Aspects of second language speech prosody: data from research in progress

Juhani Toivanen
Diaconia University of Applied Sciences, Finland
juhani.toivanen@diak.fi

Abstract

In this paper, data from three preliminary studies concerning prosody in second language (L2) speech is described. Firstly, ways of quantifying L2 speech prosody are described. Secondly, a new way of formally describing L2 speech prosody is described. Thirdly, an experimental situation is introduced proving that L2 speech prosody is a many-faceted phenomenon, which is affected by several factors.

Introduction and ongoing research

Prosodic features of second language speech have not been extensively described in SLA (second language acquisition) literature. Recently, however, there has been increasing interest in L2 prosody (Toivanen, 2001; Hincks, 2004), and it seems that the development of an appropriate methodological apparatus in this research area is an important prerequisite of progress. In this paper, the focus is on three separate research projects dealing with different aspects of L2 speech prosody. The data was collected at Oulu University, Finland, during 2000-2005, and some of the results were presented in Toivanen & Henrichsen (2006). In the present paper, a systematic overview is presented, along with a discussion of further implications.

The data set

L2 speech prosody variation using different scales, a framework for L2 speech prosody description, and an experimental scenario involving contextually relevant L2 speech prosody variation are presented in the following sections.

L2 speech prosody variation: some quantitative alternatives

In studies on second language acquisition, prosody, if dealt with at all, is often described in an anecdotal and impressionistic way. On finds descriptions such as “narrow voice range”, “flat pitch”, etc. These descriptions are often pedagogically relevant, and they may make the transcription system more accessible to the non-experts but a considerable amount of subjectivity is a corollary of this approach. However, even a more phonetically oriented descriptive system based on “more objective” labels such as high/modal/low mean, range and variability of pitch may be confusing if the analysis is not based on any concrete anchor values or baseline data.

While non-numerical, non-experimental investigations of L2 speech prosody have an important role in the study of situated language use, for example, a more quantitative approach is also needed. The first approach is to describe pitch range with the linear Hertz scale. A number of acoustic studies of pitch range, mostly dealing with L1, have utilized this strategy but the problem is that this scale fails to make an appropriate normalization for the non-linearity of pitch perception: a larger change in frequency at the higher absolute pitch range is needed to produce the same perceptual effect as a smaller change at the lower absolute pitch range. Thus with the linear scale, comparisons of pitch ranges between males vs. females in pitch range are almost pointless.

The second option is to convert the Hertz values into semitone values; the logarithmic semitone scale has been extensively used in investigations of L1 pitch range but even this scale is not completely appropriate from the viewpoint of perception.

The third, and evidently the best, strategy is to use ERB measurements (Equivalent Rectangular Bandwidth). The ERB scale is based on the frequency selectivity of the human auditory system and the scale is perceptually more relevant than either the linear Hertz scale or the logarithmic semitone scale (Hermes & van Gestel, 1991).

Toivanen (2001) investigated the prosody of Finnish English L2 speech in an experimental setting involving native English speech as baseline data. Two groups of speakers, advanced L2 English speakers and native speakers of British English (near-RP) read out a set of short standard texts (of the Rainbow Passage type), and the recorded speech data was analyzed acoustically.

Pitch range was described with the semitone scale and the ERB scale; the linear scale was used in some preliminary comparisons. A number of unsystematic differences in pitch variation between the two groups were found with the linear scale, while the semitone scale produced much more consistent differences. The most systematic differences throughout the data, however, were detected using ERB measurements. The ERB scale enabled the conclusion that pitch variation in Finnish English L2 speech is indeed significantly more limited than in native English speech. Clearly, the type of scale used for pitch analysis is critical, and it seems obvious that in comparative cross-linguistic investigations of the prosody and pitch variation in L2 speech, the ERB scale should be considered as a first choice.

Phonological transcription of L2 speech prosody

ToBI labeling is commonly used in the prosodic transcription of (L1) English, and good inter-transcriber consistency can be achieved as long as the voice quality represents normal (modal) phonation. Certain discourse situations and varieties of English, however, probably involve voice qualities different from modal phonation, and the prosodic analysis of such speech data with traditional ToBI labeling may be problematic. Typical examples are breathy, creaky and harsh voice qualities. Pitch analysis algorithms, which are used to produce a record of the fundamental frequency (f0) contour of the utterance to assist the ToBI labeling, yield a messy or lacking f0 track on non-modal voice segments. Non-modal voice qualities may represent habitual speaking styles or idiosyncrasies or they are characteristics of emotional discourse. Typically, non-modal voice segments occur in Finnish speech, as well as in the L2 English of Finns.

In Toivanen & Henrichsen (2006), a 4-Tone Emotional Voice Transcription Framework was introduced. The framework is intended for transcribing the prosody of modal/non-modal voice in (emotional) English speech. As in the original ToBI system, intonation is described as a sequence of pitch accents and boundary pitch movements (phrase accents and boundary tones). The original ToBI break index tier (with four strengths of boundaries) is also used. The fundamental difference between the 4-tone framework and the original ToBI is that four main tones (H, L, h, l) are used instead of two (H, L). In the 4-tone framework, “H” and “L” are high and low tones, respectively, as are “h” and “l”, but “h” is a high tone with non-modal phonation and “l” a low tone with non-modal phonation. Basically, “h” is “H” without a clear pitch representation in the f0 contour record, and “l” is a similar variant of “L”.

To assess the usefulness of the 4-tone descriptive framework, informal interviews in English with Finnish students at a university of applied sciences were used. The speakers talked about their exchange studies experiences

abroad. The discussions were recorded in a sound-treated room; the speakers' speech data was recorded directly to hard disk (44.1 kHz, 16 bit) using a high-quality microphone. The speech data consisted of 574 orthographic words (82 utterances) produced by three female students (20-27 years old). Five Finnish students of linguistics/phonetics listened to the tapes; the subjects transcribed the data prosodically using the 4-tone descriptive framework. The transcribers had been given a short training session in the 4-tone style labeling. Each subject transcribed the data material independently of one another.

As in the evaluation studies of the original ToBI, a pairwise analysis was used to evaluate the consistency of the transcribers: the label data of each transcriber was compared against the labels of every other transcriber for the particular aspect of the utterance. The 574 words were transcribed by five subjects; thus a total of 5740 (574x10 pairs of transcribers) transcriber-pair-words were produced. The following consistency levels were obtained: presence of pitch accent 73 %, choice of pitch accent 69 %, presence of phrase accent 82 %, presence of boundary tone 89 %, choice of phrase accent 78 %, choice of boundary tone 85 %, choice of boundary tone 85 %, and choice of break index 68 %.

The level of consistency achieved for the 4-tone descriptive framework was somewhat lower than that reported for the original ToBI system. However, the differences in the agreement levels seem quite insignificant bearing in mind that the 4-tone system uses four tones instead of two. Importantly, it can be concluded that a descriptive system of speech prosody especially tailored for L2 speech seems feasible. In Finnish English speech, "l" typically and systematically occurs, often with a decelerating speech tempo, in the vicinity of a transition relevance place, with or without a change of speaker. A traditional ToBI-based transcription system would seem to miss an important point here.

Speech situation and the prosody of L2 speech

The third aspect of L2 speech prosody to be dealt with is the effect of the speech situation on the pitch range and variation. The speech data was produced by seventeen Finnish students of business administration at a university of applied sciences (all females in their early twenties). The subjects took voluntary Spanish courses as part of their general language studies. The subjects had studied Spanish 3-5 years on an average, and they could be described as semi-fluent in ordinary L2 language use situations. The speakers read out a short emotionally charged (joyful) passage of some 50 words from a Spanish translation of a well-known Finnish novel. Each subject read out the passage nine times in two different sessions on separate days; the speakers were allowed to read out the text at their own pace with suitable breaks between the readings. The instructions were given by two different persons. In the first session, the instructions (basically stating that the text and its repetitions should be read out in a manner "natural and comfortable" to the speaker) were given (in Spanish) by a Finnish person: a female college lecturer in her thirties teaching Spanish to the subjects at the time. In the second session, the instructions were given (in Spanish) by a native speaker of Spanish, a female in her thirties, who the speakers had not met before. The speakers' speech was data was recorded directly to hard disk (44.1 kHz, 16 bit) using a high-quality microphone.

The total set of materials consisted of 17x9x2 (306) tokens (passages). The data was analyzed acoustically with CSL (Kay Elemetrics) in terms of the following speech measures: speaking fundamental frequency (f0), f0 range and jitter. Each f0 value from the pitch analysis was converted to ERB using the formula given by Hermes & van

Gestel (1991). ANOVA was used for statistical analysis of the Instructor (2) x Repetitions (9) design. Instructor effects were significant for f0, f0 range and jitter. For each measure, the "Spanish Instructor" condition produced a higher value than the "Finnish Instructor" condition with every repetition. Repetitions effects were significant for f0, f0 range and jitter. For each measure, the value became progressively higher with continued repetitions. Interaction effects were significant for f0, f0 range and jitter. For each measure, the differences between the Spanish Instructor condition and the Finnish Instructor condition became greater as the repetitions progressed.

The instructor/interlocutor or the targeted audience clearly affected the L2 speech prosody in this setting: more lively prosody could be observed with the native speaker. It also seems that the non-native speakers needed some time to "get going" prosodically. Larger pitch variation was produced as the text got progressively more familiar. The most lively speech prosody occurred when the non-native speakers accosted to a native speaker and had overcome the initial tension. In the research on the topic, a large amount of jitter is generally associated with natural relaxed communication (Scherer, 1995) – a trend found in the present investigation as well.

All in all, these findings support the conclusion that L2 Spanish speakers are sensitive to their interlocutors. There is some evidence elsewhere that L2 speakers become more hesitant (prosodically) when they address a listener with the same L1 background (Takahashi, 1989).

Discussion and conclusion

The points and research data presented in this paper have touched upon some aspects that are relevant when the prosody of L2 speech is discussed. On the one hand, one should be aware of the parameters with which prosody can be described. Should they be exact and rigorously defined in L2 speech prosody research, or can one do with a more impressionistic apparatus? On the other hand, should one, perhaps, develop new methods and analytic frameworks for L2 speech prosody research in general? Are the current tools entirely functional? Finally, one should realize that L2 speech prosody is highly dependent on the situational factors. One could hypothesize that there are more factors involved than in most situations with native language speech prosody.

References

Hermes D. & van Gestel, J.C. (1991). The frequency scale of speech perception. *Journal of the Acoustical Society of America*, 90, 97-103.

Hincks. R. (2004). Standard deviation of f0 in student monologue. *Proceedings of FONETIK 2004*. Stockholm University: Department of Linguistics, 132-135.

Takahashi, T. (1989). The influence of the listener on L2 speech. *Variation in Second Language Acquisition*, 1, 66-80.

Toivanen, J. (2001). *Perspectives on Intonation: English, Finnish and English Spoken by Finns*. Frankfurt am Main: Peter Lang GmbH.

Toivanen, J. & Henrichsen, P.J. (eds.). (2006). *Current Trends in Research on Spoken Language in the Nordic Countries*. Oulu University Press.