Englishville: A new way of practising prosody

Kizzi Edensor Costille

Edensor Costille, K. (2023). Englishville: A new way of practising prosody. In A. Henderson & A. Kirkova-Naskova (Eds.), *Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices* (pp. 61-69). Université Grenoble-Alpes. https://doi.org/10.5281/zenodo.8173981

Englishville: A new way of practising prosody

Kizzi Edensor Costille
University of Caen Normandy

Despite evidence that prosody plays an important role in the intelligibility, comprehensibility and accentedness of non-native discourse (Munro & Derwing, 1995, 1998), it is seen as difficult to teach (Setter et al., 2010). One way of making prosody easier to teach and understand is by using a real-time 3D spectrogram such as the one used on the website Englishville (Costille, 2020). Four groups of French students, enrolled in their third year of a BA in English, took part in this experiment. Thirty short sentences focusing on intonation were recorded by a female native British speaker. All participants read and recorded the same phrases as they appeared on the screen and groups 3 and 4 received specific explanations regarding the spectrogram and intonation contours. The first group simply read the phrases (limited input) and recorded their own productions. The other 3 groups received supplementary input: group 2 read the text and heard the corresponding audio recordings (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). The recordings were then compared in Englishville to the expected intonation pattern and given one point per matching pattern. The results do not show that seeing speech systematically improves students' intonation but did show that the students felt the tool was useful and easy to use.

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

1 Introduction

Teaching and research in second language acquisition has long 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 second language (L2) learning. For segmentals, teachers have found ways to render the theoretical aspects of phonemes more comprehensible by using multi-sensorial techniques – be it with their own mouth, videos of another's mouth or by demonstrating on sagittal sections. Learners can simultaneously see the different positions of jaw, lips, and tongue, and hear the phonemes. The representation of intonation or lexical stress is more abstract and tools such as Praat (Boersma & Weenink, 2001), however helpful for research, remain difficult to use if learners are not previously trained how to use them (Setter & Jenkins, 2005; Setter et al., 2010).

It might be that acquiring or improving prosody is complex partly because of the lack of physical or visual aids available to learners. It could even be argued that its abstractness dissuades teachers from teaching it. This led to the creation of Englishville (Costille, 2020) – a website dedicated to practising prosody by means of a 3D spectrogram. An experiment was set up to test its usability and usefulness regarding prosody.

This chapter gives a brief overview of prosody and research on L2 learners with a special focus on the prosodic elements in the experiment. It also discusses research that has used auditory and multi-sensorial tools. It then describes the methodology used, followed by the experiment and results. Lastly, the findings are discussed in light of the possible contribution of Englishville to the field.

2 L2 prosody

Prosody, also known as suprasegmentals (to be understood as all that is not segmental), includes elements such as: rhythm, intonation, stress, and pauses. Drawing on research studies investigating first language (L1) acquisition, Hirst and Di Cristo (1998) explain that prosody is likely to be acquired by a child before any other phonetic features, moreover it is likely to be the last feature lost when aphasia strikes or when another language is learnt. The fact that we learn the prosody of our L1 in the very early stages of development can explain why it is difficult to learn later in an L2. Previous studies have confirmed that acquiring L2 prosody is challenging even for advanced learners (Colantoni et al., 2014). However, it is generally accepted that using inaccurate intonation patterns, i.e., such that differ from native productions, can lead native listeners' to either misinterpret the intended meaning or show negative stereotyping towards the L2 speaker.

Studies focusing on L2 pronunciation instruction and the relevance of L2 speech in speaker interactions have investigated various prosodic features. Jenkins (2000) finds the most important suprasegmental features in NNS–NNS interactions to be contrastive stress, the direction of pitch movements, word stress placement, and stress-timed rhythm. Other authors share a similar point of view on prosody. For instance, Munro and Derwing (1995, 1998) and Hardison (2004, 2010), argue that prosody plays a significant 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. Furthermore, Derwing and Rossiter (2003) have shown that L2 fluency and comprehensibility significantly improved after a 12-week instruction period on the pronunciation of prosodic features.

2.1 Research on visual and auditory tools

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 (de Bot, 1983; James, 1976). 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 (Imber et al., 2017; Kartushina, et al., 2015; Offerman, & Olson, 2016; Olson, 2014, Setter et al., 2010). Gorjian et al. (2013) compared two methods of teaching stress and intonation: a) a traditional one that uses repetition and explanations about acoustic properties of speech; and b) a computer-assisted one that uses Praat software. The results showed that learning prosody with Praat was significantly more beneficial. As the authors point out, the first method is generally teacher-centred, leaving students passive in the classroom. In contrast, the use of multi-media tools and multi-sensorial software places the student at the centre of their acquisition.

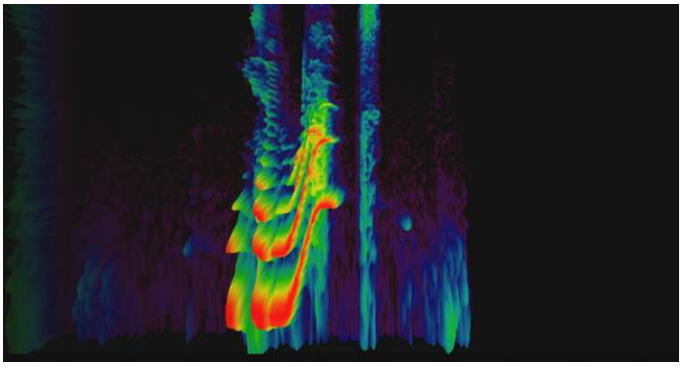
Software, such as Praat, require practice and training to use and some research has concluded that combining sound and image led to slightly more mixed results in learning prosody, often due to the complexity of the software used (Setter et al., 2010). Given the technical side of these tools, it can be difficult to motivate students to familiarise themselves with them and then work on intonation. The results of certain studies (Gorjian et al., 2013) showed that learners improved when using Praat, which could be explained simply by the additional time spent working on prosody to comprehend what they were seeing. Perhaps using real-time displays of intonation might prove to be even more comprehensible for the learners – they can see their speech appear as they speak, enabling them to test different intonation patterns more easily. We argue 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 aural perception. Therefore, depending on the pronunciation feature practised, the use of multi-sensorial tools and methods may be helpful, even having a global positive effect on L2 speech production in general.

3 Englishville: Methodological approach

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 called "Englishville" (Costille, 2020). Englishville uses a 3D spectrogram and facilitates 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 a corpus, set up experiments and participate in them. One advantage of Englishville is that L2 learners can hear the audio, see the spectrogram and intonation contour and the corresponding text, then repeat and save their own productions. 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 simultaneously compare it to one's own melodic pattern. The spectrogram used in Englishville can be seen in Figure 1 with an example of rising intonation.

Figure 1

A Real-time Display of Pitch Contour and Intensity in Englishville (Costille, 2020) of the Utterance "He said what?" Pronounced with a Rising Intonation



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 speech in order to improve their pronunciation of the different melodic patterns. 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 (open questions) are normally said with a falling tone on the last lexical item whereas Yes/No-questions (closed questions) are normally uttered with a rising tone on the nucleus. As for intonation in lists, Wells (2007) 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 and explains that "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 ↘orange juice" (p. 75).

(1) You can have ↗ coffee | or ↘ tea.
(2) You can have ↗ coffee | or ↗ tea.

The structure of the sentences used in the intonation task of the Englishville experiment correspond to the two main intonation patterns (fall, rise) and example (1) in the case of a closed list. French L2 speakers of English tend to struggle most with falling intonation and use rising intonation for all types of statements. This is typically noticeable when a speaker concludes an oral presentation with a rise, which can leave the listener frustrated and/or surprised when they realise that the presentation is, in fact, finished (the non-finality effect of rising intonation).

3.1 Research questions

The study aimed therefore to investigate the following research questions:

RQ1: Can seeing a real-time 3D spectrogram enable L2 learners of English to reproduce certain intonation patterns?
RQ2: What kind of input is more beneficial out of limited, audio, visual and multi-sensorial input?

3.2 Experiment

An experiment was set up with four groups of participants (see Table 1): Group 1 (control group) received limited input and read the sentences as they appeared on the computer screen; Group 2 had audio input, i.e., they heard the sentences as recorded by the female native speaker and saw the text for each sentence on the screen; Group 3 had visual input which corresponded to seeing the spectrogram and the melodic pattern (as in Figure 1) and read the text without hearing the audio recordings, and Group 4 had access to both aural and visual input. The aim of the experiment was to test if the participants who received multi-sensorial input (Group 4) produced better results, i.e., those closest to the expected intonation pattern.

Table 1

Experiment Conditions for each Testing Group: Type of Input and Procedure

Group	Type of input	Experiment procedure
1	limited	Records words and sentences.
2	auditory	Hears the utterances before recording them.
3	visual	Sees the spectrograms in Englishville.
4	multi-sensorial	Hears the utterances and sees the corresponding spectrograms on Englishville.

At the beginning of the experiment, the participants had one example to familiarise themselves with the user interface. Each participant read the same sentences in the same order as they appeared on the screen. They only heard or saw each item once, before recording their own production. They had to click the button *play* to hear or see the visual model recording and then had to click the microphone button on and off to record their own speech. They could not listen to the same sentence more than once, but they could pronounce it as many times as they liked while their microphone was activated. By pressing the microphone button a second time, their recording was saved, and the next utterance automatically appeared. At the beginning of the experiment, Groups 3 and 4 received supplementary information about what they were about to see. For example, they were informed that they would see the movement of the tone of voice in the spectrogram (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.

In light of previous research, it was hypothesised that the combination of both audio and visual input would yield the best results. It was therefore expected that Group 4 would have 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.

3.3 Participants

Twenty French students (n = 5 in each Group 1–4; M = 8, F = 12) at the beginning of their third year of a BA in English Language and Literature at the University of Caen Normandy participated in this experiment. They were, on average, 20 years old and had been learning English for at least ten years. Their level of English was estimated to be B2+, based on their teacher's experience with the CEFR scales. They had all studied phonology and phonetics (including intonation). For these classes, the teaching model was British English. Prior to the

experiment, each participant completed an online questionnaire about their language and personal background.

The learner's evaluation of Englishville was also of high interest in this study, therefore, after the experiment they were asked to give feedback via an online questionnaire about Englishville and their impression of it and the experiment.

3.4 Stimuli

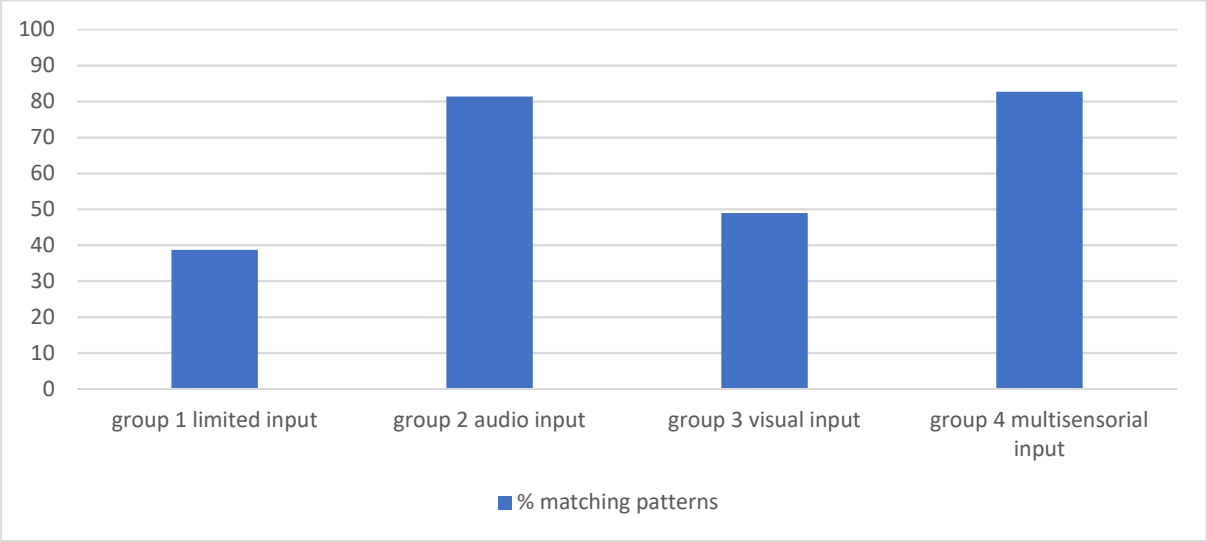
Thirty sentences consisting of simple patterns and short utterances were recorded by a female British native speaker, also the creator of Englishville. The sentences included ten statements and five of each of the following sentence types: Wh-questions, Yes/No-questions, echo questions and a two-element closed list. For example: *We live in London* (statement with falling intonation), *Where's the manual?* (Wh-question said with falling intonation), *May I lean on the railings?* (Yes/No-question said with rising intonation), *He is on the computer?* (echo-question with rising intonation), *Are you growing oranges or lemons?* (two-element closed list said with a rise followed by a fall).

4 Results

Recordings from the four groups were compared with the original recording and spectrogram to give an auditive and visual analysis. Intonation patterns found to match the model were awarded one point, for example, a final fall, rise or rise+fall (closed lists) whereas those that differed got zero. The number of matching realisations for intonation patterns were then calculated for each group, as can be seen in Figure 2.

Figure 2

Percentage of Matching Realisations of Intonation Pattern per Input Type (out of 30 sentences)



In general, the results show that having any kind of input is beneficial as Group 1, who received limited input, only matched 38.67% of the 30 sentences. For Group 3, seeing the spectrogram slightly improved their overall percentage of matching realisations (48.9%)

compared to Group 1 (marking an increase of 10.23% points). Group 2, who only heard the corresponding audio, had nearly the same percentage of matching realisations (81.33%) as Group 4 (82.67%), who benefitted from multi-sensorial input.

In the post-test questionnaire, which was completed by 10 participants from Groups 3 and 4, there was much positive feedback. Those who saw the spectrogram (Groups 3 and 4), found the tool useful (n = 7), very good (n = 7), easy both to comprehend (n = 7) and to use (n = 5). Only one participant found it hard to use (n = 1).

5 Discussion

The study aims 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 participants' views and impressions of Englishville as a teaching/learning tool. The results suggest that merely visualising the corresponding spectrogram is slightly more beneficial than having no input at all, but barely makes a difference when the participants had access to the audio recordings. The fact that Group 3 did better than Group 1 suggests that some participants were able to learn from what they were seeing and improve some of their productions.

The minor difference between the results of Group 2 and Group 4 begs the question of sensorial overload. It is possible that having to read the text, pay attention to the spectrogram and to the audio resulted in too much information for the participants and may have led them to ignore some or all the input. The sentences were simple and short, and the written text was probably superfluous. It would have been possible for this level of L2 learners to repeat the text simply from the audio recordings, without the written text. In hindsight, Groups 1 and 2 are said to have limited input or only audio input, where it could be argued that both also had visual input because they saw/read the written text for each sentence, no visual information about intonation accompanied the written forms.

Given the small sample and with only five participants per group, it is difficult to draw any firm conclusions due to speaker variation. For example, it is possible that the spectrogram helped certain speakers in Groups 3 and 4, but it is also possible that some participants paid no attention to it at all. In addition to speaker variation, the fact that the participants could only listen/see each item once and could only record their own speech once before moving on to the next phrase, left them with little possibility to improve. They could of course repeat their productions several times before validation, but few did this. The main and global objective of Englishville is to enable learners to improve their prosody because they can see and use the spectrogram. For future experiments, it would be important to provide learners with several opportunities to do the tasks.

6 Conclusion and further directions

This article addresses the difficulty of acquiring English prosody for L2 learners and the use of a multi-sensorial tool to improve it. Previous research has shown that using visual aids often yields better results in training sessions when learning prosody and even has lasting effects on speech production in general (Derwing & Rossiter, 2003). Englishville was designed to explore this issue. 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.

Despite our inconclusive results regarding the benefit of visualising speech, the students' feedback leads us to believe that Englishville corresponds to their desire for technological teaching tools. Our overall objective is to provide a tool which is easy to use for both teachers and learners and is also a useful way for L2 learners to practice prosody and raise their

awareness of it. This is motivated by the observation that teaching and learning prosody seems challenging, and the fact that software such as Praat is too complicated to be used by untrained learners. Learners should, however, be at the centre of any modern pedagogical approach and Englishville makes this possible. Therefore, we recommend that future use of this tool – whether for research or teaching – should allow students to hear, see the model, and record their own productions as many times as they want.

Various factors had an impact on the participants, including the number of times they could listen to the recordings, speaker variation, and sensorial overload. While the overall results do not allow us to confirm that seeing the melodic pattern by means of a spectrogram improves learners' productions, positive feedback was given by those who saw and used the spectrogram. They found it especially useful when they were able to match their own spectrogram with the model provided. In this regard, Englishville can be considered successful, as it encourages learners to make autonomous, critical comparisons. Perhaps with more time, their actual productions would also show improvement. To that end, an eight-week training session is currently being carried out with first year university students specialising in English. We hope that the latter will provide a clearer picture on the potential of this multi-sensorial tool.

References

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Colantoni, L., Marasco, O., Steele, J., & Sunara, S. (2014). Learning to realize prosodic prominence in L2 French and Spanish. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. M. Miguel, A. M. Tseng, A. Tuninetti, & D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum* (pp. 15–29). Cascadilla Proceedings Project. http://www.lingref.com/cpp/slrf/2012/paper3082.pdf

Costille, K. (2020, May 6) Englishville. https://demo.englishville.ovh/

de Bot, K. (1983). Visual feedback of intonation: Effectiveness and induced practice behavior. *Language and Speech, 26*, 331–350. https://doi.org/10.1177/002383098302600402

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*, 1–17.

Gorjian, B., Hayati, A., & Pourkhoni, P. (2013). Using Praat software in teaching prosodic Features to EFL learners. *Procedia - Social and Behavioral Sciences, 84*, 34–40.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology, 8*, 34–52. http://llt.msu.edu/vol8num1/hardison/

Hardison, D. M. (2010). Visual and auditory input in second-language speech processing. *Language Teaching, 43*(1), 84–95. https://doi.org/10.1017/S0261444809990176

Hirst, D., & Di Cristo, A. (Eds). (1998). *Intonation systems: A survey of twenty languages*. Cambridge University Press. DOI: 10.1353/lan.2000.0088

James, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *14*(3), 227–243. https://doi.org/10.1515/iral.1976.14.3.227

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Imber, B., Maynard, C., & Parker, M. (2017) Using Praat to increase intelligibility through visual feedback. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference* (pp. 195–213). Iowa State University..

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817–32. https://doi.org/10.1121/1.4926561

Munro, M. J., & Derwing, T. 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, M. J., & Derwing, T. M. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(2), 393–410. https://doi.org/10.1017/S0261444811000103

Offerman, H. M., & Olson, D. 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, D. 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. http://dx.doi.org/10.1353/hpn.2014.0030

Setter, J., & Jenkins, J. (2005). Pronunciation: State-of-the-art review article. *Language Teaching, 38*(1), 1–17. DOI: 10.1017/S026144480500251X

Setter, J., Stojanovik, V., & Martínez-Castilla, P. (2010). Evaluating the intonation of non-native speakers of English using a computerized test battery. *International Journal of Applied Linguistics*, 20(3): 368–385. http://dx.doi.org/10.1111/j.1473.

Wells, J. C. (2007). *English intonation: An introduction*. Cambridge University Press.

About the author

Kizzi Edensor-Costille is a Lecturer at the University of Caen, Normandy where she teaches in the Department of English and Language Science. She did her thesis in the Speech and Language Lab (LPL) at Aix-Marseille University and specialised in regional accents in the UK and Ireland. She has also developed an interest in the field of L2 English learning, particularly in terms of perception and pronunciation. More recently, she has been working on the acquisition of prosody by non-native speakers. For this purpose, she developed a website called Englishville where L2 learners can visualise a real-time 3D spectrogram to practice different intonation patterns by seeing both the model and their own intonation.

Email: Kizzi.edensor-costille@unicaen.fr