Using Pitch and Length Information to Assess Speech Prosody

CHAN, Hang

Published in:
English Teaching and Learning

DOI:
10.1007/s42321-018-0017-1

Published: 01/06/2019

Document Version:
Peer reviewed version

Link to publication

Citation for published version (APA):
CHAN, H. (2019). Using Pitch and Length Information to Assess Speech Prosody: a Parallel Approach. *English Teaching and Learning*, *43*(2), 125-146. https://doi.org/10.1007/s42321-018-0017-1

Using pitch and length information to assess speech prosody: A parallel approach

Abstract

Recent studies on pronunciation instruction underscore the effectiveness of using pitch variation contours to enhance language learners' awareness of English intonation. Little is known of the situation in language production: Do learners adjust their pitch levels, or lengthen or shorten a sound in order to vary their prosody? The current study posits that both pitch and length variations are involved in encoding speech prosody; it purports to quantify their relative importance by using Hincks's (2005) Pitch Variation Quotient (PVQ) and a self-created Length Variation Quotient (LVQ). The methods section introduces four quantitative methods, using data from the participants' verbal speech as well as the training material (a song) to compare pitch and length behaviours. Results from three of the four methods support that the learners, after undergoing a period of training (a singing class), improved their length variation but not their pitch variation over a reading task. The fourth method (a between-within ANOVA) also confirms the learners' general prosodic improvement, but the test raises curious points about the complex relationship between pitch and length that are worth further consideration by pronunciation assessors. Overall, this study provides initial evidence of the importance of length variation for assessing prosodic change.

Keywords

Pitch; length; variation quotients; prosodic assessment.

1. Introduction

Prosodic cues, including pitch, length, and intensity, play important roles in the acoustic instantiation of a sound (e.g., Chun, 2002; Cruttenden, 1997; Hirst & Di Cristo, 1998; Ladefoged, 2003; Rogerson-Revell, 2011). These factors can cause a sound to be perceived as sharper or duller, stronger or weaker, or longer or shorter, while during the process the sound itself is unchanged (e.g., Levelt, 1989). The proposal advanced by this study is this: Given that these cues (pitch, length, and intensity) are responsible for the aural shape of a sound, analysing how they are patterned in the voice may give hints as to how they are employed by a speaker. Analysing how these cues are used before and after a pronunciation activity can further reveal if they are amenable to training. The analytical framework presented in this study involves identifying these "common denominators", pitch and length (hereafter, P and L)[1], in verbal speech. The obtained P and L values were further converted into a Pitch Variation Quotient (PVQ, a measure used by Hincks, 2005) and a Length Variation Quotient (LVQ, an alternative measure proposed by this study). These quotients indicate the degree of prosodic variations exhibited by a stretch of speech, with an assumption that more prosodic variation indicates more lively speech[2]. The ultimate question posed by this paper is whether variations in both

[1] This study excludes intensity (loudness) from its analysis, as its realisation can be affected by external environmental factors (see the end of Section 1.1).

[2] Although a higher level of prosodic variation is generally desirable (Hincks, 2005), this paper does not wish to equate prosodic variation with "better-sounding speech". To investigate this link would require perceptual judgement tasks, which are beyond what the current methods can do.

pitch and length (as indexed by PVQ and LVQ) were equally drawn upon by the participants as they produced connected speech. Because the present study is interested in broad changes in prosodic cues over a stretch of speech, it will not examine cue use in specific environments, such as in sentence-initial or sentence-final positions (e.g., Kim, Broersma & Cho, 2012) or for specific discourse functions, such as to signal contrasts (e.g., Brazil, 1997).

The following sections highlight three observations in regard to the status quo of prosodic training before the paper describes the assessment framework. First, the existing literature contains a dominant focus on the teaching of pitch movements, to the extent that the possible contribution of another cue, such as length, is somewhat neglected (Section 1.1). Second, the recurrent interest in looking for computer solutions to prosodic training is unparalleled by an interest in the use of a particular text type or a particular activity for prosodic training (Section 1.2). Third, although pedagogical innovators promote the pronunciation prospects of a classroom activity, such as singing, experimental studies involving this activity focus elsewhere – chiefly on its mnemonic benefits. This paper argues that the pronunciation prospects of singing are still unknown, as the directions of these two groups do not always converge. The first of these observations explains the two-pronged focus of this study on P and L; the next two observations explain why the current study drew data from a singing classroom.

1.1 Pitch and Length as Two Prosodic Cues

Fundamental frequency [F0] (or its perceptual correlate, pitch)[3] has long been a subject of theoretical interest in phonetics research. What is pertinent to the present study is its suprasegmental function of marking a word's saliency. Pitch is the most important factor in causing stress sensation (e.g., Chun, 2002; Cruttenden, 1997; Fry, 1979). The phonetic definition of pitch is the rate at which vocal fold pulses occur (Ladefoged, 2005; Rogerson-Revell, 2011). When the frequency of these pulses is high, a distinct, high-pitched sound can be heard. Pitch variation has been used in major models of discourse intonation (e.g., Abercrombie, 1967; Brazil, 1980, 1997), such as how the changing of pitch levels can invite different interpretations of an utterance. In regard to speech perception, pitch is said to have an "all-or-none" effect (Fry, 1958), meaning that a slight movement is enough to induce attention. In production, "pitch accent will overrule everything else" (Levelt, 1989, p. 305), making a sound more noticeable than other means. Experimental studies employing a stress judgement task (i.e., a listener judges where a stressed sound lies) have supported pitch as an important clue to stress. Chrabaszcz and Winn (2014) found that, after vowel quality (i.e., the use of strong or weak nuclei), native English speakers and Mandarin speakers relied on pitch when deciding where a stressed sound fell. It is worth mentioning that a recurrent trend in the pronunciation training literature has been to advocate the display of pitch variation to make learners aware of how they should adjust their voices to give more intonationally varied sentences (e.g., Chun, 2002; Hardison, 2004; Hincks, 2005, 2015; Hincks & Edlund, 2009; James, 1976). In these studies, pitch tracings are displayed on the screen to assist learners' emulations, and the learning outcome (indexed by a score, such as the PVQ, or rated by human judges) is taken as an indication of speech liveliness (e.g., Hincks, 2005; Hincks & Edlund, 2009). The centrality of pitch in phonological descriptions makes it a natural choice for

[3] The term "pitch", rather than "fundamental frequency" ("F0", or the physical property of voicing), is used in this paper. This usage is to follow the naming of the metric PVQ by Hincks (2005), in which "pitch" is used.

inclusion in the current assessment framework, which evaluates if learners make adjustments to the pitch levels of their voices after training.

Whereas pitch is directly caused by vocal fold movements, length, the next cue for inclusion in the framework, represents the displacement of time from one sound to the next. Sometimes deemed to be secondary to pitch in terms of its effectiveness in cuing stress sensations (e.g., Chun, 2002; Fry, 1955, 1958), syllable lengths can also evoke the sensation of stress; for example, long vowels and diphthongs can be perceived as being more stressed than short vowels are (Rogerson-Revell, 2011). Some experimental studies, in fact, have found length to be a main cue of stress. Yu and Andruski's (2010) native English speakers consistently used length to determine where stress lay in a stress judgement test. Adopting a similar method, Chan (2018) reported that native English speakers judged both pitch and length as important in indicating stress, and the differences in their preferences were minimal. Ladefoged (2003) and Levis (1999) caution against treating pitch as the only cue of stress. For them, length is substitutable for pitch in many cases (e.g., a combination of low pitch and lengthening can also signal stress). It may be said that length is no less important than pitch in evoking stress sensation; however, length may be less consistent as a cue, since pitch accent may override length's effectiveness in, for example, larger prosodic structures (e.g., in sentences; Fry, 1958; Morrill, 2011).

An impression that emanates from these studies is that current endeavours seem to be more interested in finding out the relative importance of prosodic cues, with an aim to prioritise their involvement, than in studying how they may be responsible for a prosodic contour. The pitch tracing studies, for example, foreground pitch movements in their perceptual feedback, whereas the way in which length variations work in the background is largely unknown. Also common in these studies is the adoption of stress perception tests; there is a chance that their results may not be generalisable to speech production (as is the focus of this study). This section aims to explain why it is useful to include both pitch and length in the current assessment framework. This study excludes a discussion of the role of intensity, however. Intensity is likely to be affected by external factors (e.g., Beckman, 1986), such as the acoustics of the room may affect whether a learner makes louder sounds.

1.2 The Case of Singing and Three General Predictions

Despite recent advances in using technological means to enhance learners' awareness of pitch contours (see the last section), little progress has been made in the analysis of a specific material or an activity in regard to how they can improve pronunciation, while singing along to a song, drama performances, and poetry readings are all intuitively sound activities for this purpose and are commonly discussed in the teacher-oriented literature (e.g., Brown, 2012; Celce-Murcia, Brinton, & Goodwin, 2010; Goodwin, 2014; Kenworthy, 1987).

Singing is somewhat different from other training activities. Singing itself is a language arts activity, carried out in a context-rich and stress-free environment (Tompkins, 2013); it is widely practised in schools (e.g., Gan & Chong, 1998; Iwasaki, Rasinski, Yildirim, & Zimmerman, 2013), and has both hedonic values and mnemonic functions (Good, Russo, & Sullivan, 2015). It is perhaps a common observation that singing or using songs is associated with a number of benefits, such as its ability to improve rhyming scheme, automaticity, fluency, appreciation of assonance, alliteration, stress patterns, intonation, rhythm, and awareness of phonological rules (e.g., assimilation and elision) (e.g., P. K. W. Chan, 1997; Gan & Chong, 1998; Iwasaki et al., 2013; Paquette & Rieg, 2008; Schoepp, 2001; Shen, 2009). More complicated, however, is the fact that these usual expectations of singing have not been matched with the common research goals of empirical studies, where keen attention has been paid to singing's effect on memory recall (e.g., Good et al., 2015; Lehmann & Seufert, 2018; Ludke, Ferreira, & Overy, 2014; Racette & Peretz, 2007; Rainey & Larsen, 2002; Wallace, 1994). To give an example, Good et al. (2015) assessed the lyric recall ability of a group of learners after they had learned the lyrics in different conditions (a group sang the lyrics and another group read the lyrics as a poem). They concluded that singing was superior in terms of aiding recall, as the melody provided "structure cues to support memory" (p. 637). Overall, the ways in which using a song promotes the pronunciation skills are often not the theme of interest in laboratory-based studies. This study considers it as important to return to the key concern of practitioners and pronunciation assessors – that is, to understand English learners' pronunciation changes after a period of training.

There is also one feature of singing that is relevant to the training of prosodic cues. Typically, the musical notations on a score sheet describe the pitch variations required for the given words. Ladefoged (2011, p. 118) calls them "steady-state pitches" – there is an expectation that a singer should vary his/her voice to reach those pitch levels (reaching perfect pitch levels, however, may be rare for average learners who do not aim to be professional singers). At the same time, lengths will co-vary. Given this characteristic, singing therefore represents an appropriate means for motivating learners to make elaborate use of their vocal space and speech apparatus for oral production (Christiner & Reiterer, 2013). Producing normal speech, on the other hand, does not require conforming to any exonormative reference standard, so there is less motivation for a learner to make full use of his/her voice. What is of interest for the current study is exploring whether this unique context – that is, the elaborated prosodic arrangement of a song (see Christiner & Reiterer's quotation below) and its mnemonic effect (see the previous paragraph) – will predispose learners to use prosodic cues in specific manners (see Schön & François's quotation below).

> The input infants receive from adults is exaggerated, simplified and highlighted and more song-like in its nature. There is a greater variation of pitch, longer vowels and/or slower pace. […] singing education is similar to L1 acquisition as it aims to create awareness about one's vocal apparatus and one's orofacial motor abilities.
> (Christiner & Reiterer, 2013, pp. 8-9)

> Singing is particularly well-suited to the study of the relation between language and music, the advantage being that both linguistic and musical information are merged into one acoustic signal with two salient dimensions, allowing for a direct comparison within the same experimental material.
> (Schön & François, 2011, p. 2)

The characteristics of the present song, as well as its suitability as a training material, will be discussed in Sections 2 and 3.3. Before explaining those details, I will make several predictions as to what could happen to learners' P and L measurements after a training activity, as these are key concerns of this study: (a) prosodic training could promote greater P and L variations simultaneously (for example, the P and L from one syllable to the next will be more varied), causing the PVQ and LVQ to increase; (b) prosodic training could widen weak-strong syllable differences or enlarge or shorten the size of a foot, thereby resulting in improved rhythm; and/or (c) prosodic training may cause the use of prosodic cues to be reprioritised (for example, a given activity may favour the use of P or another activity may promote the use of L). The present article will focus on the first of these predictions, whereas the third prediction has been discussed by Chan (2018). All of these are logical deductions, and spelling out the details of the possible consequences should help us theorise a reasonable assessment plan.

In summary, the current trend toward teaching pronunciation by using computer-based feedback means that ordinary classroom activities receive less attention from researchers. The psycho-cognitive orientation of singing studies also means that its pronunciation effects are less known than its mnemonic effects. Although the current research drew data from a singing class (apart from the native English speakers' data), many of the upcoming discussions are relevant to prosodic assessment at large, such as the relationship between P and L, and the use of different data sources to triangulate the effects of prosodic training.

1.3 A Note on Terminology

Since this paper examines P and L variations, I shall call the object of analysis "prosody" or "prosodic change". The word "prosody" is used in two senses. In one sense it refers to the holistic quality of speech (e.g., Chun, 2002; Crystal, 1969; Pennington, 1996; Trask, 1996), representing the collective outcome of various inner contributing systems, including word stress, intonation (pitch change), and rhythm. In the second sense, prosodic change refers to the acoustic instantiations of sounds, as modulated by both P and L variations, in the forward undulation of speech. Overall, this paper presents a prosodic assessment framework through the lens of a given activity. The way prosody is measured here does not mean to be the only (or the last) method of measurement. Other ways of investigating prosody may be to examine rhythm, vowel quality, or fluency (e.g., Deterding, 2001; Kormos & Dénes, 2004; Low, Grabe, & Nolan, 2000; Setter, 2006), but to do so would involve looking at other aspects of the activity, which may not be P and L variations exactly.

2. The Proposed Framework: Using P and L Information to Assess Prosodic Change

The current assessment framework makes use of multiple quantitative methods to compare P and L behaviours. Three of the methods involve variation quotients (PVQ and LVQ), and one involves raw data (P and L fluctuations). As stated at the beginning of this paper, the current study followed Hincks (2005) to derive the Pitch Variation Quotient (PVQ) as a speech liveliness index, using this equation: the standard deviation (SD) of the pitch variation of a given speaker/his or her mean pitch (i.e., the same equation used to calculate the "coefficient of variation" in mathematics). The resultant value is unitless and can be used to compare the different kinds of data, such as the pitch variations between a man and a woman, and the different dimensions of a performance (e.g., whether the P variation is more or less dramatic than the L variation). A smaller quotient is taken to mean less variation, and a bigger quotient indicates otherwise. This study extends the equation to derive what is called the Length Variation Quotient (LVQ): the SD of L/the mean L of a speaker across a stretch of speech. Since it is known that different people produce different P and L values in different situations (e.g., a singer will naturally produce higher pitch levels during a performance, whereas the current participants would speak in a calmer manner in a test room), PVQ and LVQ are more objective indicators of voice variation than the raw P and L values are. The former ones are used for comparison in the following.

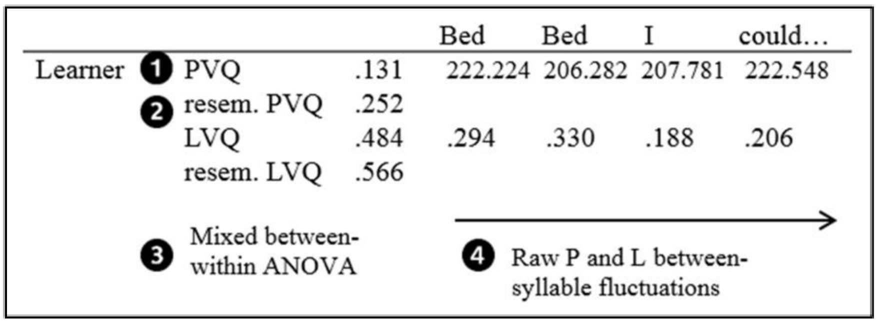
To obtain the variation quotients from the song, the researcher first cleansed the audio recording of the song using the Vocal Isolation function (under Effect tab)[4] of the software Audacity (version 2.1.1; Audacity Team, 2016). The resultant file was then submitted to the speech-analysis software Praat (version 6.0.20; Boersma & Weenink, 2016) to extract the average P and L from each syllable of the whole song (94 syllables, with the final syllable of each line removed). The P and L values (Table 1) could then be used to calculate the PVQ and LVQ. The last column of Table 1 shows that the current learning material (the song) has a PVQ of .383 and an LVQ of .867. The participants' audio recordings were processed in Praat in a similar manner (see further information in Section 3.5). The obtained quotients were subjected to statistical testing (to be introduced below). Note that the current analysis assesses between-syllable variation (i.e., how P and L change from one syllable to the next) rather than within-syllable variation (i.e., how a syllable is uttered by different people). Using between-syllable variation was congruent with the goal of examining the degree to which a person varies his or her voice while reading a sentence forward. With within-syllable variation, the focus is not on how a person reads a sentence forward, but on how a particular syllable is rendered differently across speakers.

Table 1
The raw P and L values of the song and the calculation of variation quotients

	Bed	Bed	I	could	n't	go		
Pitch (Hz)	518.464	346.126	358.662	482.458	236.319	458.445	…	PVQ = .383
Length (sec)	.494	.474	.180	.173	.177	.206	…	LVQ = .867

2.1 Methods of Assessing Prosodic Change

This section introduces the four assessment methods to be performed on the learners' data. Figure 1 gives a summary of these methods and the data required.



				Bed	Bed	I	could…
Learner	❶	PVQ	.131	222.224	206.282	207.781	222.548
	❷	resem. PVQ	.252				
		LVQ	.484	.294	.330	.188	.206
		resem. LVQ	.566				

❸ Mixed between-within ANOVA
❹ Raw P and L between-syllable fluctuations

Figure 1. Four assessment methods are employed to assess prosody

Method 1: Using Original Quotients

If we assume that prosodic training can promote voice variation, the learners' PVQs and LVQs should be higher in the post-test than in the pre-test. The pre-post differences (i.e., the two circles on the left in Figure 2) can be verified through paired-sample *t*-tests.

Method 2: Using Resemblance Quotients

[4] This process improves the quality of the sung words and facilitates more accurate extraction of the P and L values.

Comparing the original quotients excludes the song (the target) from the picture. In the second method, the song's PVQ and LVQ were used as "reference frames" from which it could be deduced how much a learner's performance differed from the song. For this purpose, the researcher derived a new set of quotients, called the "resemblance quotients", by directly subtracting a learner's quotient from the song's (e.g., the song's PVQ .383 minus a learner's PVQ .136 = .247 = resemblance PVQ)[5]. Resemblance quotients should drop in order to reflect an increase in song-speech resemblance. Similarly, the pre-post differences (i.e., the lines on top of the circles in Figure 2) can be verified by paired-sample *t*-tests.

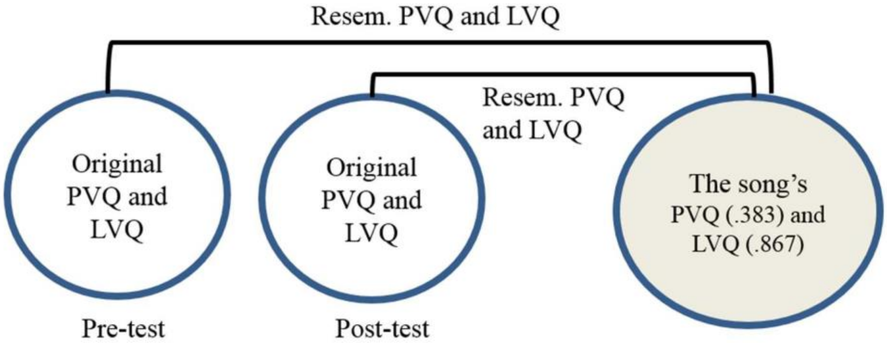


Figure 2. Original and resemblance quotients

Method 3: Mixed Between-Within Analysis of Variance (ANOVA)

The benefit of using a more advanced test, such as a mixed between-within ANOVA (Pallant, 2016), is that all the data – that is, the two factors (P and L) and the two test points – could be considered at the same time. Overall, the test aimed to find out whether the progressions of PVQs and LVQs would be different between two time points (i.e., whether PVQs and/or LVQs progressed quickly from time 1 to time 2). Whereas time was a within-subjects variable (i.e., the same participant was tested twice), the between-subjects variable, which has traditionally meant different treatments undergone by different participants, requires careful theorisation, as there was only one treatment in the current study (only the learners completed the test twice, whereas the native English speakers gave their data once). Given that P and L are different notions with different units of measurement, they can be seen as two independent factors exerting force at the same time[6], directly influencing how a learner uses his or her voice. It is also reasonable to assume that most speakers are generally unaware of their P and L when speaking; it would be interesting to explore, through Method 3, how the sounds they emit may indicate the relative importance of P and L. By using a mixed between-within ANOVA (the General Linear Model tab, then the Repeated Measures tab, in SPSS version 24), the P and L were between-subjects variables, time was a within-subjects variable, and the quotients (original and resemblance) were dependent variables. The learners' raw P and L values were not used, as these do not reflect degrees of variation. A robust training effect should cause PVQ and/or LVQ to change significantly in the post-test. Submitting the data in this way enables the researcher to intuit how P and L variations might change. Although such changes are presumably related to the prosodic training, caution should be exercised in the

[5] As suggested in Section 2, these quotients are unitless; direct additions or subtractions can be performed.

[6] It is possible to imagine that a singer uses P and L in particular manners, while he or she may or may not be aware of how they are used. The P and L are like forces intermingled in the voice.

interpretation of the results, as there was only one treatment group involved (i.e., to fully understand the effect of singing vis-à-vis that of no singing would require another treatment group).

Method 4: Using Raw P and L Fluctuations

Method 4 investigated P and L behaviours by comparing the learners' between-syllable P and L fluctuations with those of the song. The advantage of this method was that there were 94 data points (i.e., the whole song) with which to make comparisons. Pearson correlation coefficients could indicate the degrees of resemblance.

3. Research Methods

This study hypothesises that the melody of a song elaborates the acoustics of the words, stretching the P and L in various manners. Overall, we predict that prosodic training would promote greater P and L variations at the same time, causing PVQ and LVQ to increase in similar manners (i.e., the first general prediction discussed in Section 1.2).

3.1 Research Questions (RQ)

RQ 1: Does the song (i.e., do the sung words) display a stronger prosodic pattern than speech does?

□ **Hypothesis (H) 1:** Because the song was used as training material, it should display strong P and L variations. Its PVQ and LVQ (.383 and .867, as presented in Section 2) should be higher than those of normal speech.

RQ 2: Will the Cantonese learners improve their P and L variations after training? To compare P and L behaviours, four methods are used in this study:

□ **Method 1:** Original Quotients
H2a: If training can improve prosodic variation, the Cantonese learners' speech (N = 32) should exhibit more P and L variations in the post-test than they did in the pre-test, hence increasing their PVQs and LVQs.

□ **Method 2:** Resemblance Quotients
H2b: If training can improve prosodic variation, it should decrease the differences between the song and the learners' speech. The resemblance PVQs and LVQs should be lower in the post-test than in the pre-test.

□ **Method 3:** Mixed Between-Within Analysis of Variance
H2c: P and L are separate notions but are trained simultaneously. This method regards P and L as between-subjects factors, time as a within-subjects variable, and PVQs and LVQs as dependent variables. Effective training should cause PVQs and LVQs to vary in separate ways across the two time points, producing a factors □ time interaction.

□ **Method 4:** Raw P and L Fluctuations

H2d: Method 4 submits both the song's and the learners' pre- and post-test fluctuations over 94 data points (the whole song) to Pearson correlation analyses. A stronger correlation coefficient should be found between the post-test fluctuations and those of the song if the training is effective (i.e., learners' post-test fluctuations should increasingly resemble the song's fluctuations).

A group of native English speakers (hereafter, NSs; N = 41) from the University of Queensland performed the lyric reading task once. Their data will be referred to in the following sections.

3.2 Participants

A total of 32 Cantonese students (30F, 2M, with a mean age of 15.09 (SD .53)) from seven local schools participated in the training activity. All were secondary three students, approximate to the ninth grade in the US educational system. Each was given a HKD200 coffee coupon at the end of the nine-hour activity. The comparison-group data were collected on the campus of the University of Queensland in Australia. A total of 41 participants (25F; 16M, with a mean age of 20.73 (SD 4.71)) were recruited with the assistance of the university's TESOL Education Institute. They completed an informant questionnaire and a stress judgement test (not reported in this paper) and individually read the lyrics in a quiet classroom. Of these, 39 had been born in Australia and two in the UK. The majority were undergraduate students (33); there were also a few postgraduates (8). They came from a wide range of educational backgrounds: education (10), engineering (6), arts/communication (5), health science (5), natural science (3), business (3), music (3), law (2), linguistics (2), and psychology (2). Each participant was given AUD25 at the end of the meeting, which lasted for an hour and a half. They did not undergo any training, nor were they told about the use of a song in the test.

3.3 Training Material

The main training material used was the song "I Could Have Danced All Night" from the musical *My Fair Lady*. The current study also examined the material in an effort to understand how P and L are patterned in it (see Section 4.1). For that purpose, the original version sung by Julie Andrews was submitted to Audacity and to Praat for the analysis of P and L variations (already discussed in Section 2). The song is listed as suggested material in the Grade 5 singing examination by the Associated Board of the Royal Schools of Music (ABRSM) (2009), and the musical itself is a recommended learning material by the local examination authority (HKEAA, 1999). A musician with 15 years of coaching experience (in piano and vocal singing) pointed out that it is a syllabic song, in which each syllable is sung with a separate note (e.g., "tonight" = two notes). A syllabic song allows feasible assessment of P and L values, as the researcher can obtain such information from each syllable (i.e., it would be difficult to obtain P and L values from words sung in changing notes, as is the case in "melismatic" songs). Along with the song's topical relatedness to learning English and its steady speed (98 words for 68 seconds)[7], the material was considered suitable to be used in the treatment.

[7] Humans normally produce 120-150 words per minute (de Bot, 1992; Levelt, 1989). This is a useful guide for evaluating whether the speed of the material is too high. In general, the present song is easy to listen to and to follow.

In contemplating the current method, the researcher is aware that the choice of a particular song may have a major influence on learning outcomes and of the fact that sentence intonation may be more “open-ended” than inherent word stress patterns (e.g., Ashby & Maidment, 2005). The latter point concerns whether the open-ended intonational variation in the song is useful to learn[8]. To reconcile these variables (the material’s own characteristics and sentence intonation) and to make the current experience as generalisable as possible to other teaching contexts, a principle followed by this study was to select a song that has reasonable stress arrangements. A desirable outcome would be one in which the song’s varying of P and L would not lead to unnatural speech prosody. The musician was asked to examine the original score and listen to the song to detect any incongruity in the musical setting (see, e.g., Gingold & Abravanel, 1987), such as whether the melody was compatible with word stress patterns and whether non-focal, function words had been unduly stressed. Only one word, *jewel*, was suggested to be removed from the analysis because the original stress pattern (□□ for “jewel”) is incongruent with the musical cue (jew□al □). The song’s musical setting was overall compliant with normal word stress placements. In the end, although the full lyrics consisted of 110 syllables, only 94 syllables were analysed after excluding “jewel” and the final syllable of each line (which was removed to avoid the lengthening effect common in such places) (Ferreira, 2005) (see Appendix I).

3.4 The Singing Sessions

Two identical singing sessions were run at the author’s university, ensuring a small class arrangement. The vocal singer (different from the musician) was a tenor from Hong Kong Youth Windophilics. The pre- and post-reading tasks were held at 9AM and 5PM. This was an individual test; each learner read the lyrics in the presence of the researcher or the vocal singer (two separate rooms were arranged). Shure USB microphones were used to record their voices. After the test, the warm-up activities included vocal exercises and discussing the storyline of the song. The awareness-raising activity asked the learners to listen to lines of the song, hum them, and then sing the lines. Singing occurred in different manners – first in a big group and then in small groups of four. This cycle was repeated after a lunch break. The participants were seated in small semi-circles and were given group names. This created a motivated yet slightly competitive atmosphere. Overall, the singer led the participants in singing the whole song at least six times. The collection of classroom data yields recommendations that are more generalisable to real-life conditions. Responding to computer-generated visual contours or singing to respond to prompts in a computer program, whilst ensuring a degree of reliability, cannot be sustained over a few hours. Lengths of the singing sessions in published studies vary, such as from Ludke et al.’s (2014) 15 minutes, to Lehmann and Seufert’s (2018) 45 minutes, to Good et al.’s (2015) 400 minutes. The reader can compare these lengths with that of the present study.

3.5 Prosodic Analyses and Interrater Reliability

The researcher and two research assistants extracted the P and L values from the learners’ and NSs’ recordings. The P values were automatically generated by Praat after the voiced section of a syllable was highlighted (i.e., average syllable pitch was used). Length was the region that covered all of the audibly produced consonant and vowel sounds of a syllable

[8] By being “open-ended”, I mean that sentence intonation (or the song) allows a speaker a certain level of freedom to add emotion colour to an utterance. This, however, does not mean that sentence intonation would not respect word stress placements. On the contrary, it is still contingent upon word stress patterns, e.g., content words/main ideas, rather than function words, tend to be stressed (Celce-Murcia et al., 2010).

(e.g., /bed/ for “bed”; /dɑːnst/ for “danced”) and had to be extracted manually. Common problems observed during the process included elisions, devoicing, and blending of sounds. In such cases, the decisions were aided by the spectrograms and pitch tracings displayed on Praat. Consistency between the researcher and the assistants’ decisions was checked by submitting 940 pairs of P values (i.e., equivalent to 10 participants’ data, separately analysed by the researcher and an assistant) and 940 pairs of L values into SPSS to calculate intraclass correlation coefficients (suitable for comparing continuous data). The coefficient returned for the P values was .999 and that for the L values was .960, indicating a high degree of consistency between the raters’ extraction processes.

4. Results

4.1 The Prosody of the Song (RQ 1)

If we assume that a song has a more elaborated prosodic arrangement than normal speech has, it should exhibit stronger prosodic variation than normal speech does. Table 2 presents the song’s raw P and L values and those from the learners and the NSs. Because these absolute values (second column) were not suitable for analysis, the PVQs and LVQs in the third and fourth columns were used for between-group comparisons. A quick survey of the data shows that the song had more dramatic values than those of the NSs and the learners (e.g., its PVQ and LVQ are .383 and .867, respectively), indicating that the song had strong P and L variations. Although both the quotients are strong (supporting **H1**), their disparity also suggests that the singer varied L to a greater extent than P (i.e., the coefficients of variations, .383 and .867, can be compared directly).

Table 2
Variation quotients and resemblance quotients

	Mean (SD)	Original quotients	Resemblance quotients
Song			
Pitch (Hz)	330.984 Hz (126.766)	PVQ .383 †	---
Length (sec)	.584 (.506)	LVQ .867	---
Cantonese Learners (N = 32)			
Pre - Pitch (Hz)	213.769 (25.364)	PVQ .122 (.060)	PVQ .261 (.060)
Post - Pitch (Hz)	218.666 (27.181)	PVQ .131 (.063) □ *t*(31) = -.983; *p* = .333; eta^2 = .030	PVQ .252 (.063) □ *t*(31) = -.983; *p* = .333; eta^2 = .030
Pre - Length (sec)	.284 (.118)	LVQ .419 (.063)	LVQ .448 (.063)
Post - Length (sec)	.291 (.127)	LVQ .442 (.074) □ *t*(31) = -2.250; *p* = .032*; eta^2 = .140	LVQ .425 (.074) □ *t*(31) = -2.250; *p* = .032*; eta^2 = .140
Native Speakers (N = 41)			
Pitch (Hz)	164.908 (41.635)	PVQ .263 (.112)	PVQ .120 (.122)
Length (sec)	.225 (.127)	LVQ .565 (.060)	LVQ .302 (.060)

Key: $p \leq .05$*; $p \leq .01$**; $p \leq .001$***
† The standard deviation (SD) is not available for the song, as only one song was used.

4.2 Methods 1 and 2 (RQ 2)

Given that the song exhibits more prosodic variations than verbal speech does, the next question to explore is whether singing along to the song leads to more voice variation. The third column of Table 2 shows that the learners' mean PVQ rose from .122 to .131. This mean PVQ is comparable to Hincks's (2005) range of .06-.30 (her data were drawn from oral presentations). However, such an improvement by the learners was statistically insignificant (i.e., PVQ .122 □ .131; p = .333). In contrast, the learners made significant improvement in their LVQs, at an α value of .05 (i.e., LVQ .419 □ .442; p = .032, eta^2 (effect size) = .140), indicating that they increased their L variation more than their P variation. This improvement from .419 to .442 moved in the direction of the NSs' .565 (bottom of column 3) but was still short of it. Thus, **H2a** is only partially supported, as only the LVQ was found to have improved. The patterns of the resemblance quotients (which take into account song-speech resemblances) mirror those of the original quotients. Both the PVQ and LVQ dropped, but only the LVQ was significantly lower in the post-test. Again, **H2b** was partially supported; only the L variation had changed markedly. Overall, the results from Methods 1 and 2 indicate that the learners' improved prosody came from the fact that they shortened and lengthened sounds to vary their voices, but they did not significantly readjust their pitch levels after training.

4.3 Method 3 (RQ 2)

The purpose of Method 3 (a mixed between-within ANOVA) was to examine whether the PVQs and LVQs progressed differently between two time points. The main conceptual challenge in this operation involved the possibility of P and L being independent notions, both of which exerted force during the training and caused the variation quotients to change over time. Levene's test of equality was non-significant (pre p = .185; post p = .103), indicating that the other test statistics were ready for interpretation. Overall, the test failed to find an interaction between training factors (P and L) □ time (i.e., Wilks' Lambda = .976, F (1, 62) = 1.542, p = .219, eta^2 = .024). However, there was a main effect for time (Wilks' Lambda = .897; F (1, 62) = 7.143, p = .010, eta^2 = .103), suggesting there were positive changes in the quotients. The test also confirms that the two sets of quotients (depicted in the "between-subjects effects" table) had very different values (F (1, 62) = 401.524, p = .000, eta^2 = .866). This may relate to the fact that the LVQs were often two to three times bigger than the PVQs (see Table 2). Overall, this analysis rejects the hypothesis that singing could cause PVQs and LVQs to vary differently between the two time points (thus, **H2c,** which predicts an interaction effect, is unsupported). However, the test confirms the learners' improved prosody (i.e., toward more variation), and agrees that the PVQs and LVQs were of different magnitudes.

The resemblance quotients (fourth column of Table 2) were also submitted to an ANOVA as dependent variables. Similar results were returned: An interaction was not found (Wilks' Lambda = .976, F (1, 62) = 1.542, p = .219, eta^2 = .024); a main effect for time was found (Wilks' Lambda = .897, F (1, 62) = 7.143, p = .010, eta^2 = .103); and a between-group difference was found (F (1, 62) = 139.846, p = .000, eta^2 = .693). In a similar vein, this test supports a general decrease of the quotients but denies a difference between the PVQ and LVQ progressions (thus, **H2c** is unsupported).

4.4 Method 4 (RQ 2)

Method 4 compared the raw P and L fluctuations in the learners' verbal data with those of the song. Stronger correlations between the song's and the learners' post-test fluctuations would be taken as signs of a training effect. Before singing, the learners' L fluctuations already resembled those of the song (*r* = .572, column 3 of Table 3). After singing, their L fluctuations were even closer to those of the song (*r* = .630, column 3 in Table 3). The learners' P fluctuations (pre- and post-test), however, were not correlated with those of the song. This finding is the same as those by Methods 1 and 2; it suggests that singing had an influence on L variation but not on P variation. The native English speakers' data likewise show a resemblance between the L fluctuations and those of the song's L (*r* = .606). Overall, these results may go against the usual expectation that learners vary pitch levels mainly during speaking (see Section 1.1). This curious finding was further investigated by examining the original speech data, which are presented in the next section.

Table 3
Pearson correlations coefficients between the song's and the learners' P and L values over 94 syllables

Song's pitch	330.984 Hz (126.766)	
Learners' pitch (pre)	213.769 (25.364)	*r* = -.031; *p* = .765
Learners' pitch (post)	218.666 (27.181)	*r* = .010; *p* = .922
Native speakers' pitch	164.908 (41.635)	*r* = .047; *p* = .646
Song's length	.584 sec (.506)	
Learners' length (pre)	.284 (.118)	*r* = .572; *p* = .000***
Learners' length (post)	.291 (.127)	*r* = .630; *p* = .000***
Native speakers' length	.225 (.127)	*r* = .606; *p* = .000***

Key: *p* ≤ .05*; *p* ≤ .01**; *p* ≤ .001***

4.5 Further Analysis of the Speech Data

One way to find out why the prosodic training has a partial effect on L variation (as revealed by Method 4, as well as Methods 1 and 2) would be to return to the original material. In this analysis, the researcher selected two lines of the song – the refrain ("I could have danced all night, I could have danced all night"; see Figure 3) – and examined their P and L contours. Hypothetically, the song's L contour should be similar to the learners' L contour, given the strong correlations found in Method 4. This prediction was confirmed by the visual representations of the contours (as can be seen in the bottom two graphs of Figure 4), indicating that the way the learners varied L was similar to how L was patterned in the song. The same analysis also found evidence that the learners' P variation and the song's P variation had different shapes (as can be seen in the upper two graphs of Figure 4). It seems that the way in which P was used to encode speech was different from the way in which it is patterned in the song. Whereas the song has sharp rises and falls in the pitch levels of the three words "danced", "all", and "night", the learners' P contour shows a general, gradual drop in the pitch levels over an intonation unit (upper-right graph), a pattern commonly observed in declarative sentences (Brazil, 1980; Coulthard, 1992).

Further characteristics of the graphs are worth noting. First, the P contour and the L contour could vary somewhat independently; for example, the word "all" is short but had a moderate pitch level, and the second “night” is lengthy but not high-pitched (see left column of Figure 4). Second, the L contours (lower graphs) seem to add emphasis to particular content words (“danced” and “night”), but what is stressed or not stressed is not so clear in the P contours (in particular in the learners’ P contour). Third, visually the P contours (upper graphs) showed fewer volatile movements, while the L contours (lower graphs) showed more volatile movements. Judging from the refrain of the song, it seems that P and L did not vary in completely uniform manners. One reason for our learners’ reliance on L could be that they might find L variation to be easier to emulate, as they produced a similar L contour in the reading test. The learners did not seem to apply what they had learned from the P contour of the song; they produced a different P contour in the reading test. Overall, contrasting the song’s P and L contours with the learners’ shows the extent of song-speech transfer that had occurred. The next section will discuss why P seemed to be difficult for the learners to adopt.

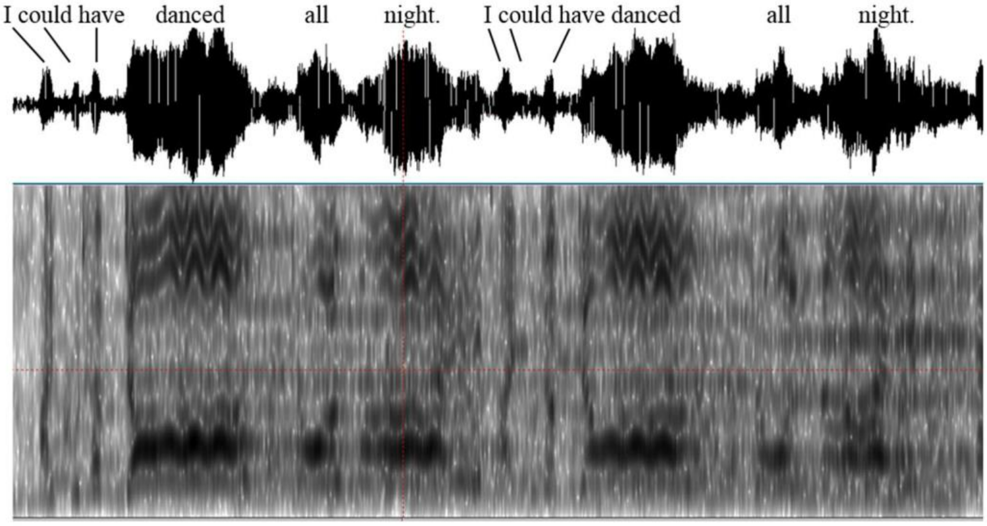


Figure 3. The refrain of the song on the screen of the speech-analysis software Praat.

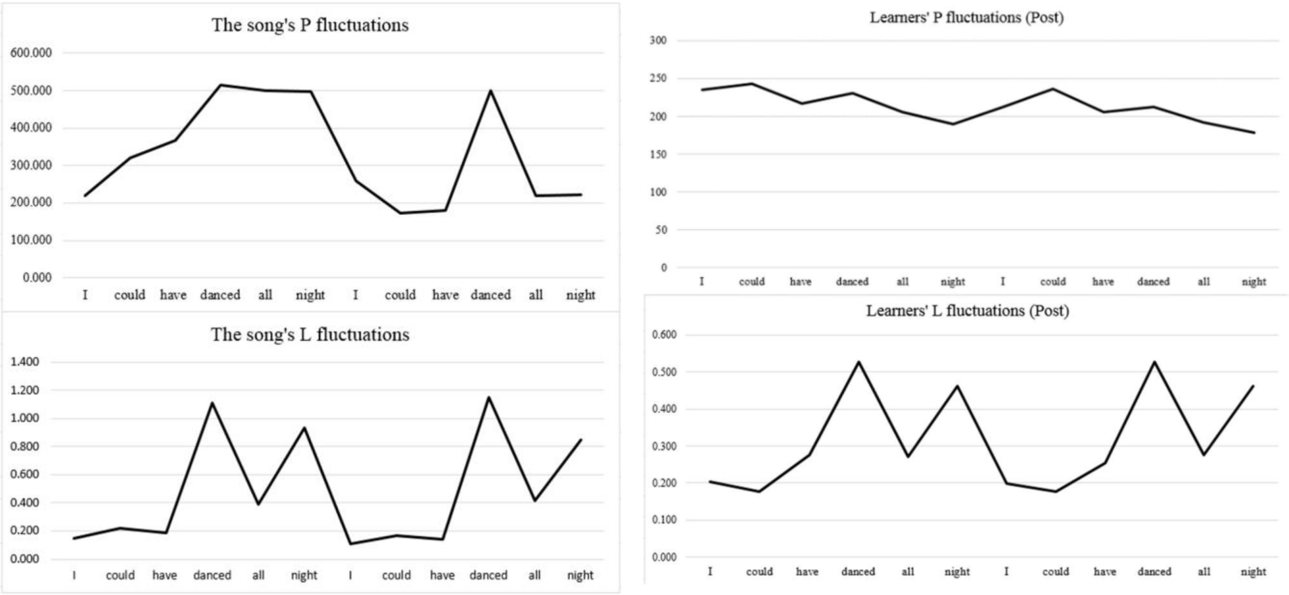


Figure 4. The song's (left) and the learners' (right) pitch and length contours.

5. Discussion

This study demonstrates how variations of pitch and length in one's voice can provide a detailed analysis of connected speech prosody. The benefit of using a song was its ability to encourage learners to process more extreme levels of pitch (Section 1.2). Although the majority of the data were obtained from a singing classroom (except for the NSs' data), the results and issues discussed are relevant to prosodic assessment at large.

We may be reminded that the quotients give summative indications of the liveliness of speech (used in Methods 1, 2, and 3), whereas in Method 4 the raw P and L fluctuations were used to examine the resemblance between the song and verbal speech. On the whole, three of the four tests (Methods 1, 2, and 4) suggest that the current prosodic training activity may have had a partial effect on the learners' L variation. First, analysis of the song found that this material has high PVQ (.383) and LVQ (.867), meaning that the song displays more prosodic variation than normal speech does, thereby supporting **H1**. Hypothetically, singing to it should lead to parallel improvements of the learners' PVQs and LVQs. This, however, was not the case. Methods 1 and 2 found that only the improvement across the LVQs was significant at the .05 level[9]. **H2a** and **H2b** (which predict simultaneous improvements of PVQs and LVQs) are thus only partially supported, and the first prediction of prosodic training (Section 1.2) is partially supported. Methods 1 and 2 also demonstrated that the "resemblance quotients" could be alternative prosodic measures if an investigator wishes to include the training material in the calculation of the quotients. Method 3 implemented a between-within ANOVA. The operation of this test involved seeing the treatment as being capable of exerting influences on both P and L simultaneously. This procedure made the researcher reflect on the complex relationship between P and L – dimensions that are clinically separable but that also act together

[9] Existing pitch visualisation studies often use human judges' ratings, mean scores, or raw PVQs to indicate prosodic changes. If we examine only the raw PVQs (Table 2), the present learners' post-test speech also showed more pitch variation.

in effecting prosody. Method 3 failed to find an interaction between factors (P and L) and time, thereby leading to the rejection of **H2c**; that is, PVQs and LVQs did not differ in significant ways between the pre-test and the post-test. However, Method 3 did detect an overall prosodic improvement (i.e., a main effect for time was found) and that PVQs and LVQs were different in magnitude (i.e., it found a between-subjects effect). At this point, we cannot fully account for why there was no interaction effect. It could be due to the current teaching condition (e.g., the duration of the training and the material used), or the complex relationship between P and L. The fact that P and L (as well as PVQ and LVQ) were somewhat related may have had an effect on the results of this robust test. Method 4 found that the learners' post-test L variation increasingly resembled that of the song. Thus, **H2d** (which predicts that both P and L variations would resemble those of the song) was partially supported; rather unexpectedly, the learners' use of pitch was unrelated to the pitch patterns of the song. Finally, examining the verbal data (the refrain of the song) revealed further information as to why L variation was preferred by the learners. Figure 4 shows that the sung words and verbal speech both contained similarly shaped L contours, suggesting that the material itself had arranged L variation in a way akin to verbal speech. This similarity could be a reason for the learners' preference for L variation, as they only needed to lengthen or shorten a sound to increase or reduce the stress sensation. On the other hand, there is less similarity in the pitch patterns between the song and verbal speech. That is, although the song has a high PVQ, the learners did not vary pitch in the manner they encountered in the song. Analysing the refrain also discovers that P and L variations do not correspond neatly (e.g., the word "all" is short but had a moderate pitch level). One tentative explanation is that the learners followed one style instead of the other, and that in the present case, they tended to follow the L variation pattern.

Two broad implications can be drawn from these findings. The present research found that the speakers varied L to a greater degree than P when encoding prosody. Apart from the chance that the length patterns may be easier to imitate (a reason given in the previous paragraph), this may also be because pitch has a powerful leverage effect (e.g., Fry, 1958); a slight increase is obvious and overusing it could make one sound un-nativelike (Zhang, Nissen, & Francis, 2008). If that is the case, the current learners might find it difficult to make "just enough" pitch adjustments in order not to give any awkward-sounding speech. In a pronunciation classroom, asking learners to raise their voices for stress marking should be accompanied by stretching or prolonging the sounds.

Second, the quotients showed that the song attained extreme scores in both aspects of P and L, but it had limited power to predispose the learners to use their pitch. Then, does the lack of transfer of P in the present case challenge the basic assumption of the pitch tracings studies (Section 1.1), which show that pitch information is useful for pronunciation instruction? One tentative explanation could be that the song as a training material may contain specific P variation patterns that are not likely to be adoptable in speech. As shown in Section 4.5, some high-pitched words in the song were not adopted by the learners, who continued to speak in a manner that showed a gentle decrease of pitch levels. On the other hand, their adoption of the L variation pattern is well worth investigation. If these speculations are further proven, the current finding may constitute initial evidence that singing may only have a partial effect on prosodic variation, despite its overall strong prosody. Currently, the lack of similar studies and the inadequate understanding of singing as a pronunciation activity leave much to be discovered about what would occur during song-speech transfer (most studies mentioned in Section 1.1 used human spoken sentences for training, and transfer from speech to speech was assumed to be more direct). Further inquiry into a pronunciation activity (be it singing, poetry reading, or drama performances) should aim to improve our understanding of a given activity's effects on the use of prosodic cues, stress placements, and the information structure of speech. The knowledge of whether the activity can build awareness of specific or general aspects of English pronunciation (such as singing may promote the use of L in the present study) will help teachers develop realistic expectations of different training techniques.

Before concluding this paper, the researcher would like to highlight a few reasons why the importance of length variation has not sparked much discussion in the literature. The first is that most extant studies relied on human raters to judge whether there was prosodic improvement in learners' speech (e.g., Hincks, 2005; Hincks & Edlund, 2009; James, 1976; Levis & Pickering, 2004). It is reasonable to assume that human raters will not be able to differentiate the fine differences occurring between pitch and length variations, and such differences would have to be detected by speech analysis software. Furthermore, existing computer programmes (e.g., Praat & VisiPitch; see Chun, Hardison, & Pennington, 2008) are able to extract fundamental frequency automatically but have not been programmed to extract length information, so length has to be done manually by a researcher. The labour-intensiveness of this task may explain the lack of attention paid to this dimension. Finally, most studies on the promotion of pitch visualisation techniques (e.g., Hardison, 2004; Hincks, 2005; Hincks & Edlund, 2009) do not make multiple comparisons between the obtained data. The different comparisons undertaken by this study (e.g., P and L, pre-test and post-test, native and non-native, and song and speech), as well as the use of multiple quantitative methods, may explain why the current study found more intricate patterns of prosodic cue use. On the whole, while we cannot tell from the current study whether a different training activity can result in different P and L variations, the obtained results can still offer advice that is generalisable to future researchers' own investigations. For example, future researchers should be more aware of both factors – P and L – in regard to assessing prosody. There is also a need to examine the original training material in order to gain a fuller picture of any gains or changes.

6. Conclusion

This study argued that the exclusive focus on pitch in existing studies has in fact not eliminated the effect of length, because it is always present, orchestrating alongside pitch movements. It proposed a consideration of the metric LVQ, in addition to the existing PVQ, in regard to prosodic assessment. All the current methods of analysis supported more prosodic variation in the learners' speech after singing, and Methods 1, 2, and 4 in particular demonstrated that this improvement was related to the learners' ability to use length variation. The different analyses also prompted a consideration of the relationship between P and L. These two notions may not be clinically separable in their realisation of prosody (a discussion prompted by Method 3), and yet they showed somewhat independent responses to the training (i.e., P was less transferred than L, according to the results from Methods 1, 2, and 4). If researchers consider P and L as being related, it is worth including both factors for assessing prosody in the future. If researchers are more interested in the independent nature of P and L, it is worth further investigating how they are learned in different ways through different training tasks and materials.

7. Limitations

This study has several limitations. First, we only included one song in the training. We thus do not know whether the current experience can be extended to other songs. Second, only one learning activity was conducted. We do not know how singing may compare with another prosodic activity. As pointed out previously, this study focused on how prosodic training may orchestrate the use of inner prosodic cues; to compare different learning activities is beyond the scale and capability of this work. Third, the present study investigated whether prosodic patterns in the song are transferred to verbal speech (i.e., via cross-modal transfer), but it did not involve a delayed post-test. Including a delayed post-test would strengthen the design and help investigate the mnemonic effect of the song. Despite these limitations, however, the assessment framework presented here provides insight into the complex relationship between P and L, demonstrating that it is possible to quantify one's prosody using multiple quantitative methods.

8. Conflict of interest

The corresponding author states that there is no conflict of interest in this research.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh, UK: Edinburgh University Press.

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*. Cambridge, UK: Cambridge University Press.

Associated Board of the Royal Schools of Music (AMRSM). (2009). Grade 5 singing examination. Retrieved from http://us.abrsm.org/en/our-exams/singing-exams/.

Audacity Team. (2016). Audio editor and recorder [computer programme]. (Version 2.1.1). Retrieved from http://audacityteam.org/.

Beckman, M. E. (1986). *Stress and non-stress accent*. Riverton, NJ: Foris Publications.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [computer programme]. (Version 6.0.21). Retrieved from http://www.praat.org/.

Brazil, D. (1980). *Discourse intonation and language teaching*. London: Longman.

Brazil, D. (1997). *The communicative value of intonation in English*. Cambridge, UK; New York, NY: Cambridge University Press.

Brown, J. D. (2012). *New ways in teaching connected speech*. Alexandria, VA: TESOL International Association.

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (Vol. 2). New York, NY: Cambridge University Press.

Chan, H. (2018). A method of prosodic assessment: Insights from a singing workshop. *Cogent Education, 5*(1). doi:10.1080/2331186X.2018.1461047.

Chan, P. K. W. (1997). Using songs in the English language classroom. In P. Falvey & P. Kennedy (Eds.), *Learning language through literature: A sourcebook for teachers of English in Hong Kong* (pp. 107-115). Hong Kong: Hong Kong University Press.

Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in psychology, 4*, 1-11. doi:10.3389/fpsyg.2013.00874.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins Publishing Company.

Chun, D. M., Hardison, D. M., & Pennington, M. C. (2008). Technologies for prosody in context: Past and future of L2 research and practice. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 323-346). Amsterdam; Philadelphia, PA: John Benjamins Publishing.

Coulthard, M. (1992). The significance of intonation in discourse. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 35-49). London; New York, NY: Routledge.

Cruttenden, A. (1997). *Intonation* (Vol. 2). Cambridge, UK: Cambridge University Press.

Crystal, D. (1969). *Prosodic systems and intonation in English*. London, UK: Cambridge University Press.

de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics, 13*(1), 1-24.

Deterding, D. (2001). Letter to the editor: The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics, 29*(2), 217.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27*(4), 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech, 1*(2), 126-152.

Fry, D. B. (1979). *The physics of speech*. Cambridge, UK: Cambridge University Press.

Gan, L., & Chong, S. (1998). The rhythm of language: Fostering oral and listening skills in Singapore pre-school children through an integrated music and language arts program. *Early Child Development and Care, 144*, 39-45.

Gingold, H., & Abravanel, E. (1987). Music as a mnemonic: The effects of good- and bad-music settings on verbatim recall of short passages by young children. *Psychomusicology: A Journal of Research in Music Cognition, 7*(1), 25-39. doi:10.1037/h0094188

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music, 43*(5), 627-640. doi:10.1177/0305735614528833.

Goodwin, J. (2014). Teaching pronunciation. In M. Celce-Murcia, D. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language* (4th ed., pp. 136-152). Boston, MA: National Geographic Learning.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology, 8*(1), 34-52.

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System: An International Journal of Educational Technology and Applied Linguistics, 33*(4), 575-591. doi:10.1016/j.system.2005.04.002.

Hincks, R. (2015). Technology and learning pronunciation. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 505-519). Malden, MA: Wiley Blackwell.

Hincks, R., & Edlund, J. (2009). Promoting increased pitch variation in oral presentations with transient visual feedback. *Language Learning & Technology, 13*(3), 32-50.

Hirst, D., & Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 1-44). Cambridge: Cambridge University Press.

Hong Kong Examinations and Assessment Authority (HKEAA). (1999). *HKDSE English Language: Recommended texts for the school-based assessment component*. Retrieved from http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng-SBA_Recommended_Texts-091008.pdf

Iwasaki, B., Rasinski, T., Yildirim, K., & Zimmerman, B. S. (2013). Let's bring back the magic of song for teaching reading. *Reading Teacher, 67*(2), 137-141.

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

Kenworthy, J. (1987). *Teaching English pronunciation*. London: Longman.

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in learning new words in an unfamiliar language. *Studies in Second Language Acquisition, 34*(3), 415-444. doi:10.1017/s0272263112000137

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*(2), 145-164.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Malden, MA: Blackwell Publishing.

Ladefoged, P. (2005). *Vowels and consonants: An introduction to the sounds of languages* (2nd ed.). Oxford; Malden, MA: Blackwell.

Lehmann, J. A. M., & Seufert, T. (2018). Can music foster learning - Effects of different text modalities on learning and information retrieval. *Frontiers in psychology, 8*. doi:10.3389/fpsyg.2017.02305.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly, 33*(1), 37-63.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*(4), 505-524. doi:10.1016/j.system.2004.09.009

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and speech, 43*(4), 377-401.

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory Cognition, 42*(1), 41. doi:10.3758/s13421-013-0342-5.

Morrill, T. (2011). Acoustic correlates of stress in English adjective-noun compounds. *Language and speech, 55*(2), 167-201. doi:10.1177/0023830911417251

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using SPSS* (6th ed.). Maidenhead, UK: Open University Press.

Paquette, K. R., & Rieg, S. A. (2008). Using music to support the literacy development of young English language learners. *Early Childhood Education Journal, 36*(3), 227-232. doi:10.1007/s10643-008-0277-9

Pennington, M. C. (1996). *Phonology in English language teaching: An international approach*. London; New York, NY: Longman.

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition, 35*(2), 242-253.

Rainey, D. W., & Larsen, J. D. (2002). The effect of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception, 20*(2), 173-186. doi:10.1525/mp.2002.20.2.173.

Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. New York; London: Continuum.

Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in psychology, 2*, 167. doi:10.3389/fpsyg.2011.00167

Schoepp, K. (2001). Reasons for using songs in the ESL/EFL classroom. *The Internet TESL Journal, 7*(2), retrieved from http://iteslj.org/Articles/Schoepp-Songs.html.

Setter, J. (2006). Speech rhythm in World Englishes: The case of Hong Kong. *TESOL Quarterly, 40*(4), 763-782.

Shen, C. (2009). Using English songs: An enjoyable and effective approach to ELT. *English Language Teaching, 2*(1), retrieved from http://www.ccsenet.org/journal/index.php/elt/article/view/341/305.

Tompkins, G. E. (2013). *Language arts: Patterns of practice* (8th ed.). London; Boston, MA: Pearson Education.

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. London: Routledge.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1471-1485. doi:10.1037/0278-7393.20.6.1471.

Yu, V. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research, 39*(4), 323-344. doi:10.1007/s10936-009-9142-2.

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America, 123*(6), 4498-4513.

Appendix I: Lyric Reading Test

I understand, dear	not analysed	
It’s all been grand, dear	not analysed	
But now it’s time to sleep	not analysed	
Bed* bed I couldn’t go to bed	7	(3.08s)
My head’s too light to try to set it down	9	(3.71s)
Sleep sleep I couldn’t sleep tonight	7	(3.37s)
Not for all the jewels **in the** crown	6 (exclude ‘jewels’)	(4.71s)
I could have danced all night	5	(3.20s)
I could have danced all night	5	(3.32s)
And still have begged for more	5	(5.60s)
I could have spread my wings	5	(3.15s)
And done a thousand things	5	(3.25s)
I’ve never done before	5	(6.32s)
I’ll never know what made it so exciting	10	(6.44s)
Why all at once my heart took flight	7	(5.93s)
I only know when he began to dance with me	11	(8.91s)
I could have danced danced danced all night.	7	(7.57s)
	94 syllables analysed	

*The words in bold represent the syllables analysed for P and L.