ORIGINAL ARTICLE　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

The acquisition of question intonation by Vietnamese learners of English

Anh-Thư T. Nguyễn[1*] and Đích Mục Đào[2]

* Correspondence:
thunguyen1972@gmail.com
[1]Mountain Creek, Sunshine Coast,
Australia
Full list of author information is
available at the end of the article

Abstract

This paper examines the intonation of English statements and questions produced by Vietnamese speakers at two differing levels of proficiency. The goal of the study is three-fold: (1) analysing the final tunes and the prosodic structure observed in information-seeking questions, namely Yes-No question, Or-question, Tag-question and Wh-question, (2) evaluating which characteristics of the L2 English intonation can be clearly derived from the observation of the data, and (3) whether the L2 English intonation patterns are transferred from Vietnamese. A data set of 25 sentences that included 5 statements and 20 information-seeking questions were constructed. Ten native Australian English speakers as a control group and 20 Southern Vietnamese speakers of English (10 beginners and 10 advanced speakers) were recorded. The final tunes (the direction of the final F0 contours) of the sentences were analysed. The result showed that while the advanced speakers of English mostly produced intonation patterns that are typically used by native English speakers, beginning speakers of English used a variety of tunes, several of which are deviate from the native-like standard and clearly transferred from the tone contours in Vietnamese. The findings of this study have an original and significant contribution to the literature because it investigated into the prosodic transfer of intonation patterns between two typologically distinct languages: English, a stress accent language and Vietnamese, a contrastive contour tone language and has implications for intonation teaching.

Keywords: Prosodic transfer, Intonation, English, Vietnamese, Second language acquisition

Introduction

Research on interlanguage intonation has shown that the intonational patterns observed in learners' productions are often influenced by their first languages (L1 s) (Mennen, 2004; Jilka, 2000; Rasier and Hiligsmann, 2007). Therefore, the notion of L1 transfer is often used to account for the observed patterns. As Mennen (2007) has pointed out, transfer may apply at the phonological as well as at the phonetic level. Transfers at the phonological level result from differences in the metrical structure or the tonal inventory. For example, in a study on the intonation of tag questions in English, Ramírez and Romero (2005) have shown that Spanish speakers of English use rises at the end of the question tag for confirmation request, whereas native English speakers will use falls; these patterns were thus analysed as resulting from a phonological transfer. By contrast, transfers at the phonetic level occur when an identical phonological form differs in the way it is phonetically implemented in both languages.

For instance, differences in the temporal alignment of pitch accents may be a case of phonetic transfer (Mennen, 2004). As pointed out by Mennen (2015), the distinction between different types of transfers or deviations is of great help to study interlanguage intonation.

Previous studies have shown that second language (L2) speakers have difficulties with selecting appropriate intonation contours for sentences (He et al. 2012) and that their usage of pitch can show cross-linguistic influence (Gut 2009). One of the first studies that found the cross-linguistic influence on L2 intonation was by Wennerstrom (1994), who compared the pitch height at the end of a yes–no question in a reading passage produced by native English speakers to that produced by Thai, Japanese and Spanish L2 speakers of English. Her results show that the Thai native speakers did not mark the question with a high ending rise as the native English speakers did, while the other two learner groups produced rises like the native speakers. She speculated that these differences between L2 speakers might be due to L1 influences, and specifically 'the fact that in Thai, a tone language, pitch functions to distinguish lexical rather than discourse meaning'. Goh (2001)'s study reports a high frequency of rising tones in questions produced by both Malay and Singaporean speakers of English, whereas it was found by Lim (2002) that while the overall intonation contours of the question *Where are you going*? was similar among Malay, Indian and Chinese Singaporeans, there were differences in pitch alignment on the final lexical item. Whilst all three groups displayed a final rise-fall contour, the F0 peak was found to occur much later for the Malay speakers. Even though Lim does not suggest that this is due to the influence of Malay, she does argue that this phenomenon may be a distinguishing factor of interethnic variation in Singapore English. In another study, Lim (2009) demonstrated that ethnically Chinese Singaporeans produce tones from the tone language Chinese on some particles when speaking English. Moreover, their intonation in English consists of sustained tone movements rather than pitch contour movements. Similarly, it was proposed by Gut (2005) that Nigerians who have a tone language as their first language show cross-linguistic influence in their L2 English: First, it has a reduced inventory of pitch movements compared to British English; and second, high and low pitch on syllables seem to be used mainly for the function of accentuation. Furthermore, the domain of pitch appears to be the word rather than the utterance in Nigerian English. Ramirez Verdugo (2002) found that Spanish L2 speakers of English show little difference in their use of intonation in read out wh-questions and yes–no questions, marking the former with falls and the latter with rises like native English speakers. However, the L2 speakers overused rises in tag questions compared to native speakers. This was also found by Hewings (1995), who asked English native speakers as well as Korean, Indonesian and Greek learners of English to read out a scripted dialogue containing one tag question. While the native speakers all produced a fall, ten out of the twelve L2 learners produced a rise. Similarly, in the wh-question *Which one will you go for*? five learners produced a rising pitch movement.

This paper examines the intonation of English statements and questions produced by Vietnamese speakers at two differing levels of proficiency. The goal of the study is three-fold: (1) analysing the final tunes and the prosodic structure observed in information-seeking questions, namely Yes-No question, Or-question, Tag-question and Wh-question, (2) evaluating which characteristics of the L2 English intonation can

be clearly derived from the observation of the data, and (3) whether the L2 English intonation patterns are transferred from Vietnamese.

The intonation of statement and questions in Vietnamese and English

To distinguish between questions and statements, the use of high pitch has been claimed to prevail cross-linguistically. However, the implementation of the high pitch feature may differ across languages and dialects. Terminal rises for questions are probably the most widespread. Over 70% of the languages in the world are estimated to have interrogative intonation contours, which end with rising pitch (Bolinger 1978). Yet, there are many cases of languages that contradict the putatively universal pattern of rising questions (van Heuven and van Zanten 2005).

In English as well as many other non-tonal languages interrogative intonation has a rising pitch contour whereas declarative intonation has a falling pitch contour. This phenomenon has been widely studied and was generalized as the Strong Universalist Hypothesis (Ladd, 1981), according to which pitch rising indicates a question and pitch falling indicates a statement. Various researchers have suggested that specific pitch movements are associated with different syntactic types of questions. According to Halliday (1967), Ladd, (1996), Wells (2006), Halliday and Greaves (2008) and O'Connor and Arnold (1973), wh-questions typically have a falling tone while yes–no questions have a rising tone. Tag questions have a rising intonation when the speaker is genuinely asking for information, but a fall when the speaker expects that the other speaker will agree (Wells 2006). An alternative question (i.e., or-question) denotes choice and is spoken with a rising intonation in the first part and a falling intonation in the second part. These claims about the typical intonation of these different types of question have been largely validated in empirical studies (Geluykens 1988; Hirschberg 2000; Hedberg and Sosa 2002; Hedberg et al. 2004). In American English telephone conversations between friends, it is found by Hedberg et al. (2004) that wh-questions were associated with a falling tone in 82% of all cases. The wh-questions that were produced with a rise were interpreted to signal that the speakers know that they should be aware of the answer but forgot it. Yes–no questions with verb inversion were produced with a rising tone in 80% of all cases. Nevertheless, Geluykens (1988) analysed spontaneous conversations in standard British English and found that only 52.5% of them were produced with a rising pitch movement. Hedberg et al. (2004) proposed that yes–no questions that were produced with falls indicated the speaker's relative certainty of the answer.

In a tone language such as Chinese or Vietnamese, however, the difference between declarative and interrogative intonation is much more complicated because of tone and intonation interaction. For example, interrogative intonation with a final rising tone has a rising end, which is similar to English, whereas that with a final falling tone often has a falling end (Yuan et al., 2002). Studies in Chinese intonation show that the diverse surface patterns can be accounted for by two consistent features: 1. Interrogative intonation has a higher phrase curve than declarative intonation; 2. Sentence final syllables have more careful intonation and wider pitch swings in interrogative sentences (Yuan et al., 2002; Zeng, Martin and Boulakia, 2004). In Vietnamese, previous studies have shown that there are a number of acoustic correlates for realizing sentence modalities, predominantly based on global f0 and intensity and local sentence-final f0 (Đỗ et al.

1998; Nguyễn and Boulakia 1999; Vũ et al. 2006; Brunelle et al., 2012). It has been found that declaratives tend to have a slight overall f0 declination (Đỗ et al. 1998; Nguyễn and Boulakia 1999). Interrogatives are found to have a high overall range (Hoàng 1985), or a high range and a rise starting much before the sentence final question marker (Đỗ et al. 1998; Nguyễn and Boulakia 1999). However, one study suggests that the range difference between statements and questions is insignificant and that the rise of the questions is largely located towards the end of the sentence-final question particle (Vũ et al. 2006). In a recent study on Northern Vietnamese, Brunelle et al. (2012) look at the role of intonation in the realization of communicative functions in Northern Vietnamese. Their results show that there are a number of acoustic strategies for realizing communicative functions, predominantly based on global f0 and intensity and local sentence-final f0. Nevertheless, the acoustic properties associated with communicative functions are variable, or even absent in individual speakers. In another recent study on Southern Vietnamese, Dao and Nguyen (acoustic correlates of statement and question intonation in Southern Vietnamese, submitted) found that communicative functions (i.e., declarative vs. various interrogative forms) are conveyed by the global values of the three acoustic correlates: duration, f0 and intensity. First, the f0 height of the entire sentence is affected (mean f0 being lowest in statement and higher in all forms of questions). Second, compared to statements, questions are characterised by a faster tempo. Third, the intensity of the entire sentence is found to be raised in questions.

In addition to the global f0 differences between declarative and interrogatives, local f0 and duration effects are also found. First, final syllables are longer, have fuller contours, wider pitch swings and/or heightened f0 end in interrogative sentences. Particularly, heightened f0 end is found for level and falling tones in questions which normally have a falling end in statements. Second, in Or- questions, the words/syllables before and after the particle *hay* (or) are lengthened, have larger f0 range and have a fuller tonal shapes than the other syllables in the questions. Particularly, dropping and curve tones of the target words/syllables have wider pitch swing: it falls deeper and rises higher. These local effects are consistent with findings in previous studies on Vietnamese (Đỗ et al. 1998; Nguyễn and Boulakia 1999; Vũ et al. 2006; Brunelle et al., 2012) and Chinese – another tone language (Yuan et al., 2002; XiaoLi, Martin and Boulakia, 2004). Furthermore, the presence of a final rise in questions, which could be interpreted as a high boundary tone, is found in all speakers despite it occurs more frequently in some speakers and less in other speakers. Although there appears to be a great deal of variation in the use of local f0 rises sentence-finally, this can be interpreted as an optional use of intonational high boundary tones, consistent with Brunelle et al. (2012)'s results for Northern speakers.

Vietnamese acquisition of English prosody

There has not been much research that focused on Vietnamese acquisition of English prosody. Nguyễn and Ingram (2005) examined the transfer of tonal acoustic correlates in Vietnamese learners' production of English word stress. More specifically, the study examined acoustic features that native and non-native speakers (Vietnamese learners of English) use to differentiate stressed from unstressed syllables in noun-verb pairs (e.g.,

as in the words *re*cord vs. re*cord*). The results indicated that Vietnamese learners of English (both proficiencies) utilised F0 and intensity correlates similarly to native speakers. A major difference was the lack of vowel and syllable duration cues in the beginning learners' production. In another study on prosodic transfer effects in the production and perception of three English stress patterns (broad-focus noun phrase, narrow-focus noun phrase and compound) at the level of word and phrase prosody by Vietnamese learners of English, Nguyễn et al. (2008) found that Vietnamese speakers had no problem in manipulating contrastive levels of f0 and intensity on accent-bearing syllables but failed to realize the timing contrast between compound words and phrases and the syntagmatic contrast of accent in larger units such as polysyllabic words or phrases, as evidenced by their failure to deaccent the second element of the compound and narrow-focus patterns. Nevertheless, the advanced speakers' ability to compress the constituents of the compounds and to deaccent the final nouns shows the effect of language learning/experience on prosodic acquisition. At the connected speech level, Nguyễn and Ingram (2004) found that the transfer of many segmental, prosodic, timing and syllable structure from Vietnamese phonological system such as checked stop, implosive stop, vowel quality, suppression of vowel reduction and checked tones was also evidenced in advanced Vietnamese speakers of English. Particularly, the suppression of vowel reduction in unstressed syllable and the lengthening of many unstressed vowels/ function words were projected under sustained high tones in unstressed syllables in spite of an advanced level of English proficiency. In a recent study, Nguyen (F0 patterns of tone versus non-tone languages: the case of Vietnamese speakers of English, submitted) aimed to find out whether F0 patterns of L2 English produced by Vietnamese speakers are different from those of English and whether the prosodic deviation is transferred from Vietnamese. 10 native/L1 Australian English speakers, 20 Vietnamese speakers of English (10 beginners and 10 advanced speakers) and a control group of 4 native/L1 Vietnamese speakers were included. The F0 profiles (F0 maximum, F0 minimum, F0 range, F0 mean and F0 standard deviation at three levels: utterance, syllable and phoneme) were obtained from a set of 10 English sentences and 20 Vietnamese utterances. The results showed that F0 patterns of beginning-level L2 English are systematically different from those of native English speakers, which can be transferred from their native tone language. Nevertheless, the advanced speakers' ability to produce native-like F0 patterns indicates the effect of language learning/experience on prosodic acquisition.

Method

Linguistic materials

In order to pursue the aim of this study, we constructed a data set of 25 sentences that include 5 statements and 20 information-seeking questions (5 Yes-No questions, 5 Or-questions, 4 Tag-questions and 6 Wh-questions). The list of sentences is presented in Additional file 1.

Participants

We recorded 10 native Australian English speakers as a control group and 20 Southern Vietnamese speakers of English (10 beginners and 10 advanced speakers). The control group of 10 English speakers consisted of 5 males and 5 females who were students at

Macquarie University, Australia. Their age ranges from 21 to 30 (mean age: 25.4). The advanced group of Vietnamese speakers of English included 7 postgraduate students at a university and 3 high school (year 11) students (5 males and 5 females) in Australia. They were in the age range 16–32 (mean age: 24.5). Their length of residence in Australia varied from 6 months to 2 years. All of them achieved a proficiency level of 'competent' and 'good user of English language' since they had at least an average band score of 6.5 on the IELTS test (International English Language Testing System - a 9-band proficiency test of English on four skills: listening, speaking, writing and reading). All of the subjects started learning English at the age of twelve with the Grammar Translation method during the secondary and high school. However, they were exposed to communicative English learning for some time at Foreign language centres in Vietnam and English classes in Australia before entering high school or university. Their English proficiency can be said to be advanced or high level. The 10 beginners (5 males and 5 females) were students at University of Social Science and Humanities of Ho Chi Minh City, Vietnam who were in the age of 18–23. The beginners had all started learning English at the age of 12 (in secondary school) with the grammar translation method, which focuses on vocabulary and grammar learning. Their English proficiency can be said to be at low level. The purpose of including advanced Australian learners is to examine if learners at a more advanced level, who were exposed to Australian English while living in Australia, can accommodate to Australian native-like intonation patterns or not.

Procedures

The 25 sentences were randomized and presented to the subjects in print. Before the recording, they had time to familiarize with the sentences including statements and questions. To elicit the declarative statements, the subjects were instructed to speak the five declarative sentences to the researcher first as if they were telling the researcher about themselves (all five sentences were started with the first person "I" or "we"). To elicit information-seeking questions, the participants were then requested to ask the researcher 20 remaining questions for information (17/20 questions had the second person pronoun "you" and 3/20 questions had the third person pronoun "he" or "she"). Since the participants did not know anything about the researcher in advance, their questions can be considered as genuine information seeking. We assume that after being familiar with the sentences, the subjects will speak them more naturally. The sentences were recorded at 44.1 kHz using an external microphone connected to a laptop and the Praat software (Boersma & Weenink, 2009).

Analysis

The 750 utterances from the corpus (25 sentences × 3 groups × 10 speakers) were all annotated prosodically in order to allow a comparison of the productions across three speaker groups. Focus was given to the form of the final tune (i.e., the pitch movement that goes from the last pitch accent to the boundary tone) occurring at the end of questions and statements. Hedberg et al. (2004) 'es examination of the interface between intonation and meaning of English questions in real speech reveals that the prosodic structure, specifically, the direction of the final contour, is fundamental to interactional

pragmatic meaning. Therefore, in this study, the final tunes are encoded. The symbols used are F for a simple fall, R for a simple rise, HF for high fall and FR for fall-rise according to the British tradition but with a limited tone inventory, following Hirst (2005) and Gussenhoven (1984). The symbol L (for a level tune) is used for encoding the speech of Vietnamese learners. The symbol FD (falling and deaccenting: i.e., the suppression of accents on any following words.) is also used to label native English speakers' speech. The statements, Yes-No questions and WH-questions all had only one tune at the end, while the Or-questions and the tag-questions had two tunes each (see the Additional file 2 for a sample of pitch contours and their tunes encoding). The Or-question had one tune in the first part before the conjunction "or" and another tune at the end. In the tag-questions, one tune came after the statement (usually falling) to which the tag ending is attached and another tune came on the tag ending.

Sentences produced by the speakers were acoustically processed and analyzed in Praat (Boersma and Weeninck, 2009). Each sentence was segmented into syllables and acoustic analysis was carried out using PRAAT with visual pitch contour to decide the label of the tunes.

In order to see whether the differences in the choice of the final tunes across the fixed variables studied in this research were statistically significant, we calculate the proportion (expressed in percentage) of each tune (R, F FR, HF, L, FD and so on) for each sentence types (statements and four question types) for each speaker. We then constructed mixed effect models on the tune proportion for each sentence/question type. The fixed effects were groups (3 speaker groups: native English, advanced and beginner of English) and tunes. The random effect was speakers (30 speakers). The data analysis was carried out using the SPSS program. The results were reported in Table 1.

Results

As shown in Table 1, the mixed effect model results showed that there was no significance for the main factor Groups but the factor Tunes and the interaction Groups x Tunes reached significant level ($p < 0.05$). Therefore, in the following sections, we only examined the interaction effects Tunes x Groups for each sentence/question types separately.

Statements

As shown in Fig. 1 and Additional file 2, the native English speakers produced statements with either a falling final tunes (F) or falling and deaccenting (FD). The advanced groups also produced mostly falling final tunes (96%) while some speakers produced a

Table 1 Mixed effect model results

	Statement	Yes-No question	Or-question	Tag-question	WH-question
Groups	$F_{(2,35)} = 2.7$, $p = 0.08$ ns.	$F_{(2,32)} = 0.9$, $p = 0.3$ ns.	$F_{(2,56)} = 0.2$, $p = 0.7$ ns.	$F_{(2,34)} = 1$, $p = 0.36$ ns.	$F_{(2,35)} = 0.1$, $p = 0.8$ ns.
Tunes	$F_{(4,32)} = 14$, $p < 0.0001$	$F_{(5,36)} = 7.2$, $p < 0.0001$	$F_{(10,54)} = 1$, $p = 0.4$ ns.	$F_{(5,34)} = 5.7$, $p < 0.01$	$F_{(4,41)} = 3.8$, $p < 0.02$
Groups x Tunes	$F_{(2,35)} = 3.8$, $p < 0.05$	$F_{(2,35)} = 3.5$, $p < 0.05$	$F_{(8,56)} = 4.1$, $p < 0.04$	$F_{(4,34)} = 3.6$, $p < 0.05$	$F_{(4,37)} = 4.5$, $p < 0.03$

Fig. 1 Mean proportion of final tunes of statements by speaker groups. Y-axis: mean percentage. The symbol * means significant at $p < 0.01$
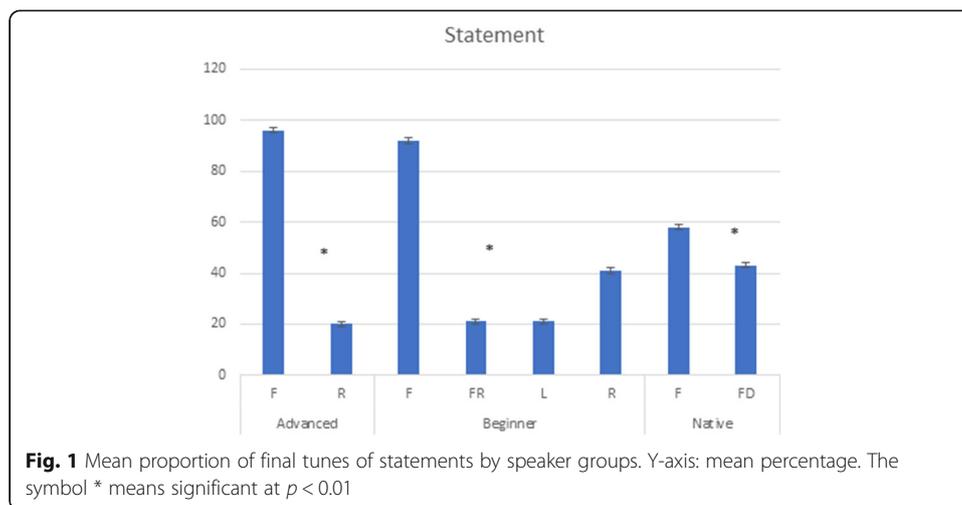
rising tune (20%). By contrast, apart from the falling tunes (92%), the beginner produced fall-rise(FR), rise(R) and level (L) tunes.

Yes-no questions

As shown in Fig. 2, the native English speakers produced Yes-No questions not only with rising final tune (80%) but also with either a falling tune or deaccenting. Advanced speakers also produced this question types mostly with a rising tunes (86%) and occasionally with a falling tune (24%). In contrast, beginners produced a variety of tunes (R, F, FR, L, and RF).

Or-question

Figure 3 and Additional file 2 showed that native English speakers produced the Or-question mostly with a rise in the first part and either a fall or deaccent in the second part (R-F: 58%, R-FD: 53%). Some of them also produced other patterns such as R-R, F-F, F-FD, and F-R. On the contrary, the advanced and beginning speakers of English produced the R-F pattern less frequently than native speakers

Fig. 2 Mean proportion of final tunes of Yes-No questions by speaker groups. Y-axis: mean percentage. The symbol * means significant at $p < 0.01$
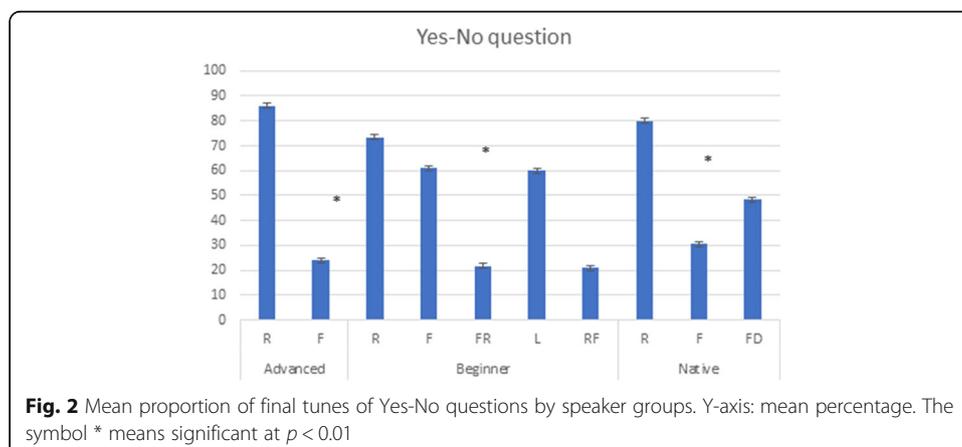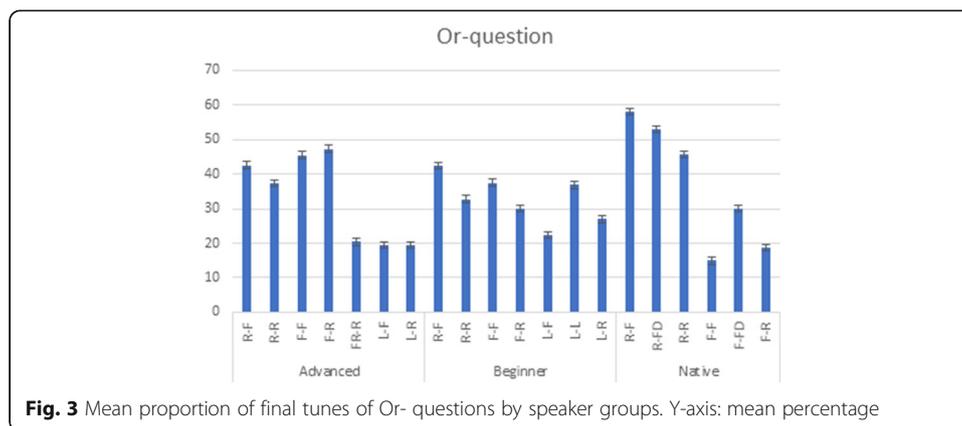
Fig. 3 Mean proportion of final tunes of Or- questions by speaker groups. Y-axis: mean percentage

(advanced:43%, beginner: 42%), while they spoke this question type with a variation of patterns (R-R, F-F, F-R, L-F, L-R).

Tag-questions

As shown in Fig. 4 and Additional file 2, the native English speakers produced the tag-question mostly with a fall at the end of the statement while they produced the tag ending with a variety of tunes such as FD (83%), R (78%) and simple F (38%). The advanced speakers also mostly used a falling tune for the statement and a rise tune for the tag (F-R: 93%). In contrast, the beginners used different tonal patterns for this question type (F-F, L-L, L-R, and R-R).

WH-questions

Figure 5 and Additional file 2 showed that native English speakers mostly used either a falling tune or deaccenting at the end of the WH-questions. Nevertheless, some of them used a fall-rise or rising tunes. The advanced speakers also mostly produced a falling tune for this question type. On the contrary, the beginners used rising tunes more frequently (R: 61%) in addition to other tunes such as F (57%) and L(17%).
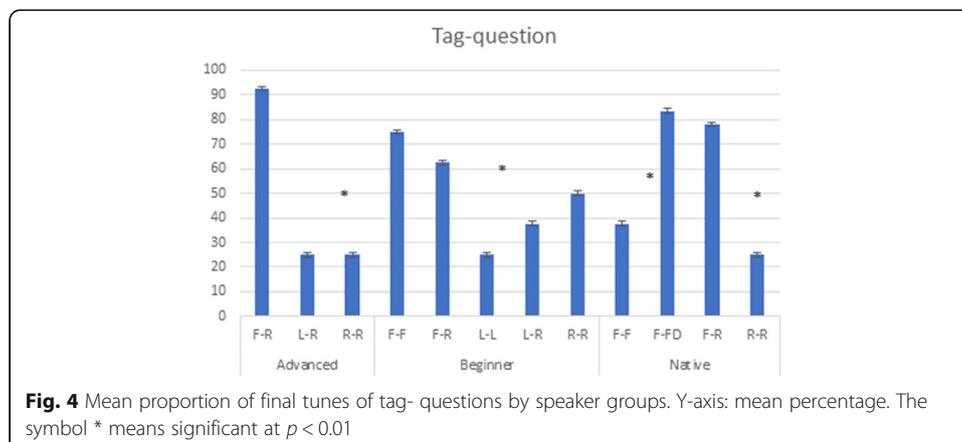
Fig. 4 Mean proportion of final tunes of tag- questions by speaker groups. Y-axis: mean percentage. The symbol * means significant at $p < 0.01$

Fig. 5 Mean proportion of final tunes of Wh- questions by speaker groups. Y-axis: mean percentage. The symbol * means significant at $p < 0.01$
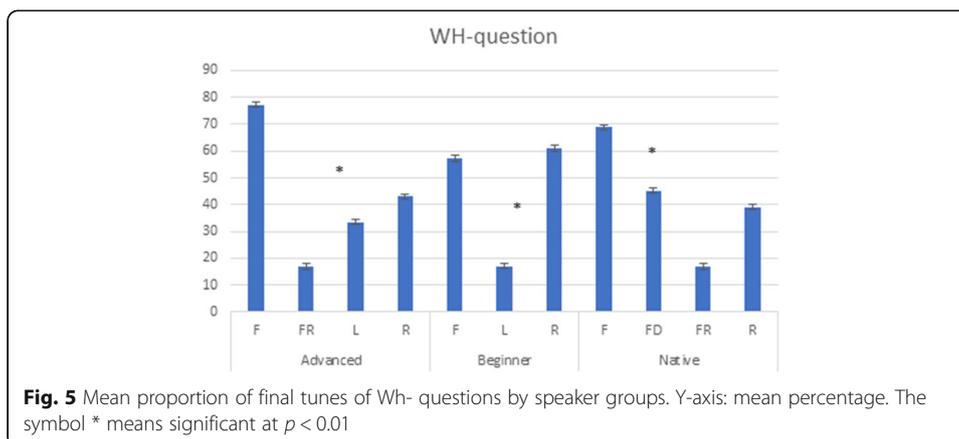
Discussion and conclusion

In this section, we summarize and discuss the results by addressing the three research questions raised in Introduction section.

First, *the analysis of the final tunes and the prosodic structure observed in English information-seeking questions, namely Yes-No question, Or-question, Tag-question and Wh-question* shows that while the advanced speakers of English mostly produced tonal patterns that are typically used by native English speakers (such as a F for statements, a R for Yes-No questions, a R-F for Or-questions, a F-R for tag-questions and a F for Wh-questions), beginning speakers of English used a variety of tunes, several of which are deviate from the native-like standard such as L, L-R, L-F, L-L. Furthermore, apart from typical falling and deaccenting final tunes, some native English speakers' Wh-questions were produced with a rise. This can be interpreted to signal that the speakers might know that they should be aware of the answer, as reported by Hedberg et al. (2004). In addition, several Yes–no questions with verb inversion in this study were produced with a falling tone or deaccenting by native Australian English speakers, in line with data by Geluykens (1988) on standard British English in which only 52.5% of them were produced with a rising pitch movement. Hedberg et al. (2004) proposed that yes–no questions that were produced with falls indicated the speaker's relative certainty of the answer. Moreover, the native English speakers in this study used many falling or deaccenting final tunes for the final tag ending of the tag-questions, suggesting that they might know the answer and expected the other speaker to agree (Wells 2006).

Second, *the evaluation of which characteristics of the L2 English intonation can be clearly derived from the observation of the data* shows the following patterns:

1.) Both groups of L2 speakers of English used the rising tunes at the end of the Wh-questions significantly more frequently than the native English speakers. This can also be attributed to the transfer of L1 patterns. In Dao and Nguyen (submitted), Southern speakers of Vietnamese tend to raise F0 at the end of questions, particularly the final words/syllables. This happens to all speakers in their study. This tendency occurs not only in yes-no question without particle but also in all other kinds of questions with particles. For example, the final words of the Wh-questions had the falling tone (i.e., *nào(which)* and *ngò(coriander)*), which should have a falling

contour, but they were produced with a rising F0 end. In addition, the tendency to mark Wh-questions with a rise appears to be a common pattern with non-native speakers of English, as observed by Gut and Pillai (2015).

2.) Both groups of Vietnamese speakers of English tended to overuse rises in tag questions compared to native speakers. This was also found by Ramirez Verdugo (2002) for Spanish speakers of English and by Hewings (1995) for Korean, Indonesian and Greek learners of English.

3.) While the native English speakers tend to deaccent the final part of the utterance when it has a falling tune, consistent with the general patterns for standard English (Ladd, 1978), neither of the L2 learner groups accommodate to this feature. This, on the one hand, is in line with the tendency for other variety of English dialects. For instance, Singaporean speakers tend not to deaccent information which is repeated at the end of the utterances as reported by Low (1994). On the other hand, this result is indicative of the transfer of L1 Vietnamese prosodic structure of a lexical tone language in which every syllable is specified for a tone and there was neither toneless syllable nor tonal reduction.

Third, *the examination of whether the L2 English intonation patterns are transferred from Vietnamese* indicates that transfer from L1 Vietnamese apply at the phonological as well as at the phonetic level. Transfers at the phonological level is shown in the differences in tonal inventory between native English speakers and L2 Vietnamese speakers. For example, L2 speakers of English tend to use rises at the end of the Wh-questions whereas native English speakers used falls. In addition, transfer at the phonetic level is indicated by the fact that Vietnamese speakers of English failed to deaccent the final falling tunes and used a variation of L1 tonal contours for their English intonation. These results are consistent with Mennen (2004, 2015).

In addition, the F0 patterns produced by beginning learners (such as L, L-R, L-F, L-L) are clearly transferred from the tone contours in Vietnamese, consistent with the transfer from lexical tones to intonation contour in English by other non-native English speaker groups such as Chinese Singaporeans (Lim, 2009) and Nigerians (Gut, 2005).

In conclusion, this study conducted a systematic investigation on F0 patterns of the final tune of English statements and different question types by Vietnamese speakers. The result has shown that intonation patterns of beginning-level L2 English produced by Vietnamese speakers are systematically different from those of native English speakers, which can be transferred from their native tone language. Nevertheless, the advanced speakers' ability to produce native-like intonation patterns indicates the effect of language learning/experience on on F0 pattern acquisition, lending further support for the L2 prosody acquisition in previous studies (Nguyễn and Ingram, 2005; Trofimovich and Baker, 2007; Nguyễn et al., 2008). The findings of this study have an original and significant contribution to the literature because it investigated into the prosodic transfer of intonation patterns between two typologically distinct languages: English, a stress accent language and Vietnamese, a contrastive contour tone language.

The implication of this study is that the obtained data can help teachers and students to identify the problems that ESL Vietnamese speakers can have when learning English as a L2, especially with regards to intonation. We strongly believe that one effective way to tackle these difficulties is an explicit teaching of English intonation using visual

and audial display of English intonation for each question type with the help of electronic devices that provide visual feedback of intonation (e.g., De Bot, 1982; Suzuki et al., 1989). That is, a target sentence is presented auditorily to the learner (by a teacher or the computer software) while at the same time the pitch contour of this utterance is shown on the screen. The learner is then asked to imitate the target sentence. After the imitation has been produced, the example and its incitation can be compared visually. The information the learner receives in this way may enable him or her to produce a better pitch contour. Comparing the speech and the visual signals simultaneously may enable the learner to see how the pitch contour develops as a function of time and how parts of the sentence are related to parts of the contour. Thus, these devices not only provide auditory but also visual feedback: the learners can hear as well as see the pitch contour in the target language as well as their success at imitating it. This additional feedback should enable learners to imitate the target sentence more accurately, because they are made aware of mistakes which they might not notice otherwise, and which they can try to correct in subsequent learning trials.

Additional files

Additional file 1: English statements and questions. (DOCX 14 kb)
Additional file 2: F0 contours and prosodic labelling of the final tunes Statements. (DOCX 60 kb)

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Mountain Creek, Sunshine Coast, Australia. [2]University of Social Sciences and Humanities, Vietnam National University, Ho Chi Minh City, Vietnam.

References
Boersma, P.& Weenink, D. (2009). Praat: Doing phonetics by computer (version 4.6.09). Computer program. Retrieved June 24, 2007, from http://www.praat.org/ California, Los Angeles.
Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed), Universals of human language, vol. 2: Phonology (pp. 471–524). Stanford: Stanford University Press.
Brunelle, M., Ha, K. P., & Grice, M. (2012). Intonation in northern Vietnamese. *The Linguistic Review, 29*(1), 3–36.
De Bot, C. (1982). *Visuele Feedback van Intonatie.* Doctoral dissertation: Nijmegen University.
Đỗ, T. D., Trần, T. H., & Boulakia, G. (1998). Intonation in Vietnamese. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A Survey of Twenty Languages* (pp. 395–416). Cambridge: Cambridge University Press.
Geluykens, R. (1988). On the myth of rising intonation in polar questions. *Journal of Pragmatics, 12*, 467–485.

Goh, C. C. M. (2001). Discourse intonation of English in Malaysia and Singapore: Implications for wider communication and teaching. *LREC Journal, 32*, 92–105.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accent*. Dordrecht: Foris.

Gut, U. (2005). Nigerian English prosody. *English World-Wide, 26*, 153–177.

Gut, U. (2009). *Non-native speech: a corpus-based analysis of phonological and phonetic properties of L2 English and German. Frankfurt A. M.* Peter Lang.

Gut, U., & Pillai, S. (2015). The question intonation of Malay speakers of English. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations* (pp. 51–70). Berlin: Springer.

Halliday, M. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

Halliday, M. A. K., & Greaves, W. S. (2008). *Intonation in the grammar of English*. London: Equinox.

He, X., van Heuven, V., & Gussenhoven, C. (2012). The selection of intonation contours by Chinese L2 speakers of Dutch: Orthographic closure vs. prosodic knowledge. *Second Language Research, 28*(3), 283–318.

Hedberg, N., & Sosa, J. (2002). The prosody of questions in naturEal discourse. *Proceedings of Speech Prosody*, 375–378.

Hedberg, N., J. Sosa, & L. Fadden. (2004). Meanings and configurations of questions in English. *Proceedings of Speech Prosody,* 309–312. Nara.

Hewings, M. (1995). Tone choice in the English intonation of non-native speakers. *International Review of Applied Linguistics in Language Teaching, 33*(3), 251–266.

Hirschberg, J. (2000). A corpus-based approach to the study of speaking style. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 271–311). Dordrecht: Kluwer.

Hirst, D. (2005) Form and function in the representation of speech prosody. In Hirose, K., Hirst, D. & Y. Sagisaka (eds.) *Quantitative prosody modeling for natural speech description and generation (Speech Communication 46 (3–4))*, 334–347.

Hoàng, C. C. (1985). Bước đầu nhận xét về đặc điểm ngữ điệu tiếng Việt (trên cứ liệu thực nghiệm). *Ngôn Ngữ, 3*, 40–49.

Jilka, M. (2000). *The Contribution of Intonation to the Perception of Foreign Accent*. Doctoral dissertation, University of Stuttgart, 2000.

Ladd, D. R. (1981). On intonational universals. In T. Myers et al. (Eds.), *The cognitive representation of speech*. Amsterdam: North Holland Publishing.

Ladd Jr., D. R. (1978). *The structure of Intonational meaning*. Bloomington: Indiana University Press.

Ladd, R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

Lim, L. (2002). Ethnic group differences aligned? Intonation patterns of Chinese, Indian and Malay Singaporean English. In. A. Brown, D. Deterding, & L. Ee-Ling (Eds.) *The English language in Singapore: Research on pronunciation* (pp.10–21). Singapore: Singapore Association for Applied Lingustics.

Lim, L. (2009). Some new Englishes as tone languages? In L. Lim and G. N. Gisborne(Eds.) Special issue on *The typology of Asian Englishes*. *English World-Wide*.30, 218–239.

Low, E-L. (1994*)*. Intonation patterns in Singapore English. M. Phil. Dissertation, Department of Linguistics, University of Cambridge.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics, 32*, 543–563.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (eds.), *Non-native Prosody: Phonetic Descriptions and Teaching Practice* (pp. 53–76). Mouton De Gruyter.

Mennen, I. (2015). Beyond segments: towards an L2 intonation learning theory (LILT). In Delais-Roussarie, E., Avanzi, M. & S. Herment (eds.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*. Springer: Verlag.

Nguyễn, T. A. T., & Ingram, J. (2004). A corpus-based analysis of transfer effects and connected speech processes in Vietnamese English. *Proceedings of the Tenth Australian International Conference on Speech Science & Technology*. Macquarie University, Sydney, 8th–10th December.

Nguyễn, T. A. T.& Ingram J. (2005). Vietnameese acquisition of English word stress. *TESOL Quarterly,* 39 (2), 309–319.

Nguyễn, T. A. T., Ingram, J. C., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *J Phonetics, 36*, 158–190.

Nguyễn, T. T. H., & Boulakia, G. (1999). Another look at Vietnamese intonation. *XIVth Int Congress of Phonetic Sci*, 2399–2402.

O'Connor, J., & Arnold, G. (1973). *Intonation of colloquial English* (2nd ed.). London: Longman.

Ramírez, D., & Romero, J. (2005). The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers. *Intercultural Pragmatics, 2*(2), 151–168.

Ramirez Verdugo, D. (2002). Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English. *ICAME Journal, 26*, 115–132.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française* 28, 41–66.

Suzuki, H., Kiritani, S., & Imagawa, H. (1989). For improvement of English intonation learning system. *Annual Bulletin RILP, 23*, 59–63.

Trofimovich, & Baker, W. (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics, 28*(2), 251–276.

van Heuven, V. J., & van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication, 47*, 87–99.

Vũ, M. Q., Trần, Đ. Đ., & Castelli, É. (2006). Intonation des phrases interrogatives et Affirmatives en langue vietnamienne. *Journées d'Étude de la Parole, 4*.

Wells, J. (2006). *English intonation, an introduction*. Cambridge: Cambridge University Press.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics, 15*, 399–420.

Yuan, J., Shih, C., & Kochanski, G. P. (2002). Comparison of declarative and interrogative intonation in Chinese. *Proceedings of Speech Prosody, 2002*, 711–714.

Zeng, Xiao-Li, Philippe Martin & Georges Boulakia. (2004). Tones and intonation in declarative and interrogative sentences in Mandarin. *International Symposium on Tonal Aspects of Languages:* With Emphasis on Tone Languages.