Using BetterAccent Tutor and Praat for Learning English Intonation

Naima Hamlaoui
Department of English
Faculty of Letters, Human and Social Sciences
El-Bouni Campus
Badji Mokhtar University,
Annaba, Algeria

Nawel Bengrait
Department of English
Faculty of Letters and Languages
University of Guelma 8 Mai 1945
Guelma, Algeria

Abstract
Learning English pronunciation necessitates a deep understanding of the prosodic features such as intonation. The present study reports the effective use of BetterAccent Tutor and Praat programs for the learning of English intonation with an experimental and a control group that involves nine students during a third semester phonetics course, at the Department of English, University of Guelma, Algeria. The research aims at investigating the usability of audio-visual feedback paradigm to learn prosody by Algerian students in order to enhance their English intonation renditions. The experimental method enabled generating, measuring and analyzing intonation contours, pitch and intensity of English wh-questions and tag-questions. The results revealed that the audio-visualization technology support helped the students to identify the errors and to understand the functionality of acoustic features they produce regarding intonation patterns.

Keywords: Audio-visual feedback, intonation software, pitch, intensity, acoustic features

Introduction

Research in applied linguistics and second language (L2) acquisition demonstrates that pronunciation instruction has been marginalized compared to morphology and syntax growing body of literature (Derwing & Munro, 2005). The secondary position given to pronunciation has urged researchers to implement new methods with the advent of speech technologies that are accessible and up-to-date. Computer-aided pronunciation tools are integrated into second language classroom in higher education to serve as the medium to display and analyze the segmental and prosodic features of English language. Concerns raised about the difficulties that users may encounter in interpreting and understanding the visual display of speech urge experts to examine pronunciation software interface.

1. Literature Review

1.1 English vs Algerian Arabic Intonational Systems

L2 learners' failure or success in acquiring the target language segmentals and prosodic features is due to the learners' dialectal background (Odisho 2005, p. 66). English and Algerian Arabic (Darija) phonetic inventory systems share similarities as well as differences. Intonation in English and Algerian Arabic is dependent on pitch levels that are ranged from low and mid to high. Native speakers may mishear an utterance when L2 learners not only mispronounce sounds but also use improper pitch patterns.

Different pitch levels depict several intentions. For instance, interjections, statements, wh-questions and commands in English are produced with a falling contour (Odisho, 2005). Algerian Arabic shows similar intonational patterns of the mentioned structures (Benrabah, 1987, p. 81) as displayed in the following examples:

/ki`bnina/ (How tasty!)
/mʃitəl `ba:rəħ / (I went yesterday)
/ʃħalxə `ɬɬə stu / (How did you pay for it?)
/bəlaʕəl ˌba:b/ (Close the door)

English and Algerian Arabic yes/no questions, requests and listing items are characterized with a rising contour (Odisho, 2005; Benrabah, 1987, p. 81). For example:

/təmʃi ˌmʕa: ja / (Do you want to go with me)
/ʃrinalba ˌta:tələ ˌʕnəbwə`ʧina / (We bought potatoes, grapes, and oranges)
/ki ˌ ʃafətni `ħarbət / (When she saw me| she ran away)

Pitch patterns in Algerian Arabic and English may serve various functions with some language structures; also, the falling pitch in Algerian Arabic does not reach a low level as in English. Pitch changes in English may occur on a single vowel or a group of vowels; however, in Algerian Arabic the movements of pitch fall on a vowel at some point in time and a sudden or unexpected change in pitch may occur in the next vowels. These differences may lead to deficiencies in mastering English intonation patterns and may result in accented speech and misunderstanding of Algerian learners' intent by native speakers of English.

Intonation comprises linguistic and phonetic components in which the former deals with the pragmatic and phonological explanation of phonetics features and how speakers and listeners

employ various intonational patterns to interact in a communicative setting, and the latter with intonation acoustic, perceptual and physiological patterns ('t Hart et al., 1990). Intonation patterns serve different paralinguistic and linguistic functions in which the former expresses attitudinal and emotional nuances (Bolinger, 1986) while the latter indicates sentence types (Celce-Murcia et al., 1996). Thus, the present study focuses on the linguistic functions of intonation contours related to specific types of sentences: wh-questions and tag-questions.

1.2 Visualization Technology of Intonation

The growing interest in computerized learning dates back to the 1970s and it focuses on the analysis of speech patterns, using different commercialized acoustic phonetics software and hardware. Advances in speech technologies make it possible to share and implement better ways to learn and teach the prosodic feature of intonation. Computer Assisted Language Learning software (CALL) allows speech to be visualized and graphically represented by demonstrating pitch contours and sounds waveforms. The innovation of the early software *Visi-Pitch* by Kay Elemetrics (1986) provided visual display and analysis of native and non-native speakers' productions (de Bot, 1983). In light of this, the prosodic features and intonation contours in particular have been the center of interest of many researchers (Anderson-Hsieh, 1992, 1994; Chun, 1989, 1998, 2002; Molholt, 1988), in which L2 learners combining auditory and visual tools achieved better results than depending only on auditory input (de Bot, 1983).

Investigating the use of visual feedback paradigm of intonation contours took place mainly in laboratories settings (de Bot, 1983) and it was directed to advanced foreign learners of phonetics-centered course (Lord, 2010). Predominant concerns were raised about implementing visual feedback model in L2 classroom to elementary level learners who are at the level of developing their language skills. Some researchers claimed that speech analysis technologies are better used by trained instructors and these devices are too advanced for students and may not suit their needs (Derwing, 2010; Setter & Jenkins, 2005; Wang and Munro, 2004). Also, the students who have limited competence in using signal analysis software may not interpret intuitively spectrograms unlike the visual display of some forms of intonation contours (Chun, 1998; Hardison, 2004). In his study based on a questionnaire to students in a third semester Spanish course to evaluate their perception of intonation contours using Praat software, Olson (2014) reported that visual feedback paradigm was beneficial to conceptualize pronunciation features in L2 classroom.

1.3 User Interface

User Interface (UI) stands for the interaction between computer program and user. It is the visual part that displays information on the computer screen. The design of UI depends on three rules (Bahrami, 2012): (1) the user is placed in control; (2) the user's memory load is reduced; and (3) the interface is to be consistent. The process of User Interface Design (UID) involves four basic activities (Pressman, 2009): (1) interface design; (2) implementation; (3) interface validation; and (4) user, task and environment analysis.

The software application failure or success is related to UID. Designers have to take into consideration users' gender, age, personality traits, motivation and educational level, and ethnical and cultural background because the UI may not be applicable by all computer users but it may be intended for specific learners (Shneiderman, 1997). Learners' affective factors that

influence the teaching and learning process may affect the use of UI such as the pedagogical stimuli that involve coercion, requirement and motivation (Mehrdad, 2007). It is frustrating to the learner when s/he experiences difficulty in using the software especially when his/her mistakes are not overcome; the learner is more likely to dislike the program regardless of the functions offered by the application; thus the interface has to be accessible to the learner (Pressman, 2009).

2. Research Questions

Taking into consideration the necessity for systematic pronunciation methods and the applicability of visual feedback paradigm in enhancing prosodic features production, the paper addresses three research questions:

(1) Does audio-visual feedback help Algerian learners to improve English intonation production in the classroom?
(2) Do similarities and differences between English and Algerian Arabic influence the rendition of English intonational patterns?
(3) Which software user interface suits the needs of Algerian students?

3. Experimental Study
3.1 Methodological approach

The answer to the above research questions requires the exposure of the experimental group to audio-visual feedback provided by BetterAccent Tutor software while the control group listened only to audio files of the same recordings produced by native speakers of American English. The materials employed in this study were a set of three wh-questions and tag-questions that were taken from BetterAccent Tutor curriculum (see Appendix). The collected data are analyzed also with Praat visual technology support in which pitch contours, duration and intensity of the produced questions are displayed on the screen.

3.2 Systems Description: BetterAccent Tutor 2.0 Program and Praat

BetterAccent Tutor software is designed to help learners improve their American English pronunciation with primary focus on the prosodic features: intonation, stress and rhythm. The software application enables interaction between the learner and the Tutor, in which the user selects language structures from the available comprehensible curriculum that involves respectively: Word Stress, Simple Statements, Wh-Questions, General Questions, Repeated Questions, Alternative Questions, Tag Questions, Commands, Exclamations, Direct Address, Series of Items, Long Phrases and Tongue Twisters. The software UI permits the learner to listen to the audio recording of the native speaker's rendition of the selected item and to record his/her pronunciation. The software provides the learner with an audio-visual feedback and analysis of the learner produced stress, intonational and rhythmic patterns. The explanation box allows the learner to compare his/her performance with the one of the native speaker and correct it after many trials.

Praat is a speech visualization program that can be freely downloaded from the internet and installed on the computer. It was elaborated by Boersma and Weenink (2008) at the Department of Phonetics, University of Amsterdam. The program is constantly updated and it comprises acoustic features as pitch contours, spectrogram analysis, intensity analysis, formant

analysis and many other functions. It can measure maximum, average and minimum pitch levels that indicate intonation contours; in addition, the spectrogram represents in colors both the frequency and intensity in which the latter demonstrates stress placement in sequence of speech.

3.2   Participants
Two groups were drawn from second year students in the third semester taking an intermediate level phonetics course in which the students learn for the first time the acoustic characteristics of English prosodic features, mainly intonation. Five students in the experimental group received audio-visual feedback provided by BetterAccent Tutor and Praat, while four students in the control group listened only to audio-recordings.

3.3   The Task
The instructor invited the participants to record the productions of *wh-questions* and *tag-questions* during the pre-test phase (session 1), then after listening to the sample exercise in the test phase (session 2) and finally during the post-test phase after many trials (session 3). The control group recorded their renditions using the *record mono sound* function of Praat and the instant recording and analysis of data was avoided to give more credibility to the research. The experimental group used the researcher's computer to have access to BetterAccent Tutor and Praat programs in order to listen and visualize the native speaker's pronunciation, and to record and visualize their productions. The native speakers who participated in this study were 4, in addition to the model speaker on BetterAccent Tutor.

3.4 Procedure
The control and experimental groups were given equivalent amount of time for the productions of *wh-questions* and *tag-questions*. Both groups spent 180 minutes in one week organized in three sessions, in order to complete the experiment activities. Throughout the experiment phases, both groups did not receive from the instructor any verbal corrective feedback or clarification regarding the target prosodic patterns.
Pre-test recording Phase: the participants recorded in isolation the selected items prior to the class session.

Re-recording and Practising Phase: the instructor asked the participants of both groups to re-record the target questions. Unlike the control group, the experimental group participants were encouraged to visualize, compare and analyze their own productions and the native speaker's renditions of the intonational compositions.

Post-test re-recording Phase: the participants were required to submit the final recordings after many trials in the last session.

3.5 Measuring Techniques
The investigators collected the minimum and maximum pitch of the target questions (6 in total) in order to calculate the mean pitch (in Hertz or Hz) and average intensity (in decibels or db) of these values of Algerian non-native students and native speakers of English and experimental group vs control group, with the use of Praat program.

4.   Findings
(i) Wh-questions

Wh-questions are characterized with a falling intonation (Celce-murcia et al., 1996). The participants read a set of questions that were taken from *BetterAccent* Tutor (see Appendix). The following is an example of a *wh-question* produced by a native speaker of American English and an Algerian participant and examined with *BetterAcent Tutor* (Figure 1).

Example: What's this?

Native

Speaker of

English

Algerian
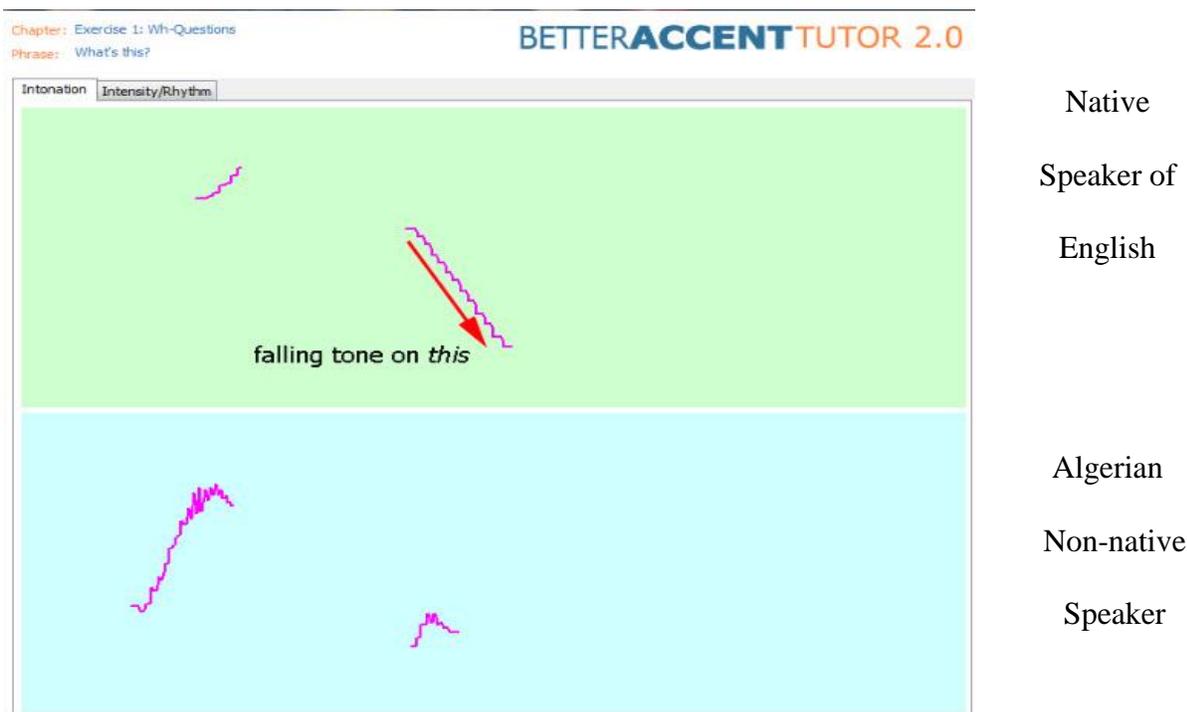
Non-native

Speaker

***Figure 1*. Wh-question Pitch Contour Productions of Native Speaker vs Algerian Non-native Speaker of English on BetterAccent Tutor Program**

The results demonstrate that English wh-questions end with a falling tone that requests more information and the center of intonation contour differs depending on the more prominent word in the string of speech. In Algerian Arabic, falling tone is associated with *wh-questions* (Benrabah, 1987); however, the majority of Algerian learners speaking English tend to place the pitch contour on the interrogative word that is pronounced with a rising tone regardless of where the center of intonation pattern is Figure 2.
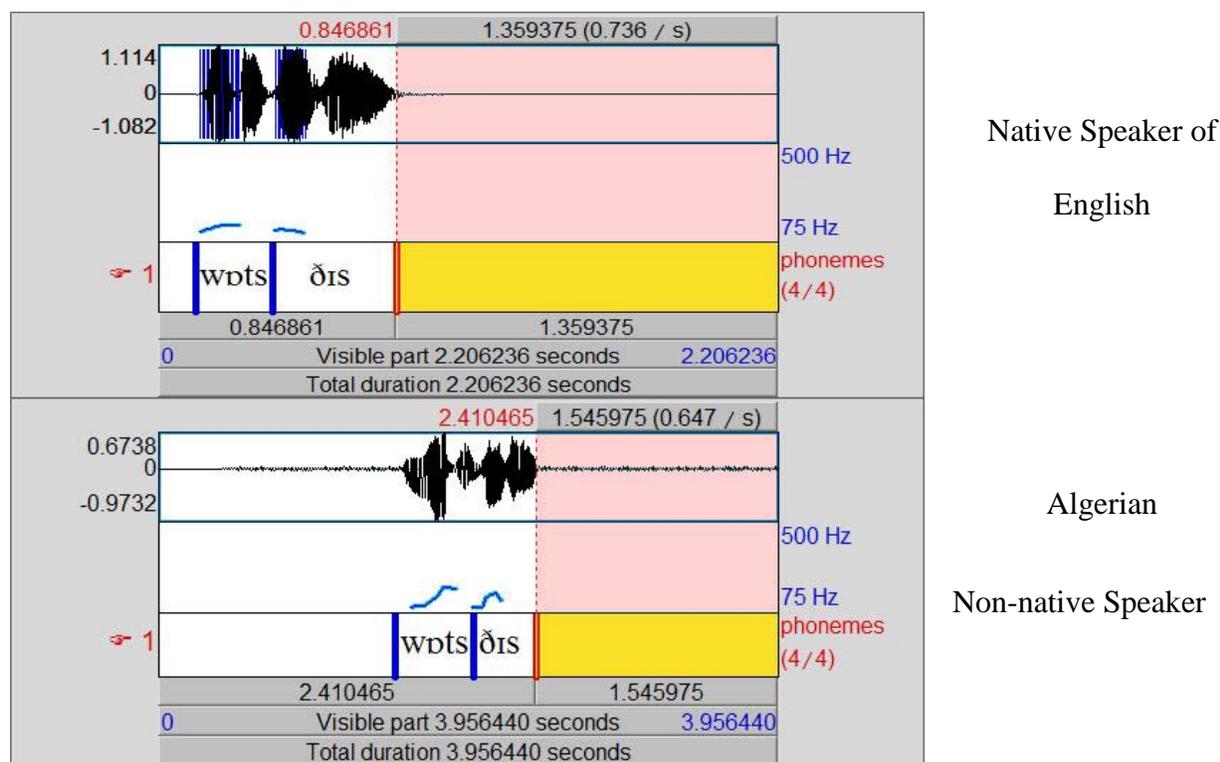
Figure 2. **Wh-question Pitch Contour Productions of Native Speaker vs Algerian Non-native Speaker of English: Analysis with *Praat***

The main difference between native speakers and Algerian learners' productions of English *wh-questions* is the mean pitch. The average pitch of English native speakers is 132 Hz that contrasts with 138 Hz of the standard pitch contour of Algerian participants. Employing rising pitch instead of falling pitch expresses confirmation rather than genuine question. The results are displayed in table 1.

Table 1.*Wh-questions Mean Pitch*

Participants	Mean Pitch (Hz)	Intensity Average (db)
Native Speakers	132	78
Algerian non-native speakers	138	72

The experimental and control groups achieved quite similar results regarding the production of wh-questions during the pre-test stage. However, there is a significant difference of experimental group wh-questions mean pitch in test and post-test phases, in contrast to the control group that were quite equivalent during the three sessions as shown in figure 3.
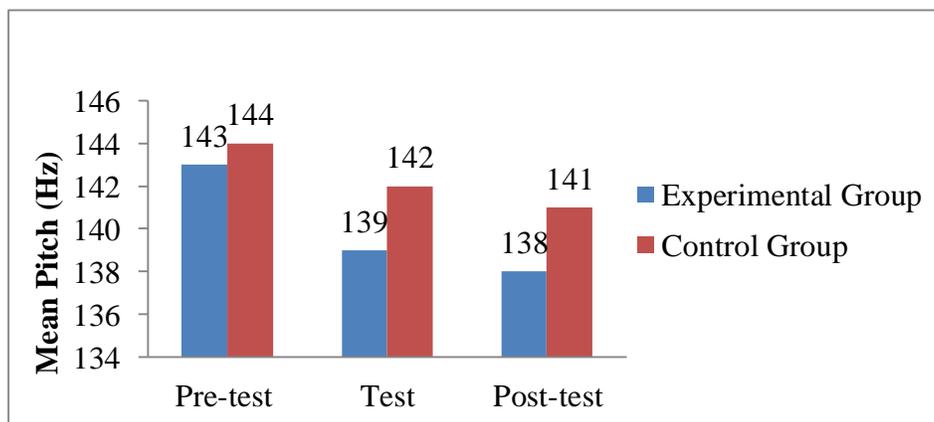
Figure 3. **Experimental and Control Groups' Mean Pitch of Wh-questions**
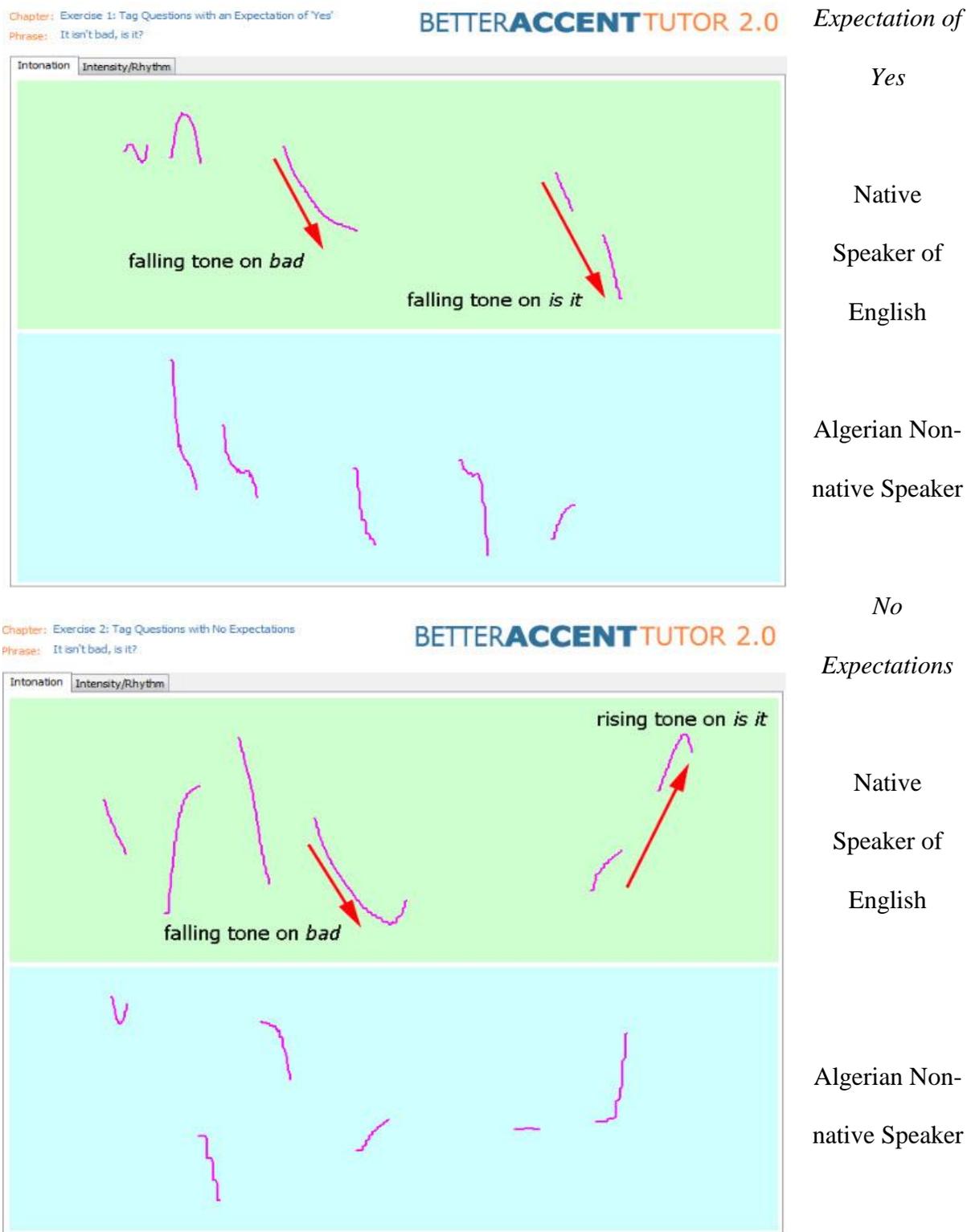
(ii) Tag-questions

Tag-questions are typically used in English to clarify information or expect confirmation (Celce-murcia et al., 1996), and they are characterized with a falling pitch final contour. Also, tag-questions can be pronounced with a rising pitch when asking for information giving the sense of yes/no questions (see figure 4). The following examples illustrate the two types of pitch contours associated with tag-questions.

*Falling Pitch*

It isn't bad, is it?   (Expectation of 'yes')

Rising Pitch

It isn't bad, is it? (No Expectation)

Figure 4. **Tag-questions Pitch Contour (Expectation of Yes/ No Expectations) of Native Speaker vs Algerian Non-native Speaker of English on BetterAccent Tutor**

The results indicate that the majority of Algerian participants performed the English tag-questions with a rising intonation contour. The participants used a rising tone in reading the tag-questions with *expectation of yes* instead of a falling tone giving the sense of yes/no sentences rather than clarification. Furthermore, more than half of the participants ended the tag-questions of *no expectations* with a rising pitch similar to native speakers' productions (see figure 5).
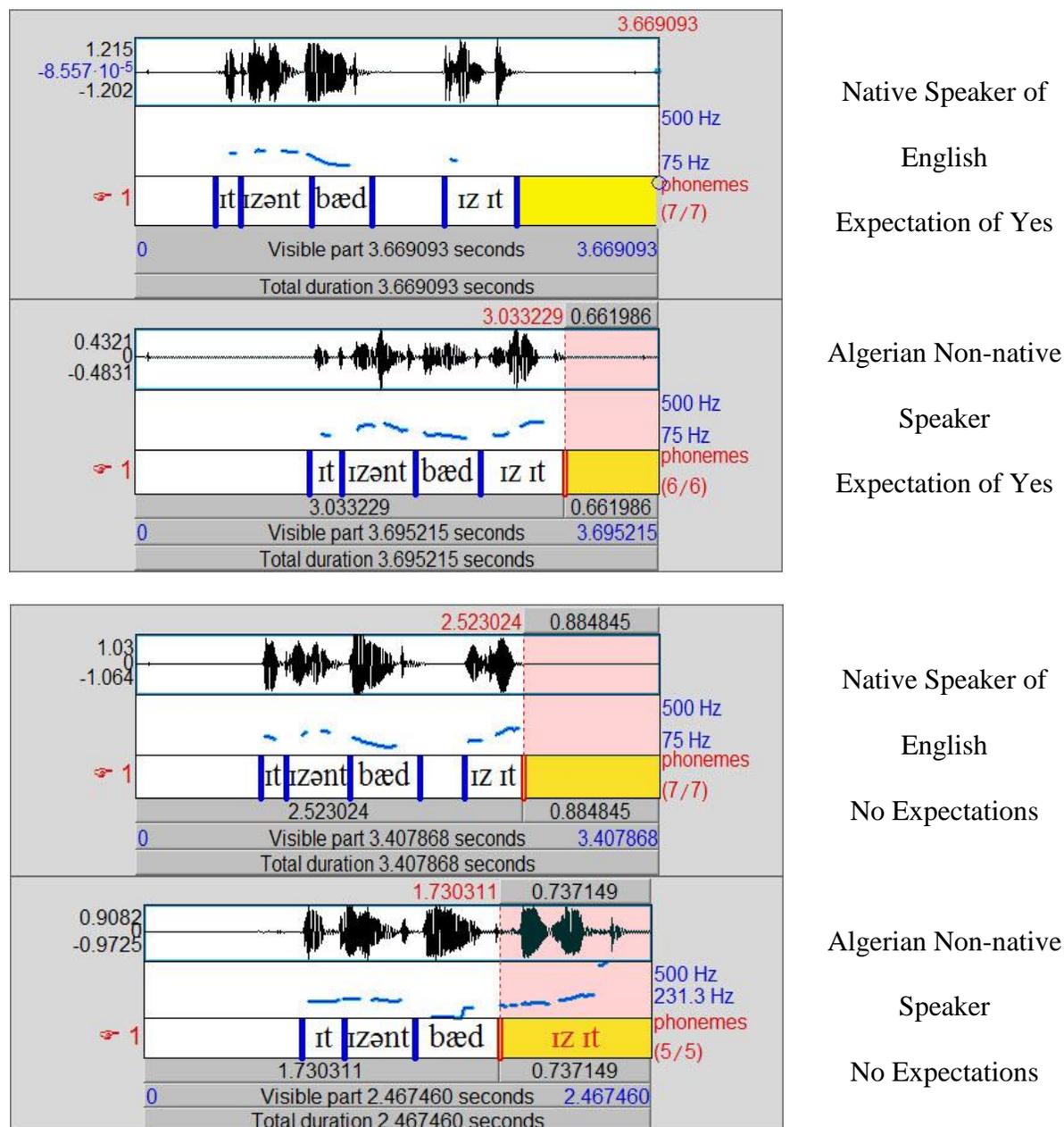

*Figure 5***. Tag-questions (Expectation of Yes/ No Expectations) Pitch Contour of Native Speaker of English vs Algerian Non-native Speaker: Analysis with*Praat*

The results show that Algerian participants'production of *tag-questions* with *expectation of yes* reached a high average pitch of 213 Hz compared to American speakers who ended this set of questions with a lower pitch of 186 Hz. Moreover, American subjects' average pitch of *tag-questions* with *no expectations* was 192 Hz which is slightly equivalent to Algerian participants mean pitch of 194 Hz. Similarly, both native speakers and Algerian participants' intensity average was ranged between 76 db and 77 db, respectively; this is displayed in Table 2.

Table 2. *Tag-questions Mean Pitch*

Participants	Mean Pitch (Hz)		Intensity Average (db)	
	Expectation of Yes	No Expectations	Expectation of Yes	No Expectations
Native Speakers	186	192	77	76
Algerian non-native speakers	213	194	65	77

Observing the average pitch of both experimental and control groups during the pre-test phase demonstrates similar results; however, there was a noticeable progress in the rendition of the tag-questions with *expectation of yes* by the experimental group (see figure 6).
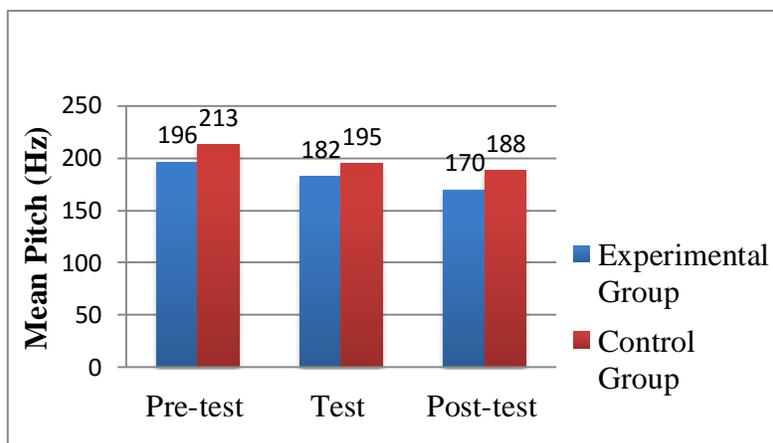
Figure 6. **Experimental and Control Groups' Mean Pitch of Tag-questions**

5. Discussion
(i) Algerian Learners' English Intonational Productions
The combination of visual feedback and auditory feedback helped the experimental group to achieve better results than the control group, in which the latter had access only to auditory feedback. The control group mimicking the native speaker's model, made use of Listen and Imitate technique, thus, combination of both visual and auditory input is found to be effective to Algerian students.

This study demonstrates the similarities and differences between English and Algerian Arabic with certain types of questions in which learners tend to transfer their native language intonational patterns into English.

(ii) Pedagogical Perspectives on UI of BetterAccent Tutor and Praat

The observation of the students' use of speech-based technology software helped to assess at pedagogical level the applicability of BetterAccent Tutor and Praat programs. The UI of BetterAccent Tutor is very accessible to the students. The program is found to be limited to compare the participants' intonational compositions with the target model and the learners' attempts to match the provided native speaker sample may not always be successful.Two renditions of intonational patterns, even of the same speaker, cannot be equivalent. The program provides the learners with explanation about the type of the intonation contour and which constituent receives the tone in the native speaker's chart of speech; however, there is no explicit clarification or corrective feedback regarding the participants' intonational compositions except the visualization of pitch curves. Thus, it is up to the learner to estimate the level of accuracy that is related to matching or non-matching of pitch patterns to those of native speaker's model. The software does not provide instructions on how to align the two produced pitch contours in order to achieve better results. The learners need to compare their productions not only to the target model but also to other regional varieties and native speakers' attitudes, that is, explanation of pitch patterns in accordance to particular situations, and to provide the learner with the environment that encourages integration into natural speech contexts. Also, the program does not focus on: phonetic transcription, waveform and articulatory representation, and video or graphical animation.

Praat allowed the experimental group to analyze and synthesize their speech, and it offered knowledge about their data computation, segmentation, manipulation and labelling (Boersma & Weenink, 2015). The produced speech signals are graphically presented, and the spectrum and speech analysis functions of *Praat* provided the learners with information about features as pitch, intensity, formant, pulses, etc. The participants need to be acknowledged enough concerning the acoustic characteristics of prosody for a better understanding and use of software functionality. The second year phonetics course curriculum is intended to instruct the students about preliminaries regarding the acoustic features at segmental and prosodic levels. Thus, the software is oriented for more advanced level of phonetics and phonological research. The visualization of pitch patterns of longer stretches of speech – discourse – is complicated as the screen window necessitates the instant display of data to be interpreted immediately by the learners. At elementary level, the visualization of intonation contours is helpful, but with longer strings of speech this technique may be difficult to apply at intermediate and advanced stages such as in oral presentations.

Conclusion

The study outcomes are very effective in answering the research questions and the results are found to be compatible with the related literature. The findings assume that visual feedback technologies are not exclusively used in laboratory-based setting but it is possible to test the potential application of these digitalized devices in English language classroom in order to enhance Algerian learners' pronunciation. The similarities and differences between English and Algerian Arabic prosody allow figuring out the sources behind the mispronunciation of English

intonational patterns. *BetterAccent Tutor* and *Praat* were the tools used to generate the intonational compositions; however, *Praat* proved to be more useful to measure the pitch contours and intensity of intonation. Finally, the present study focused on a limited group of learners, thus, instructors may implement feasible pedagogies and maximize the beneficial use of visual feedback paradigm for large group size in L2 classroom.

About the authors:

Naima Hamlaoui is a professor of Applied Linguistics in the English Language Department at the University Badji Mokhtar of Annaba (Algeria). She teaches a range of courses and supervises Doctoral students in English Language and Applied Linguistics. Her research focuses on second language writing. She is also interested in technology and foreign language learning, ESP, translation, and teacher education.

Nawel Bengrait is an assistant professor in the Department of Letters and English Language, University of Guelma 8 Mai 1945, Algeria. She holds a Magister in Language Sciences and is currently completing a Doctoral research in Applied Linguistics. Her research interests are general phonetics, TEFL and technology-enhanced language learning (TELL).

References

Anderson-Hseih, J. (1992). Using electronic visual feedback to teach suprasegmentals.
        System, 20, 51–62.
Anderson-Hseih, J. (1994). Interpreting visual feedback on suprasegmentals in computer
        assisted pronunciation instruction. *CALICO Journal, 11*(4), 5–22.
Bahrami, M., & Bahrami, M. (2012). A Review of Software Architecture for   Collaborative
Software's.*Advanced Materials Research, 433*, 2372-2376.
Benrabah, M. (1987).*The Intelligibility of Algerian Speakers of English: a Phonetic/
        Phonological Study* (Unpublished doctoral dissertation). London University College.
Boersma, P., & Weenink, D. (2015).*Praat: Doing phonetics by computer* [Computer
        software]. Retrieved from
        http://www.praat.org.
Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English.* Stanford, CA:
        Stanford Press.
Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation: A reference
        for teachers of English to speakers of other languages*. New York: Cambridge
        University Press.
Chun, D. (1989). Teaching tone with microcomputers. *CALICO Journal, 7*(1), 21–47.
Chun, D. (1998). Signal analysis software for teaching discourse intonation.*Language Learning
& Technology, 2*(1), 61–77. Retrieved from http://llt.msu.edu/vol2num1/article4/index.html
Chun, D. (2002). *Discourse intonation in L2: From theory and research to practice*.
        Amsterdam, The Netherlands: John Benjamins.
de Bot, K. (1983). Visual feedback of intonation: Effectiveness and induced practice
        behavior. *Language and Speech, 26*, 331–350.

Derwing, T. M. & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39,*379-397.

Derwing, T. M. & Munro, M. J. (2010).Utopian goals for pronunciation teaching. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 24–37). Ames, IA: Iowa State University.

Hardison, D. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology, 8*(1), 34–52. Retrieved from http://llt.msu.edu/vol8num1/hardison/default.html

Lord, G. (2010). The combined effects of instruction and immersion on second language pronunciation.*Foreign Language Annals, 43*(4), 488–503.

Mehrdad, H. (2007). *Art of Teaching*. Tehran: Ravan.

Molholt, G. (1988). Computer assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly, 22*(1), 91–111.

Odisho, E. Y. (2005). *Techniques of teaching comparative pronunciation in Arabic and English*. New Jersey: Gorgias Press.

Olson, D. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in L2 pronunciation instruction. *Hispania, 97*(1), 1–25.

Pressman, R., S. (Ed.). (2009). *Software engineering: a practitioner's approach* (7[th]ed.). New York: McGraw-Hill.

Setter, J., & Jenkins, J. (2005). State of the art review: Pronunciation. *Language Teaching, 38*, 1–17.

Shneiderman, B. (Ed.). (1997). *Designing the User Interface* (3[rd]ed). Addison-Wesley.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An   experimental-phonetic approach to speech melody*. Cambridge: Cambridge University   Press.

Wang, X., & Munro, M. (2004). Computer-based training for learning English vowel contrasts. *System, 32*, 539–552.

Appendix

English Questions

1.   Wh-questions:
   - What's this?
   - Where did you go?
   - Who is this woman?

2.   Tag-questions
   - The weather is fine today, isn't it?
   - It isn't bad, is it?
   - She has a sister too, hasn't she?