Syllable Circles for Pronunciation Learning and Teaching

Charlie Cullen
*Technological University Dublin*, charlie.cullen@tudublin.ie

Keith Gardiner
*Technological University Dublin*, keith.gardiner@tudublin.ie

John B. Whipple
john.b.whipple@gmail.com

Follow this and additional works at: [https://arrow.tudublin.ie/aaschmedart](https://arrow.tudublin.ie/aaschmedart)

Part of the Communication Commons, and the Education Commons

**Recommended Citation**

This Article is brought to you for free and open access by the School of Media at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License
Syllable Circles for Pronunciation Learning and Teaching

John B. Whipple

Charlie Cullen

Keith Gardiner
Syllable Circles for Pronunciation Learning and Teaching

John Whipple, Charlie Cullen and Keith Gardiner

Syllable Circles is an interactive visualization representing prominence as a feature in short phrases or multi-syllable words. They were designed for Computer Aided Pronunciation Teaching as a part of English Language Teaching. This study explores the question of if and how interactive visualizations can affect English Language Learners’ awareness of prominence, or stress, in English pronunciation. The study followed seven learners and three teachers. Think-aloud protocols, notes from direct observation and interviews of two groups allowed for six streams of data. It was found that interactive visualizations of syllable circles facilitate noticing prominence. Learners and teachers believed interactive visualizations were a useful means for presenting prominence and other suprasegmental features and would be valuable learning and teaching resources.

Changes in Pronunciation Learning Objectives

We are now at a point where most L2 teachers recognize that there is nothing wrong with having an accent, and that intelligibility and comprehensibility should be the goals of L2 speakers, not native-like status. Technology is advancing; there is a real role for... other sorts of practice opportunities, informed by research. (Derwing, 2009)

Intelligibility should be primary goal for English language learners and teachers (Levis 2005; Munro and Derwing 2011). To have a greater impact on intelligibility, research suggests using methods and tools to focus learning on suprasegmentals: features such as rhythm, intonation, syllable stress and sentence-level word stress.

Stress in both sentences and words will be referred to in this study as prominence. Noticing prominence is fundamental to thinking about suprasegmental features. Unfortunately, like other suprasegmental features, prominence is not visible in standard written English and is noted in the International Phonetic Alphabet (IPA) as mere punctuation despite its impact on intelligibility. To help learners understand what makes one speaker more intelligible than another, teachers can address the perception, conceptualization and description of suprasegmentals.

Thoughtful teaching and the use of tools like the IPA aid pronunciation learning. Despite the systematic presentation of the IPA lending to a perceived teachability, phonetic alphabets like the IPA do not, unfortunately, focus visually on suprasegmentals in a way proportionate to their impact on learner intelligibility.
For best effect on intelligibility, learners should be primarily affected by the impact of suprasegmental features (Anderson-Hsieh, et al 1992; Hahn 2004; Levis 2005).

**Noticing Prominence**

The faculties of learners and teachers to notice speech phenomena can have an effect on learner intelligibility (Couper 2006). Hahn's study demonstrates that 'primary stress' or prominence is a worthwhile learning target because of its significant effect on intelligibility (2004). She found that an unusual use or absence of prominence was a key factor in speakers being rated as less intelligible.

Schmidt's *Noticing Hypothesis* posits that any learning must be preceded by either conscious or unconscious noticing (1990). Whether conscious or unconscious, the learner's observation of suprasegmental features makes an impact on their intelligibility.

If prominence is given more attention visually, learners may be enabled to notice prominence. Once noticed, awareness and stronger criteria for acceptable production can facilitate greater intelligibility.

**Computer Aided Pronunciation Teaching**

Munro and Derwing outline how intelligibility-focused pronunciation teaching is progressing and is facilitated by technology (2011). They note the utility of recordings, speech analysis and speech visualization while agreeing that more appropriate learning objectives and tools are needed.

Grantham-O'Brien describes how multimedia technology has been used in descriptive language teaching. She synthesizes a progression of developments to suggest guidelines for future applications of multimedia technology (2011). These echo Levis in his call for pronunciation learning to mean increasing intelligibility (2005). This includes creating tools to give simple feedback and resources designed expressly for language learning.

Their findings also reflect the idea that language learning is not a linear process. DeBot suggests that language learning and acquisition is perhaps best described through Dynamic Systems Theory, implying that language learning needs 'external resources' like 'spatial environments to explore' and 'internal resources' like conceptual knowledge of key concepts and terms (DeBot, et al 2007). However, few external resources dealing with suprasegmentals are available. The teacher is left to address these needs alone to in the absence of tools clearly requested in research and practice.

Attempts have been made to use standing technologies for Computer Assisted Pronunciation Teaching. One example tool is PRAAT, used memorably by Brett
(2004). PRAAT is a freely downloadable, powerful speech analysis and visualization tool. His study clearly described how PRAAT helped learners further understand discrete segmental features. Though somewhat successful, the feedback was difficult to interpret without certain theoretical grounding. Brett reminds readers that PRAAT was designed for speech scientists not language learners and teachers. Its feedback remains complex and perhaps not suited to language learning.

Visualizations of speech samples designed to be clear, to be memorable and to facilitate discussion may help learners do what they do in classrooms already: discover, share, discuss and try out their observations. Multimedia tools designed to facilitate learners noticing prominence should positively affect the learners’ ‘internal abilities’ by giving them an ‘external resource’, a ‘spatial environment’, to explore.

This should facilitate teachers as they help learners to notice, examine, discuss and socially explore features of speech shown to positively affect intelligibility.

**Syllable Circles: Interactive Visualizations to Help Learners Notice Prominence**

Three interfaces or components were designed to demonstrate syllables and prominence visually and interactively for learners.

Component 1 was based on a single audio recording of a short phrase. The recording was divided into syllable-sized audio segments. Each syllable segment was represented visually through circular buttons of different sizes. A click on any button played the syllable it was drawn from in the short phrase. The buttons were aligned on screen in sequence. Larger buttons for more prominent syllables and smaller for less prominent. Circle size related to three factors:

1. Its duration;
2. Its pitch change;
3. Its volume.

The placement of the syllable circles matched a computer-generated waveform of the recording and text. The final interface appeared as below and could be enlarged to fill a full screen:
Figure 1 Component 1: Mouse over second syllable circle

Each syllable circle button functioned interactively to avoid explicit instructions. When the mouse was rolled over the circle, the circle filled and the corresponding letters of the word or phrase appeared above it (see Figure 1).

The offset button (upper left corner) played the entire phrase and cued a playhead to move horizontally across the three vertically aligned elements: the syllable circles, the waveform and the text. The movement of the playhead matched the progress of the voice through the sample. It associated the elements for the user and reduced the need for identifiers or explanations (see Figure 2).

Component 2 and Component 1 were both created in Adobe Flash. Component 2 was designed to facilitate comparison between instances of the same phrase spoken by the same speaker. It demonstrates decreasing formality and
increasing speed. Some irregular spellings were used to demonstrate and emphasize the elision of some sounds. See Figure 3 below.

Figure 3. Component 2: Multiple Samples allow the user to compare different instances of the same phrase spoken by the same user.

Component 3 was created in partnership with the Digital Media Centre in Dublin Institute of Technology. It created syllable circles automatically through a program using a speech analysis framework adjusted to detect syllables based on the presence of vowel sound indicators. Once detected, the framework sends data to an interface, automatically rendering interactive syllable circle visualizations.

Fifty phrases were pre-recorded to be available for the study. The same interactive syllable circle concept featured, with size indicating prominence and each circle played its respective syllable-sized audio segment when clicked. The circles filled with a new color to indicate they were being sounded. They were positioned sequentially, but prominent circles were placed higher than less prominent ones. See below:
Component 3: A full phrase being played with the third syllable sounding

Component 3: Displaying 'Sorry... how much?' Note the sample phrase menu for the next phrase to be displayed and 'Record' and 'Play' buttons for the lower display space.

Component 3 some presented drawbacks and benefits. It did not reliably recognize all syllables in the phrase and frequently failed to recognize the final syllable in some longer phrases. It was missing three significant visual elements: 1) the text, 2) the waveform images and 3) the playhead. Notable advantages included the number of samples available; the speed of creation; the potential for the learner to record, analyse and compare their own samples.

Presented with Components 1 and 2, the set was deemed to present a robust sample of the syllable circle concept as a worthwhile intervention for a case study investigating how learners may be affected.
Case Study

A qualitative case study was designed to determine if and how users feel they were affected by interactive visualizations of syllable circles illustrating prominence. As an educational case study the research relied on the views of the participants (Cresswell 2008). Its findings should be combined with relevant findings from similar studies of Computer Aided Pronunciation Teaching to lead to ‘fuzzy’ generalizations about educational practice as suggested by Bassey (2006).

Seven learners and three teachers explored the components and were subsequently interviewed to determine its potential for pronunciation learning.

The study was carried out in a self-access environment in a private English language school.

Ethical procedures for university study were followed. Permission and informed consent were obtained from the school and the participants individually.

The seven learner-participants were between 18-35 and had an English proficiency level of B1 on the Common European Framework of Reference attested by their school placement procedures. The three teacher-participants had all been teaching for four years or more. Each had achieved or was studying for a higher qualification in English Language Teaching.

Step 1: Each participant had a 10-minute introduction to the three components.

Step 2: The researcher audio recorded 20-minute ‘think-aloud’ exploration for transcription providing the first data stream. Each participant engaged with the three components using the ‘think-aloud’ protocol. They were invited to spend as much of their time with any of the components as they wished and to move freely amongst the them. The researcher avoided interference but answered questions when directed at him and reminded the user of the think-aloud protocol if they began working silently or used only one component.

The researcher took notes throughout providing a second data stream, observation notes.

Step 3: The researcher held a 20-minute semi-structured interview with each following Step 2. These interviews were recorded and transcribed, providing the third data stream. The questions were open-ended to seek descriptions of the users’ experiences. Questions for the teacher interviews addressed teaching and learning practice (see Appendix 1). Accordingly teacher interview data was separated from learner data.

The researcher continued taking notes through this experience adding to the observation notes.

Research with each participant lasted approximately 60 minutes. Three data streams resulted for each participant: see table below.
Transcripts from the participants were anonymized, coded and themed to reveal data leading to findings of the research question.

The research question was as follows:

*Do interactive visualizations of speech samples help learners notice prominence as a feature of English speech? If so, how?*

**Data Analysis**

Six data streams were collected from the study. The purpose, form and a description of the analysis of these streams are illustrated below in tabled examples.

### Purposes, Form and Analysis of Think-Aloud Transcripts

Learner-Participants (Px) and Teacher-Participants (Tx) used a think-aloud protocol as they explored the components to render insights regarding the individual cognitive processes of each participant and give some evidence about how the user interacted with the components. Audio recordings were made.

Each was transcribed, the statements within were coded and finally and finally the codes revealed themes as in the samples below:

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Codes</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>P5:</td>
<td>Repetition in Observation; Observation-Hypothesis-Experiment; Learner engagement</td>
<td>Value of recordings</td>
</tr>
<tr>
<td>[Repeats short recordings 10x] The same? Hmm. [Repeats short recordings 15x] Ok. Ok, this one? Oh, 7?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P6:</td>
<td>Perceived possible learner benefits; Forecast prominence patterns; Hypothesis statement</td>
<td>Perceived Benefits</td>
</tr>
<tr>
<td>And now I know the tense (P6’s word for prominence)… So the tense is on ‘can’t’ and the ‘lie’ of ‘believe’ and ‘it’s not butter’. And do it again. (clicks) Recording: I CAN’T BELIEVE IT’S NOT BUTTER.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L6:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can’t believe it’s not butter. And I’ve got the rhythm of the sentence…</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Participants</th>
<th>Activity</th>
<th>Time</th>
<th>Data Stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td>Think-Aloud (Entire Experience)</td>
<td>20-30 Minutes</td>
<td>Recording Transcripts</td>
</tr>
<tr>
<td></td>
<td>Semi-Structured Interview</td>
<td>Continuous</td>
<td>Researcher Notes</td>
</tr>
<tr>
<td>Teachers</td>
<td>Think-Aloud (Whole Experience)</td>
<td>20-30 Minutes</td>
<td>Recording Transcripts</td>
</tr>
<tr>
<td></td>
<td>Semi-Structured Interview</td>
<td>Continuous</td>
<td>Researcher Notes</td>
</tr>
</tbody>
</table>

**Table 1 Summary of Data Collection**
P3: It’s very short like when I speak with my friends. And we can hear, but if I think if I listen, I don’t understand, it’s impossible. It’s too fast for me.

T3: (The concept of prominence) is there. It’s there for them.

<table>
<thead>
<tr>
<th>Table 2: Samples of Think-Aloud Transcripts with Codes and Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1: The circles show me how each word you have to say- you have to enunciate- each word because (...) they are together and some of them they are not together and they show how…</td>
</tr>
<tr>
<td>Codes: Circle size as prominence guide; Linking issues; Written vs Spoken English; Learner experience shaping concepts;</td>
</tr>
<tr>
<td>Theme: Concepts noticed</td>
</tr>
<tr>
<td>P4: I think it’s more about speaking… because I don’t really learn about the rules for example, for a word with two syllables, three. So, it’s really</td>
</tr>
<tr>
<td>Codes: Importance of listening; Unconscious learning; Lack of guidance/instruction; Need for speaking work</td>
</tr>
<tr>
<td>Theme: Learning difficulties</td>
</tr>
</tbody>
</table>

Purposes, Form and Analysis of Observation Notes
The initial coding process suggested a great variety of instance specific codes. Readings and re-readings led to an increased familiarity with the data. Gradually certain codes appeared more frequently with instance specific codes being revisited to see if they matched codes that had emerged after initial readings or to see if they shaped the developing themes. Overlap indicated themes. These led to evidence for the findings.

Purposes, Form and Analysis of Semi-Structured Interviews
Following the observations each learner and teacher was administered a semi-structured interview lasting approximately 20 minutes. Five questions were asked.

The purpose of the semi-structured interviews was to allow the users opportunities develop a descriptive discussion of their views and experience. The questions allowed the opportunity to discuss the interface as a learning aid, comparing its content and utility to previous learning experiences.

After coding all of the content from the interview transcripts, fifteen themes were revealed including the following included as examples:
by listening to people that you can just pick up this kind of thing.

P5: If we realise that’s the problem, we can forecast this one. So I recommend to record fast and compare and fix and find the faster problem...you know what I mean?... If we know what’s the difference, we can forecast more... the stress.

Critical thinking; Perceived possible learner benefits; Forecast prominence patterns; Hypothesis statement; Own preferences

Perceived Benefits

Table 2: Samples of Learner Interview Transcripts with Codes and Themes

Teachers were asked a separate set of questions to inform as experts on the learner’s process and the utility of the components as teaching tools. The teacher data revealed numerous codes from which three major themes emerged. These were as follows and are reflected in the Findings:

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Code</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: I think teaching sentence stress for people to give... certainly for visual learners it’d be very good... ’cause eh... it’s sometimes very difficult even just with underlining things on the board...</td>
<td>Design approval; Learner preferences; Teaching challenges for pronunciation teaching;</td>
<td>Teaching and learning issues around pronunciation</td>
</tr>
<tr>
<td>T3: I’ve never even considered visualizing it, and I don’t think my students would... It’s fantastic.</td>
<td>Teacher/learners wouldn’t visualize; Design approval; Demonstrate prominence</td>
<td>Benefits and potential for teaching and learning</td>
</tr>
<tr>
<td>T3: Let’s say I was doing a reading or there’s a potentially blocking piece of vocabulary and I could type in... and suddenly you have all of these images</td>
<td>Overlap for utility; Teacher ideas</td>
<td>Creative and critical engagement</td>
</tr>
</tbody>
</table>

Table 3: Significant Teacher Interview Themes

Findings

Primary finding

The purpose of this investigation was to determine if and how interactive visualizations of speech samples help English language learners notice prominence. The data strongly suggests that interactive visualizations of syllables do.

They do so memorably by facilitating observation, hypothesizing and experimentation while enabling them to notice other speech phenomena.

Secondary findings from learner data

Data analysis leads the researcher to suggest that use of the components allowed learners:
1. To recognize prominence independently through comparative size and interactivity;
2. To compare expectations, the sample and own production;
3. To conceptualize a shared, persistent, visual symbol of a significant, invisible phenomena;
4. To notice linking and elision.

The use of the components also:
5. Provides conceptual clarity for the learner;
6. Makes the concept memorable visually;
7. Provides the learner a welcome laboratory for testing and demonstrating observations;
8. Enables an Observe-Hypothesize-Experiment cycle promoting various discoveries and questions that can be used as starting points for various lessons.

**Secondary findings from teacher data**

Teachers confirmed learner views based on their previous experience using Computer Aided Pronunciation Teaching and their work with the components. They found that the components
1. Provide a useful way to illustrate prominence;
2. Enhance learners’ abilities to compare speech phenomena;
3. Facilitate learners’ efforts in noticing and conceptualizing syllable stress, linking phenomena, weak forms and typical phonological features of lexical chunks.

Teachers believe learners want to improve their pronunciation. They confirm that it remains difficult to suggest self-study Computer Aided Pronunciation Teaching resources. Explicit pronunciation teaching is currently central to raising learner pronunciation awareness.

**Conclusions and possibilities from teacher data**

**Teachers**

Teachers demonstrated engagement with the components through praise, criticism, ideas and requests for further samples. Below are some quotes from the teachers:

‘If you had something like this (a similar affordance)... before class, you could analyse the stress, the intonation... that would make you think more deeply about incorporating pronunciation.’

‘This would be quite good... to demonstrate clearly... how we reduce the number of syllables in each of the sentences. I think this is good. Very, very clear.’

‘This is a demonstrative application. You’re presenting... It helps them notice.’

‘I could have used this in my class this morning.’
Learners

The learner-participants displayed additional interest in noticing and recognizing other taught pronunciation phenomena in addition to prominence. They replayed syllable-level recordings a surprising number of times and seemed particularly struck noticing linking phenomena. None of the learners reported having access to this type of functionality previously.

In the semi-structured interviews, participants explained their thinking and learning frequently with the use the interface itself as a tool to illustrate their observations:

(W)hen people also speak very fast... it's different...it's (a) different sentence. First, I think (the slowly spoken sample is) for me: 'I will go and find them'.

(Learner clicks) HE WILL GO AND FIND THEM.

So, it's very slowly with a right word. ...The second sentence, it's a short sentence with the verb ‘will’ is not written, just a double ‘L’. ‘I’ll go an’ find ‘em’.

(Learner clicks) HE’LL GO AND FIN’ ‘EM.

And the verb is very, very short. I think it's like an expression when English people speaks very fast with his friends. I don’t know how in the conversation, it’s like made in French. ...it’s not my real language, so the words are different from my language and it’s very short like when I speak with my friends. And we can hear, but if I think- if I listen, I don’t understand, it’s impossible. It’s too fast for me.

The level of control afforded to the learner by the interface presented an opportunity to engage in an Observe-Hypothesize-Experiment cycle as suggested by Lewis (2008). The learner naturally and spontaneously listens, hypothesizes internally and experiments in speech. This cycle was observed with each learner.

The learner highlights one problem which constantly affects all language learners: the need to examine what is being said while simultaneously gleaning lessons about how to pronounce phrases intelligibly and deal with meaning and culture. Learners need space and time to create their internal resources and conceptual knowledge.

Perhaps the most welcome features were the degree of control over the speech sample, its visual navigability and persistence on screen.

Another advantage seemed to be the empowerment to share and demonstrate their observations because of the sample’s navigability and persistence.

Comparison, the social act of description and discussion about these observations, reflection contribute immensely to the learning process and need fostering. An example from Learner 1 demonstrates how the interface enabled the learner to observe. They goes on hypothesize and share their findings regarding. The learner used Component 2 actively in conversation to illustrate their findings.
Eh, because the circle are, I don't know, something like locked or... Yeah, and what does that mean to you? It mean there is a liaison, I have to contract them. Go-an'. Not say 'go and', say 'go-an'.

Good.

Emm.

(Learner clicks) HE'LL GO AN' FIN' 'EM.

He'll go an' fi' 'em. But there is a difference of pronunciation if I go to the first sentence.

(Learner clicks) HE WILL GO AND FIND THEM.

'He will go and find them.' So, maybe when you speak with your friend, you say: 'He'll go and fin' 'em'. And of course, if you want to have a good pronunciation, it's better to say: 'He will go and find them'.

The learner used the component spontaneously to describe a precise finding regarding pronunciation much as teachers use examples in text to illustrate grammar and lexis. This was observed very frequently.

Cautions and Caveats
A warning should also be taken: in the example above the learner says 'have to'. Do they regard the recorded samples as ideal examples? They are not ideal. They are samples. Discussion should be fostered for learning and assistance is needed in analysis. There is a need for guidance on how to interpret the significance of the sample: perfect or passable, target or topic of discussion? Language learning exists in a social context. This tool does not provide the all the guidance needed. Tools are not teachers, but tools can help teachers and learners to observe their language more clearly.

A halo effect may have been created by the presence of the researcher or the engagement in the think-aloud protocol. The attention and questions of the researcher, as well as the think-aloud protocol, should be noted as much as the persistence and navigability of the design as learning aids. However, the technology affords the learner notable new faculties.

Recommendations for development
This case study had the limited goal of examining how using this intervention might help learners notice prominence in order to affect intelligibility. Though designed for this specific purpose the ancillary benefits and themes revealed show potential for further exploration. As teachers and learners work to develop ways to explore their language it is hoped that Syllable Circles may be considered as an example for future Computer Aided Pronunciation Teaching design.

Appendix 1
Learner Interview Questions
1. Have you ever used worked on your pronunciation? (Please describe.)
2. What did the circles mean, show or represent to you?
3. What did you notice while you were working with the application?
4. Can you describe any changes in how you think about how we use syllables in English?
5. What kind of questions do you have now?

Teacher Interview Questions
1. Could you briefly describe how pronunciation teaching fits into English Language Teaching today?
2. Learning situations are very different: one-to-one, mixed language groups, single language groups: How should self-study pronunciation work be approached in ELT?
3. What should be the main focus of activity in individual learner pronunciation work in your opinion? What tools do they use?
4. What kind of tools can be used to help students in become more aware of their pronunciation? How could applications like this be made helpful?
5. What are your most successful learners doing about their pronunciation?

References
