

ENGLISHVILLE: A MULTI-SENSORIAL TOOL FOR PROSODY

KIZZI EDENSOR COSTILLE

University of Caen Normandie, France

kizzi.edensor-costille@unicaen.fr

Abstract

Research has shown that prosody plays an important role in the intelligibility, comprehensibility and accentedness of non-native discourse (Munro and Derwing, 1995, 1998). Yet prosody is deemed difficult to teach (Setter et al., 2010). Previous studies have used software such as PRAAT (Setter et al., 2010, Olson, 2014, Imber et al., 2017,) but they can be complex to use (Setter et al., 2010; Setter and Jenkins, 2005). Could a more comprehensive tool be useful to L2 learners? Englishville is a website where it is possible for the learner to see a real-time 3D spectrogram.

An experiment was set up to determine whether multi-sensorial input, available via Englishville can help learners of English. Eight French students enrolled in a BA in English took part in this initial trial experiment (2 per group). They were divided into four groups. The corpus is divided into 2 parts. The first focuses on lexical word stress (72 words) and the second on intonation in 30 short sentences. The corpus was recorded by a female native British speaker. All participants had one trial at the beginning of the experiment to familiarise themselves with the tool and they all read and recorded the words and phrases as they appeared on the screen. The first group only had access to this text (no input) before recording their own productions whereas the other 3 groups received supplementary input. Group 2 read the text and heard the corresponding audio (audio input), group 3 read the text and saw the corresponding 3D spectrogram (visual input) and group 4 read the text, heard the audio and saw the corresponding 3D spectrogram (multi-sensorial input).

An auditive analysis leads us to believe that both hearing speech and seeing the corresponding spectrogram is beneficial, especially for intonation. Positive results came from the students' feedback; they generally found the tool useful, easy to use, fun and interesting.

Key words: prosody, L2 learners of English, multi-sensorial input, Englishville

1. Introduction

For a long time, teaching and research in second language acquisition has focused on auditory sources of input for both training sessions and experiments. More recently, linguists have started to use multi-sensorial modalities, looking for ways to enhance L2 learning.

For segmentals, teachers have found ways to render the theoretical aspects of phonemes more comprehensible by using multi-sensorial techniques - be it their own mouth or images of the mouth and tongue (sagittal sections). Learners can

simultaneously see the different positions of mouth, lips, and tongue in order to correctly form phonemes whilst simultaneously hearing them. The representation of intonation or lexical stress is more abstract and tools such as Praat (Boersma and Weenink, 2001), however helpful for research, remain difficult to use for the uninitiated (Setter et al., 2010; Setter, and Jenkins, 2005).

We believe that acquiring or improving prosody is complex partly because of the lack of physical or visual aids. It could even be argued that its abstractness dissuades teachers from teaching it and many teachers and learners find perceiving whether the voice rises or falls difficult. With this in mind, we started to look for more apprehensible solution. This led us to create Englishville – a webpage dedicated to practising prosody by means of a 3D spectrogram. Before implementing this tool into a university course, we wanted to test the tool by setting up a micro trial experiment to see how effective the spectrogram function is in the case of L2 prosody, to test whether it was easy to use and corresponds to the learners' needs, but also to obtain feedback on the participant's perception of the tool. We believe that a multi-sensorial tool can greatly help learners practice their prosody, but what do students think?

The main aim of this chapter is to present Englishville and the thought process behind it, from both its conception and implementation. The overall aim of this research is to contribute to the teaching of prosody to L2 learners by providing an easier, less daunting tool.

Despite a clear increase in research on prosody, it is still less studied than other areas of linguistics (Derwing and Munro, 2005). We shall therefore start with a brief overview of prosody and research on L2 learners with a special focus on the prosodic elements in this trial experiment and the difficulties French learners of English face in terms of pronunciation. Secondly, we present research on auditory and multi-sensorial tools in order to highlight the existing advances in this domain and the functionalities of the different available software. For example, not all of the software includes the possibility of seeing speech in real-time, which strikes us as a key element in perceiving and practising prosodic features such as intonation. It is from this analysis that Englishville was created in the attempt to improve on the existing software. Part three explains in detail the methodology used and this is followed by the experiment and results. Lastly, we discuss our findings in the attempt to have a clearer overview of the useability and usefulness of Englishville from a learner's point of view.

1.1. Research on prosody

Linguists have approached prosody from several points of view; therefore, a precise and concise definition is complex. The definition of prosody - in its broadest sense – is that it is specific to spoken language. Prosody, also known as suprasegmentals (to be understood as all that is not segmental) include for example: rhythm, intonation, stress, intensity, and pauses. The two main elements

of prosody on which we shall focus here are lexical stress (intensity) and intonation (pitch). Through experience, these two elements are often difficult for learners of English to acquire.

1.2. Lexical stress

Hirst (1998) notes several linguists who have dealt with English word stress rules (cf. Kingdon 1958; Chomsky and Halle 1968; Guierre 1979; Hayes 1984; Fudge 1984; Halle and Vergnaud 1986). In France, the teaching of lexical stress is based on the rules taken from Guierre (1979) who systematized and theorized accentual regularities based on phonographematics. This work was later taken up by others, notably Ginesy (1998) whose books are widely used to teach and practice lexical stress in French universities. An example for the rule governing the -ic ending is:

-ic(s) → /10/. The accent is placed on the vowel that precedes the suffix (ic) e.g.: /fæn 'tæst ik (Ginésy, 1998:14).

The students learn the rules and exceptions with the aim of implementing them into their speech through exercises.

1.3. Intonation

A great amount of literature exists on British intonation: Halliday (2008), Tench (1996), Hirst, (1998), Wells (2007). It is often considered difficult to perceive pitch movement for learners and even for teachers. Therefore, the main objective of Englishville and this experiment is to help learners see the different melodic patterns. As this chapter deals with the acquisition of intonation by L2 learners, we start with a concise summary of the patterns used in this study. The intonation patterns used in the experiment are simple. For example, it is generally acknowledged that the use of a falling tone on the nucleus indicates finality, and that a rising tone indicates non-finality (Wells, 2007). Wh-questions (closed questions) are normally said with a falling tone on the last lexical item whereas Yes-No questions (open questions) are normally uttered with a rising tone on the nucleus. As for intonation in lists, Wells (2007:75) differentiates between 2 types: those that are finished and those that are not. To illustrate this point, he gives two versions of the same utterance:

- (1) You can have / coffee | or \ tea.
- (2) You can have / coffee | or \ tea”.

According to Wells (2007:75),

the fall on tea in (1) signals that there are no more options: you must choose either tea or coffee. The rise on tea in (2) signals that there may be other possibilities too, as yet unmentioned e.g., or you could have an \nwarrow orange juice.

It is known that in terms of lists, a rise is used on each item except for the last, which is said with a fall. The structure of the sentences used in the intonation task of our experiment correspond to the two main intonation patterns (fall, rise) and (1) in the case of a closed list.

1.4. L2 prosody

Hirst and Di Cristo (1998: 2) explain that

The prosodic characteristics of a language are not only probably the first phonetic features acquired by a child (Kaplan 1970; Crystal 1973; Lieberman 1986; Levitt 1993), but also the last to be lost either through aphasia (Caplan 1987) or during the acquisition of another language or dialect (Cruz-Ferreira 1984; Touati 1990).

The fact that we learn the prosody of our L1 in the very early stages of development makes it difficult to learn later in an L2. Previous studies have confirmed that acquiring prosody is challenging even for advanced learners (c.f. Colantoni et al., 2014:15). However, it is generally accepted that using incorrect intonation patterns, i.e., one that differs from native productions, can lead native listeners' to either misinterpret the intended meaning or show negativity towards the L2 speaker.

For Jenkins (2000), the most important suprasegmental items are: contrastive stress, the direction of pitch movements (to show attitude or grammatical meaning), word stress placement, and stress-timed rhythm. Other authors share a similar point of view on prosody. Munro and Derwing (1995, 1998); Hardison (2004, 2010), have shown that prosody plays an important role in L2 speech. Learners who received instruction on prosodic features (intonation, rhythm, word stress, and sentence stress) showed significant improvement in comprehensibility and accentedness compared with those who had only received instruction on segments. In a study by Derwing and Rossiter (2003), L2 fluency and comprehensibility significantly improved after a 12-week instruction period on the pronunciation of prosodic features. Although in their conclusion they “do not advocate eliminating segment-based instruction altogether”, they do add that:

If the goal of pronunciation teaching is to help students become more understandable, then this study suggests that it should include a stronger emphasis on prosody. (Derwing and Rossiter, 2003: 14)

1.5. Common pronunciation difficulties for French learners of English

French L2 learners of English, like many others L2 learners, find it hard to pronounce certain segments such as the distinction between /i:/ and /ɪ/, English diphthongs and for some, the realisation of /θ/ and /ð/. In our experience, the most problematic phoneme for French L2 learners of English is /ə/. This is most noticeable in both unstressed syllables and in sentence stress. Wells (2007:12) notes that French learners of English have difficulties with tonicity i.e., stressed syllables/pitch prominence (for a full explanation see Wells, 2007:6). Despite some debate, many linguists concur that on the one hand, French is more of a syllable timed language and on the other hand, English is more of a stress timed language. That means that French uses mainly full quality vowels and syllables, and English uses a mixture of strong and weak vowels and syllables – an aspect that often causes problems for learners. The tendency is to use the L1 stress pattern from French - which puts stress on the last syllable in a word and the last stressed item in an utterance – or rather they are lengthened (see Vaissière (1991), for a full discussion on accentuation and final lengthening in French). In lexical words, L2 French learners tend to use full vowels in all syllables and to shift lexical stress. For example: important (/im'pɔ:tɑ̃t/) is realised /impo'tent/, foreign ('fɔ:rən/) becomes /fo'rein/ and wonderful ('wʌndəfəl/'wʌndəfol/) as /wʌndə'fu:l/. The same issues occur in sentence stress, leading to the non-reduction of grammar words e.g., because /bi'køz/ is said /bi'koz/. As for nuclei, they are generally moved to the end of the tone unit, even if the focus or new information is on another item or if the tone unit ends with a grammatical word – for example “she lives here, does she?” the final ‘she’ (and sometimes the initial one as well) is shown as the nucleus (if the exercise is written) and transcribed or pronounced /ʃi:/, giving the impression that it is the nucleus.

Regarding intonation, French L2 speakers of English tend to struggle most with falling intonation and use rising intonation for all types of statements. This is clearly noticeable when a speaker concludes an oral presentation using a rise for the final sentence. The use of this melody can leave the listener frustrated and/or surprised when they realise that the presentation is in fact, finished (effect of non-finality of rising intonation). Some French learners of English use this tune due to the influence of American English (which uses more rises than British English), but for most, the influence comes from their L1. Perceiving stress - be it on the lexical or sentence level - is also challenging for French L2 learners. Regarding lexical stress, they are able to perceive which syllable is stressed with practice, but the realization of /ə/ - or weak vowels in general - remains problematic, even for advanced students. As highlighted in the previous transcriptions, weak vowels tend to be replaced with full ones which can sometimes impede comprehension.

We now turn to our discussion on the use of multi-media tools and how they can improve pronunciation.

2. Research on visual and auditory tools

Seeing speech is surprisingly not that new. The kymograph (or wave writer) dates back to 1847. From then onwards, the different devices used to write down speech flourished, all the way to the sonograph and spectrograms. Following the rapid progress in computer science and technology several software became available, for example Visi-pitch (1975); Speech Viewer (1985); WinPitch (Martin, 1996); and Praat (Boersma and Weenink, 2001) (c.f. Chafcouloff, (2004), for further discussions).

Major advances in speech technology have led to the increasing use of language software and technology such as computer assisted language learning (CALL) and computer assisted pronunciation training (CAPT). A few examples of these are: WASP (<https://www.speechandhearing.net/laboratory/wasp/>), Better Accent Tutor (<http://www.betteraccent.com/>), Ampitch (<https://www.speechandhearing.net/laboratory/ampitch/>). These examples are particularly interesting in regard to L2 prosody because they enable the speaker to see their intonation patterns, as can be seen in figures 1 – 3. In order to compare the different technologies and how they represent intonation, the same sentence was recorded using each of these tools. The sentence “He said what?” was said with rising intonation.

The first software is the Better Accent Tutor (Figure 1) which analyses the L2 speaker’s intonation, stress, and rhythm. It enables learners to visualise these patterns in a simple manner and offers corrective feedback. However, it is only available in American English, and it is not free.

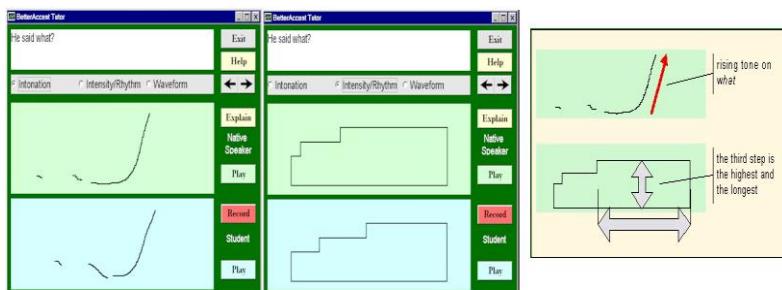


Figure 1: The representation of intonation, intensity/rhythm, and explanation of the utterance: “He said what?” in the free demo of Better Accent Tutor.

In the demo version, different representations of pitch and intensity appear on the screen for the learner to imitate and record. In the corrective phase, an explanation is given regarding the place of the nucleus, the type of tone used on the nucleus and an indication that the nucleus (third step) is also the highest and the longest of the 3 words.

AmPitch (Figure 2) is particularly noteworthy because it displays pitch in real-time. It is not possible to record one's voice or to compare it to another voice, but it is a good reference to give to L2 learners to practice prosody and *see* what their voice is doing.



Figure 2: A real-time display of pitch and amplitude in AmPitch of the utterance: “He said what?”

In this example, it is clear to see that there are 3 syllables/words, and that a rising pattern is used. The students can also note that the length of the segments vary, the last one being the longest and the one that rises (which of course corresponds to the nucleus – a fact they may or may not know). With Ampitch, L2 learners can see the different durations and observe their speech as they are speaking.

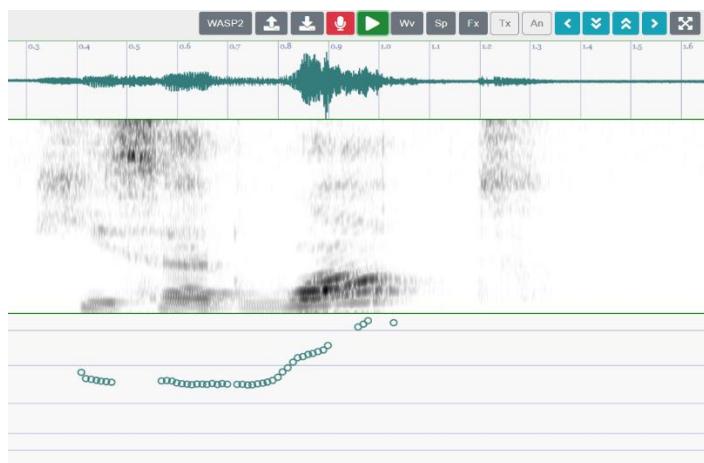


Figure 3: The speech wave, spectrogram, and pitch for: “He said what?” in the online version of WASP.

Ampitch and WASP both have a Web version and can both be found (among many other things) on the Speech and Hearing website or directly on Mark Huckvale’s research page (<http://markhuckvale.com/#research>). WASP stands for Waveform, Annotations, Spectrogram and Pitch (Figure 3) and was recently updated to enhance its usability during the pandemic, when all classes had to go

on-line. It is now possible to record, save, replay, and upload audio which is then displayed. It is also possible to annotate speech, therefore making WASP the closest – visually speaking - to Praat. In the example below of “he said what” the speech wave, spectrogram and pitch contour are visible. Contrary to Ampitch, WASP does not show speech in real-time.

The last example is of our website Englishville (Figure 4) where it is possible to hear the audio, see the spectrogram and the text to read for each item. All these factors can be changed if needed for future research. It is also possible to see one’s own spectrogram in real-time thus enabling the learner to visually compare their own production to a model to see how it differs.

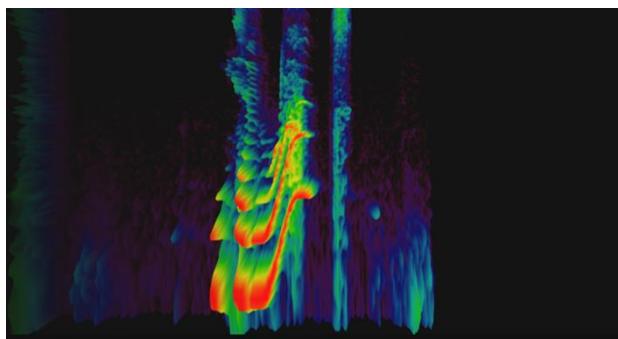


Figure 4: A real-time display of pitch contour and intensity in Englishville of the utterance: “He said what?”

Many software and Apps focus on pronunciation training at the segmental level. The first studies to use multi-sensorial tools focused on segments but were rapidly followed by research on prosody. For segmentals, the results are generally positive. For example, Offerman, and Olson, (2016) provided visual feedback on voice onset time (VOT) to L2 learners and their production of VOT improved. Even more noteworthy, training with these visual aids also had a beneficial effect on learners’ continuous and spontaneous speech.

Since the mid-1970s, there has been an on-going stream of studies which have used computer-based methods in order to test and improve the perception and production of prosody (James, 1976; De Bot, 1983). One of the first studies carried out in the early 1980s by de Bot (1983) concluded that visual feedback was more effective than auditory feedback – in other words, when the subject saw speech (in this experiment, the pitch contour in Praat was used) rather than just hearing it, the subject’s intonation improved. Pitch visualisers (such as Praat and similar software) have been used in more recent research (Setter et al., 2010, Olson, 2014, Kartushina, et al. (2015) Imber et al., 2017.). Gorjian et al. (2013), compared 2 methods of teaching stress and intonation: a traditional one that uses repetition and explanations about acoustic properties of speech, the second, using Praat. The results showed that learning prosody with Praat was significantly beneficial. As

the authors point out, the first method is generally teacher-centred, leaving students passive in the classroom. With the use of multi-media tools and multi-sensorial software, it places the student at the centre of their acquisition.

Some software, like Praat, require practice and training to use and some research concluded that combining sound and image led to slightly more mixed results in learning prosody, often due to the complexity of the software used (Setter et al., 2010). Due to the technical side of these tools, it can be difficult to motivate students to familiarise themselves with this kind of software and then work on intonation. We believe that using real-time displays of intonation is more comprehensible. Students can see their speech appear as they speak enabling them to test different intonation patterns more easily. We believe that the instantaneous effect of seeing speech makes intonation easier to perceive because not everyone can perceive it simply through their ears, some people need their eyes to validate or invalidate their perception. To conclude, depending on the pronunciation feature practised, the use of multi-sensorial tools and methods have been found to be helpful, even having a global positive effect on L2 speech production in general.

3. Method: From conception to implementation

The desire to create a free, user-friendly real-time tool for prosody in the domain of multi-sensorial learning of L2 English, and the idea that prosody should be at the centre of second language teaching motivated the creation of our website Englishville.

Englishville uses a 3D spectrogram and allows the capture of the audio stream so that it can be recorded on a server. These tools are then integrated in a website where it is possible to record corpus, set up experiments and also participate in them. The spectrogram representations are not recorded, they are always interpreted on the go. The sounds are saved in an Ogg format to minimize bandwidth usage but with a high enough bit rate so that they can be reused in other software. One advantage of Englishville is that L2 learners can see (and hear) utterances and then repeat them and even save their productions – contrary to some of the other software available. Another advantage of using a website is that it is very malleable, so that we decide what the learner sees, hears, how many times, the number of repetitions allowed, the order of stimuli etc. – everything can be controlled, therefore ensuring an easily comparable experiment or practice sessions.

The use of colour ranging from red through to blue (in order: red, orange, yellow, green, blue) renders the activity extremely comprehensible and can be used to practice lexical stress. As for intonation, the spectrogram shows the direction of the tone of voice making it possible to imitate a visual real-time model of an intonation pattern and to compare it to one's own melodic pattern.

3.1. Stimuli

Two parts were created: one for lexical stress and the second for intonation. The recordings were done by a female British native speaker, creator of Englishville. This choice was made due to technical constraints but for the future we plan on adding several speakers, both male and female and different accents.

The first part focuses on lexical stress and is composed of a list of 72 words which have been identified as problematic for French L2 learners. The initial list was made up from 56 words (collected from personal teaching experience), and 16 words from a published list (Chabert, 2018:111-119). Some examples are village, separately, Japan, Britain, effort, independent, harmonious.

The second part which focuses on intonation is made up of 30 sentences. As mentioned in the literature review, for the time being, we have chosen to focus on simple patterns and short utterances. There are ten statements and five of each of the following sentence types: Wh questions, Yes-no questions, echo questions and two-element lists. For example, we live in London/he is on the computer? / are you growing oranges or lemons? These correspond to falling and rising intonation and closed lists (cf. Intonation above).

4. Experiment

A trial experiment was set up to test if the participants who received multi-sensorial input produced better results, i.e., those closest to the expected intonation pattern, than those who had no input, or only had audio or visual input in the quest to determine the potential usefulness and effectiveness of this tool in improving prosodic features in the speech of French learners of English. In order to test multi-sensorial input, it seemed necessary to include groups that did not receive this type of input. Therefore, four groups were determined, and the following experiment devised. All of this was decided before knowing how many participants there would be. In hindsight, had we known the number was to be so small, the experiment would probably have focused exclusively on multi-sensorial input. However, to evaluate whether multi-sensorial input is effective for practising prosody it appeared necessary to have other groups who did not have the same input. It was thus decided that one group would have no input, with only the text to read in order to have a control group and assess if, without input, the participants managed to perform the task successively. It also seemed important to have one group who only had audio input and one group who only had visual input to see if the results of either of these were as good as the group who received multi-sensorial input. In light of previous research, it was hypothesized that the combination of both audio and visual input would yield the best results. However, to be able to confirm this, it was necessary to assess the

impact of all types of input. This was deemed necessary before further evaluations of Englishville.

Each participant read the same words and sentences in the same order, as they appeared on the screen but they each received different input:

1. GRP 1. Records words and sentences – no input. This is the control group.
2. GRP 2. Hears the utterances before recording them – auditory input only.
3. GRP 3. Sees the spectrograms in Englishville – visual input only.
4. GRP 4. Hears the utterances and sees the corresponding spectrograms on Englishville – multi-sensorial input.

Before the experiment started, the participants had one example word to familiarise themselves with the user interface. Each participant could only hear or see each item once, before recording their production. The participants clicked on the start button and a word or phrase appeared for them to read (or see or hear). They then had to click on the button record to activate the microphone and click the same button again once they had finished their recording before being presented with the next utterance. Those who saw the spectrogram received supplementary information about it. The following text appeared on the screen at the beginning of the experiment: *You can see a 3D spectrogram that corresponds to the text. The colours red and orange correspond to high intensity, yellow, green and blue indicate low intensity. You can also see the movement of the voice in the spectrogram, i.e., with downward or upward movement.* The participants who received input, be it auditory or visual were asked to imitate as closely as possible what they heard or saw. For this, the instructions were: *try to reproduce the effects (colours and movements) of each spectrogram you see.*

Previous research has shown that using visual aids often yields better results when learning prosody and even has lasting effects on speech production in general. It was therefore expected that Group 4 would have the better results than the other three groups because they would be able to see immediately if their spectrogram resembled or not the model and attempt to improve during the experiment.

4.1. Participants

The eight students who participated in this experiment were in their final year of a BA in English language and literature at the University of Caen Normandy. They were, on average 20 years old and had been learning English for at least ten years. They had all studied phonology and phonetics, notably intonation in their final year and the teaching model for those classes at our university is British English. Unfortunately, by the time Englishville was up and running, university classes had already finished. Therefore, finding participants was more complicated than anticipated and the experiment set up was no longer in full adequation with the number of participants. Despite this small number, we would like to present our

results, but we fully understand that it is difficult to talk about anything conclusive. Prior to the experiment, a questionnaire was given to each participant to collect information on their personal background, any difficulties they had as learners and their language practices outside the classroom. They each filled out a consent form following the standard rules in research. At the end of the experiment, participants are asked to provide feedback about the experiment in order to gather their impressions and improve the tool. As previously mentioned, the learner's evaluation of Englishville was at the heart of this trial experiment. We wanted to find out how much they understood the tool (with the brief explanation provided) and its objectives, how they chose to describe it (fun, interesting, easy to use) and how useful they thought it could be for learners of L2 English. The purpose of Englishville is to create a tool that is more than just a means of drilling but has the added element of visualising one's own speech and therefore being able to compare it to a model by means of a spectrogram.

5. Results

Recordings from the four groups were compared with the original recording and spectrogram to give an auditive and visual analysis. Stress and intonation patterns found to be the same as the model were awarded one point whereas those that differed got zero. The number of correct realisations for word stress and intonation patterns were then calculated for each group. The results for correct lexical stress can be seen in Figure 5.

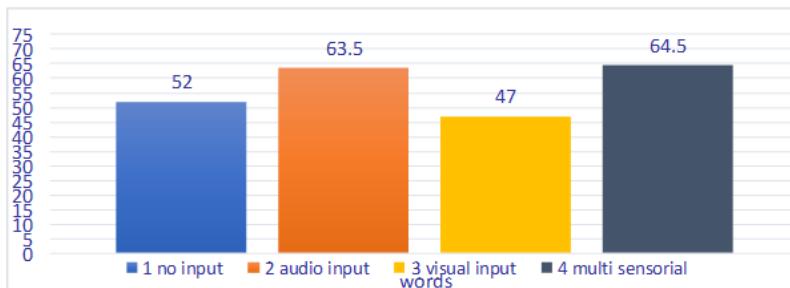


Figure 5: Average correct realisation of lexical stress (out of 72 words)

For lexical stress, it can be noted that the number of correct pronunciations is relatively high across all four groups. Group 1, who just read the words as they appeared on the screen correctly pronounced 52 out of 72 words. Group 3 who only saw the spectrogram without sound correctly pronounced less words (47 out of 72) than the control group (no-input). Group 4 – who saw the spectrogram and

heard the audio – correctly pronounced 64.5 out of 72 words on average, only 1 word (on average) more than group 2 who only received audio input (63.5/72). Even though the results for lexical stress are not significantly different, the two groups who received audio input did have better results than those who received none. However, there appears to be no extra benefit from seeing a 3D spectrogram for lexical stress.

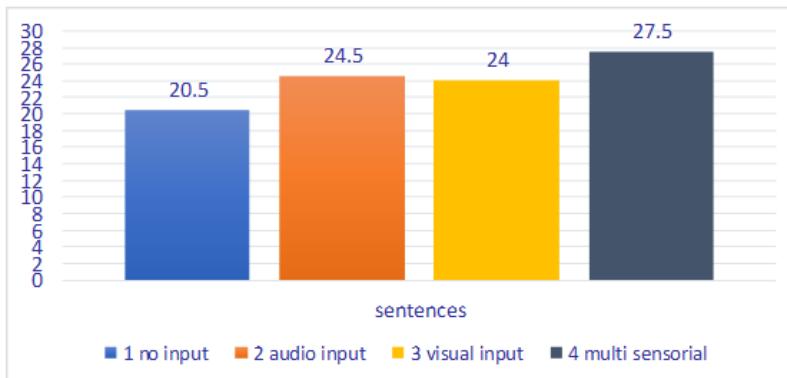


Figure 6: Average correct realisation of intonation pattern per input type (out of 30 sentences)

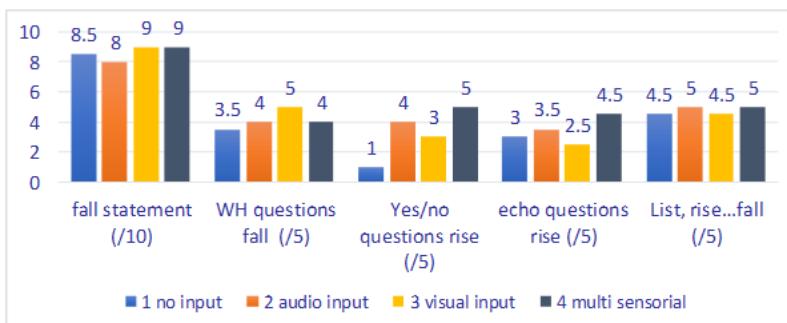


Figure 8: Average correct realisation of intonation pattern per input type

Regarding the results for intonation, group 1 realised the least number of sentences correctly (20.5 out of 30). Group 2 and 3 benefitted from having one type of input. Group 3, who only saw the spectrogram (no audio) seemed on average, to be able to use the visual intonation contour to the same extent as those who only had access to the audio (group 2). The number of correct realisations is the highest in group 4 (27.5 out of 30) who saw the spectrogram and heard the audio. The results suggest that visualising speech can be helpful in producing the correct intonation pattern. Now we can examine our results in more detail by analysing the participants' speech per sentence type to see if some intonation

patterns were more correctly realised than others. The results can be seen in Figure 3.

The results show that rising intonation is the most difficult for these French learners of English, both knowing when to use a rise and being able to realise one. The easiest pattern to produce is the one found in short lists (rise followed by a fall) both with and without input. Falling intonation associated with statements is also correctly produced. However, incorrect intonation patterns are found for questions in general, but especially those with rising intonation. There is a tendency for the participants' intonation patterns to be closer to the model when the participants receive auditory and/or visual input.

In the post-test questionnaire given to the participants, there was much positive feedback. They found the tool useful, fun, easy to comprehend and compare and modify their speech patterns. Here are a few comments that they wrote:

It allows to see immediately if the pronunciation is correct.

So armed with a visual of what is going on internally can help with better pronunciation.

To visually compare what one "must"/"wants" to say and what one actually says.

As there are 2 spectrograms, the comparison is easy.

It altered my pronunciation, I tried to follow the spectrogram

This feedback is very encouraging, and we can confirm that at least one of our objectives has been met.

6. Discussion

Prosody is one of the first linguistic elements acquired and the last lost. This can explain why it can be problematic for L2 learners and why L1 prosodic features are found in L2 productions.

The aim of this preliminary assessment of Englishville is first to see if there is a difference in the participants' oral productions depending on which type of input they received, and secondly to gather the participants perception of the tool. The results suggest that visualising the corresponding spectrogram can be more beneficial when dealing with intonation than when focusing on lexical stress in words. With such a small number of participants, we are conscious that it is impossible to draw any real conclusions due to speaker variation. For example, in group 3 the two participants had a very heterogenous level of language and also had a very different approach to the experiment. Out of the two participants, one said that they paid little attention to the spectrogram, whereas the second participant said that they tried to closely imitate the pattern and colours. The latter participant was a mature student who had quite a fossilised pronunciation. The fact that she found the tool useful and reported that she altered her pronunciation thanks to the spectrogram speaks volumes. They did however, express their frustration in not having access to the audio. The participants in this experiment

are at an advanced learning stage (final year of BA in English) which means that correct realisations were often high even without input. This is clearly a positive point for the participants but is less positive when faced with the task of evaluating the usefulness of such a tool. The words chosen were thought to be generally difficult for French learners but were clearly not that problematic for this group.

In the same vein, it was believed prior to the experiment that it would be easier for the participants to produce utterances with rising intonation than with falling intonation. This is because when French learners of English conclude at the end of a presentation, they tend to use a rise instead of a fall. It was therefore presumed that the same difficulty would be found during the experiment whereas the results show that correctly using rising intonation was more difficult – this result was unexpected. A possible explanation of this difficulty lies in the use of echo questions which may seem unnatural to produce when there is no added context. We may need to rethink the type of stimuli used, for example, imitating a model in short utterances can be found easy compared to using intonation correctly in longer phrases or in real interactions.

For lexical stress, it can be noted that the number of correct pronunciations is relatively high across all four groups. This is probably because the participants were advanced learners, having been taught word stress during their three-year degree.

The results suggest that visualising speech can be helpful in producing the correct intonation pattern. Rising intonation was more difficult for all groups but especially for group 1 who did not seem to know when to use this pattern.

As can be noted from the quotes above, overall positive feedback came from the participants – in particular, the four out of eight who saw the spectrogram. They found it especially useful when they were able to match their own spectrogram with the provided model. This leads us to infer that this tool can be both useful *and* usable. As previously mentioned, other tools exist but to our knowledge none of them have the same components, and so far, most experiments have been carried out with Praat.

7. Conclusion and further directions

This chapter discusses the process of conceiving and implementing a technological tool and an initial trial experiment which addresses the difficulty of acquiring English prosody for L2 learners. An increasing number of researchers are addressing the problems faced by learners in regard to L2 prosody by testing multi-sensorial techniques in the search of an efficient tool for the acquisition and training in prosody.

We have seen that other tools exist, but we hope to further contribute to this fast-developing field by presenting Englishville to the scientific and teaching community. We believe that Englishville corresponds to our students' needs and

desire for technological teaching tools. One of its advantages is its malleability, making it possible to add for example, words, phrases, speakers, accents, participants, or remove elements from it (for example, to only have one group) so that all participants have the same input.

Software such as Praat have been found to be too complicated for the uninitiated to use by some researchers. Our objective is to provide a tool which is easy to use for both teachers and learners and a fun way to raise awareness and practice prosody for L2 learners. This is motivated by the pedagogical analysis that teaching and learning prosody is not an easy task. It is also a quest for a more modern, student-centred approach.

It is impossible to draw any significant conclusions from such a small- scaled experiment but this is just the first step in our thought-process and the implementation of this tool. More experiments are planned, as is a training session, this time with first year university students. We are also currently working on an algorithm to automatically compare the students recording.

Further directions for Englishville are plentiful. First of all, more experiments must be carried out on a larger scale. Training sessions with a pre-test and a post-test are also planned. It is thought that the latter will provide a clearer picture on the potential of a multi-sensorial tool.

References

- Boersma, Paul. 2001. PRAAT, a System for Doing Phonetics by computer. *Glot International*, 5, 341–345.
- Chabert, Evelyne. 2018. *Bien Prononcer L'anglais – Manuel d'anglais Oral Pour les Francophones*. Génération 5.
- Chafeuloff, Michel. 2004. Voir la Parole. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)*, 23, 23–65.
- Colantoni, L., Marasco, O., Steele, J., and Sunara, S. 2014. *Learning to Realize Prosodic Prominence in L2 French and Spanish*. In Miller RT, et al (eds.), *Selected Proceedings of the 2012 Second Language Research Forum*. Cascadilla Proceedings Project, 15–29. <http://www.lingref.com/cpp/slrf/2012/paper3082.pdf>
- de Bot, K. 1983. *Visual Deedback of Intonation: Effectiveness and Induced Practice Behavior*. *Language and Speech*, 26, 331–350. <https://doi.org/10.1177/002383098302600402>
- Derwing, Tracey M., and Munro, Murray J. 2005. *Second Language Accent and Pronunciation Teaching a Research-Based Approach*. *TESOL Quarterly*, 39, 379–397. <https://doi.org/10.2307/3588486>
- Derwing, Tracey M., and Rossiter, Marian J. 2003. The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning*, 13, 1–17.
- Englishville: <https://demo.englishville.ovh/>
- Ginésy M. 1998. *Mémento de Phonétique Anglaise*, Nathan.
- Gorjian, Bahman, Hayati, Abdolmajid, and Pourkhoni, Parisa. 2013. Using Praat Software in Teaching Prosodic Features to EFL Learners. *Procedia - Social and Behavioral Sciences*, 84, 34–40. <https://doi.org/10.1016/j.sbspro.2013.06.505>

- Guierre, Lionel. 1979. *Essai sur L'accentuation en Anglais Contemporain : Eléments pour une Synthèse*. Thèse de doctorat, Université Paris-VII.
- Halliday, M.A.K., and Greaves, William S. 2008. *Intonation in the Grammar of English*. London: Equinox.
- Hardison, Debra M. 2004. Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings. *Language Learning and Technology*, 8, 34–52.
- Hardison, Debra M. 2010. Visual and Auditory Input in Second-Language Speech Processing. *Language Teaching*, 43(1), 84-95. <https://doi:10.1017/S0261444809990176>
- Hirst, Daniel. (1998) *Intonation in British English*. In Hirst, D., and Di Cristo, A. (Eds). *Intonation Systems: A Survey of Twenty Languages*, 56-77. Cambridge University Press.
- Hirst, Daniel, and Di Cristo, Albert. (eds). 1998. *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press.
- Jenkins, Jennifer. 2000. *The Phonology of English as an International Language*. Oxford: Oxford University Press.
- Imber, Brenda, Maynard, Carson, Parker, Maria. 2017. Using Praat to Increase Intelligibility through Visual Feedback. In M. O'Brien M. and J. Levis (eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, 195-213.
- James, E. Flege. 1976. The Acquisition of Prosodic Features of Speech Using a Speech Visualizer. *IRAL*, 14(3), 227-243. <https://doi.org/10.1515/iral.1976.14.3.227>
- Komissarchik, Edward, and Komissarchik Julia. 2000. BetterAccent Tutor – Analysis and Visualization of Speech Prosody. *Proceedings of InSTILL*, 86–89.
- Kartushina, et al. 2015. The Effect of Phonetic Production Training with Visual Feedback on the Perception and Production of Foreign Speech Sounds. *Journal of the Acoustical Society of America*, 138(2), 817-832. <https://doi.org/10.1121/1.4926561>
- Munro, Murray J., and Derwing, Tracey M. 1995. Foreign-Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, 45(2), 73-97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, Murray J., and Derwing, Tracey M. 1998. Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning*, 48(2), 393-410. <https://doi.org/10.1111/0023-8333.00047>
- Offerman, Heather M., and Olson, Daniel J. 2016. Visual Feedback and Second Language Segmental Production: The Generalizability of Pronunciation Gains. *System*, 59, 45-60. <https://doi.org/10.1016/j.system.2016.03.003>
- Olson, Daniel J. 2014. Phonetics and Technology in the Classroom: A Practical Approach to Using Speech Analysis Software in Second Language Pronunciation Instruction. *Hispania*, 97(1), 47-68. <https://doi.org/10.1353/hpn.2014.0030>
- Setter, Jane, and Jenkins, Jennifer. 2005. Pronunciation: State-of-the-Art Review Article. *Language Teaching*, 38(1), 1–17. <https://doi.org/10.1017/S026144480500251X>
- Setter, Jane, and Stojanovik, Vesna, and Martínez-Castilla, Pastora. 2010. *Evaluating the Intonation of Non-Native Speakers of English Using a Computerized Test Battery*. International Journal of Applied Linguistics, 368-385. <https://doi.org/10.1111/j.1473-4192.2010.00253.x>
- Tench, Paul. 1996. *The Intonation Systems of English*. London: Cassell.
- Vaissière, Jacqueline. 1991. Rhythm, Accentuation and Final Lengthening in French. In J. Sundberg, L. Nord, and R. Carlson (eds.), *Music, Language, Speech and Brain*, 108-120. Macmillan Press. https://doi.org/10.1007/978-1-349-12670-5_10
- Wells, John C. 2007. *English Intonation: An Introduction*. Cambridge: Cambridge University Press