utoronto.ca/~danhall/phonetics/sammy.html
6. Howard-Baker, T. (n.d.-a). Vocalist - Vocal Software Articles,. Retrieved from http://www.vocalist.org.uk/vocal_software.html
7. Howard-Baker, T. (n.d.-b). Vocalist - Singer - Singing Teachers Resource. Retrieved from http://www.vocalist.org.uk/index.html
8. O'Connor, K. (2014). Singwise - Vowels, Vowel Formants and Vowel Modification. Retrieved from http://www.singwise.com/cgi-bin/main.pl?section=articles&doc=VowelsFormantsAndModifications&page=2
9. Python Programming Language – Official Website. (n.d.). Retrieved from http://www.python.org/[python.org]
10. R. A., B., C. G., H., & J. B., P. (1984). Towards an auditory theory of speaker normalization. Language & Communication, 4(1), 59–69. doi:10.1016/0271-5309(84)90019-3
11. Syndal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of Acoustic Society of America*, 79, 1086. doi:10.1121/1.393381
12. Tcl Developer Site. (n.d.). Retrieved from http://www.tcl.tk/
13. Thomas, E. R., & Kendall, T. (2007). norm1_methods @ ncslaap.lib.ncsu.edu[ncslaap.lib.ncsu.edu].
14. TMH KTH :: Snack Home Page. (n.d.). Retrieved from http://www.speech.kth.se/snack/
15. Traunmuller, H. (1990). Analytical expressions for the tonotopic sensory scale, 97–100.
16. Tversky, B. (2001). Chapter 4: Spatial schemas in depictions. In M. Gattis (Ed.), Spatial Schemas and Abstract Thought (pp. 79-112). Cambridge: MIT Press.
17. UCL. (n.d.). UCL Division of Psychology and Language Sciences. Retrieved from http://www.phon.ucl.ac.uk/resource/software.php
18. UNT. (n.d.). For Singers | Voice. Retrieved from http://music.unt.edu/voice/singers[music.unt.edu