The effect of teaching prosody awareness on interpreting performance: an experimental study of consecutive interpreting from English into Farsi

Mahmood Yenkimaleki & Vincent J. van Heuven

Published online: 12 May 2017.

Submit your article to this journal ⬚

Article views: 301

View related articles ⬚

View Crossmark data ⬚

Routledge
Taylor & Francis Group

The effect of teaching prosody awareness on interpreting performance: an experimental study of consecutive interpreting from English into Farsi

Mahmood Yenkimaleki[a] and Vincent J. van Heuven[a,b]

[a]Leiden University Centre for Linguistics, Leiden, The Netherlands; [b]Dept. Hungarian and Applied Linguistics, University of Pannonia, Veszprém, Hungary

ABSTRACT

This study investigates the effect of prosodic feature awareness training on the quality of interpreting by interpreter trainees. Two groups of student interpreters were formed. Participants were assigned to groups at random, but with equal division between genders (seven males in each group). The control group was then taught interpreting skills by the routine curriculum, while the experimental group spent part of the time instead on theoretical explanation and practical exercises emphasizing prosodic differences between Farsi and English. Three raters assessed the quality of the interpreter trainees' performance in a post-test in terms of accuracy, omissions, overall coherence, grammar, expression, word choice, terminology, accentedness, pace and voice. The results show that prosodic feature awareness training did have a statistically significant effect on the quality measures: the overall assessment of the experimental group was 14 points better (on a scale between 0 and 100) than that of the control group. Moreover, the difference was larger for the phonetic/prosodic quality scales (accentedness, pace, voice) than for the other scales. These results have implications for designers of curricula for training interpreters, material producers and all who are involved in foreign-language study and pedagogy.

1. Introduction

Prosody is the ensemble of properties of speech that cannot be understood from the mere linear sequence of segments (e.g. van Heuven, 2017; Nooteboom, 1997). As a first approximation, prosody includes word and sentence stress, word tones, phrasing and intonation. Prosody helps listeners to process and understand the incoming message. Therefore, it has an important role in speech communication (e.g. Cutler, 2012). Phonological awareness is the awareness of, and the ability to manipulate, the sound structure of language (Miller & Schwanenflugel, 2006). Prosodic feature awareness would then be the conscious capability of a language user to perceive, interpret and manipulate the prosody of a language, i.e. its

temporal and melodic properties. Tunmer, Herriman, and Nesdale (1988) suggested that prosodic feature awareness develops later than awareness of properties of vowels and consonants. Shankweiler and Crain (1986) argued that conscious metalinguistic awareness of phonemic structure (i.e. vowels, consonants and syllables) develops as a by-product of language acquisition. Thompson, Tunmer, and Nicholson (1993) considered language an object of thought rather than the natural result of language acquisition. Instead, they claimed that the processing of one's second language is contingent on conscious awareness of explicitly learned differences between the native and the second or foreign language. Therefore, prosodic feature awareness is not an automatic result of (second) language learning, and can only be acquired through formal instruction (Huang, Lin, & Su, 2004).

Jackson and O'Brien (2011) claimed that the relationships between prosody, second-language speech production and comprehension are understudied. Systematic studies should be conducted to learn how interpreters may exploit the relationships between prosody and meaning when decoding messages in the source language and encoding the same message in the target language. Such studies would also help second-language learners in general. The present experimental study investigates the effects of explicit prosodic feature awareness training of interpreter trainees on the quality of their interpreting performance.

2. Phonetics and interpreting

Translating (written text) and interpreting (spoken text) are complex skills used by a bilingual intermediary when two (groups of) language users, A and B, who do not understand each other's language, want to communicate. The go-between takes text in the source language, A, as input and outputs a faithful semantic equivalent of the decoded message in the target language, B. Translators may avail themselves of dictionaries, terminology lists and encyclopedias to find the best possible translation equivalents. Interpreters generally have to deliver their rendition of the input speech on the spot, with minimal delay. Although interpreters can conduct background research in preparation for an interpreting task, there is no time to consult any resources once the interpreting has begun.

Derwing, Munro, and Wiebe (1998) showed that awareness training emphasizing word stress led to better intelligibility of second-language learners. Later, those learners were found to transfer their acquired perceptual skills to spontaneous L2 speech production as well (Field, 2005). Interpreters may be taught to take advantage of prosodic properties of the source language so as to exploit redundancies in the input and resolve ambiguous utterances (Huber, 2005). In this way, prosodic feature awareness training can be beneficial to interpreters in both speech production and speech recognition.

3. Prosodic consciousness in message perception

Prosodic features are among the first to be picked up by infants acquiring their first language. The acquisition of prosody may begin before birth, since the unborn child is exposed *in utero* to the low-frequency sounds produced by the mother, which carry information on the melody and rhythm of the mother's speech. Appealing to a first in–last out principle of learning and unlearning, adult second-language learners soon acquire an adequate perceptual representation and pronunciation of the segments of the foreign

language (L2), but experience great difficulties in replacing the prosodic properties of the native language (L1) with those of the L2 (Wanner & Gleitman, 1982). Prosody plays an instrumental role in the decoding and encoding of meaning. Segmentation of continuous speech into syllables, words and phrases, informing syntactic structure, and emphasizing content words and other salient information are some of the functions of prosody that facilitate the processing of speech (Whalley & Hansen, 2006). For successful decoding of input speech and encoding speech output in the non-native language, the L2 learner may benefit from an explicit comparison of the prosodic properties of his or her native language and those of the L2. Since the prosodic features that exist in the source language contribute to the message, they should be adequately expressed in the target language as well (Ahrens, 2004).

4. Stress

By stress, we mean the property of a syllable to be more prominent than other syllables within the same word. In most languages that have stress, the location of the stressed syllable can be derived from the application of a simple rule, or can be looked up in a pronunciation dictionary. Accent, in contrast to stress, is a property of a word that makes it prominent within a larger prosodic domain (e.g. a phrase or utterance). In most languages, including those under study in the present paper, it is difficult to predict which words will be accented; the choice depends on the communicative intentions of the speaker (cf. Bolinger's (1972) article 'Accent is predictable if you're a mind reader'). Prosody training should make the learner aware of the differences between the L1 and the L2 that determine which syllable is stressed at the word level and which words are accented at the sentence level. The L1 of the participants in the present study, Farsi, typically stresses the penultimate syllable in the word. The stress system of English is more complex (e.g. Kager, 1989). Farsi learners of English cannot routinely stress the penultimate syllable in English but should learn the stress pattern for each English word separately and store it in their bilingual mental lexicon. Stress and accent placement also differ across languages in phonetic terms. Stressed syllables may differ from unstressed syllables in loudness, duration, spectral expansion and the size and segmental alignment of pitch changes (e.g. van Heuven, 2014). Different languages may make different selections from these correlates, or weigh them in different ways. For instance, Indian speakers of English execute a late rise–fall on the stressed syllable (as in Hindi), which is perceived by British listeners as stress on the next syllable, so that the word *character* (initial stress) is perceived as *director* (stressed on the second syllable [Bansal, 1966]).

Awareness of prosodic differences between native and foreign language should pay off when interpreting from the L2 into the L1. In this process, explicit knowledge of the L2 prosody would help the interpreter process the incoming speech. In normal speech processing, under optimal listening conditions, prosody is redundant vis-à-vis the segmental information. This is why most conventional orthographies do not mark word prosody. Prosody will assume a more important role, however, when segmental information is unreliable – due to a noisy communication channel. Crucially, word recognition and speech understanding are severely compromised if incorrect (and therefore misleading) prosody is added to segmentally imperfect speech (e.g. van Heuven, 2008; Wingfield, 1975).

Speech in a non-native language is inherently noisy, since the input speech does not match the deep-rooted expectation pattern of the non-native listener (Cutler, 2012). We hypothesize that drawing the non-native listener's attention to the specific characteristics of the L2 prosody (e.g. its stress system) via intensive exposure to words with unexpected stress patterns, and/or explicitly pointing out prosodic differences between in the L1 and L2, will help the non-native listener process the L2 input speech. We further assume that knowing how to exploit the redundancies imparted by word and sentence prosody in the L2 input pays off, especially when the speech processing task is aggravated by the time pressure and heavy demands on working memory inherent to interpreting.

Cognitive psychologists consider awareness a fundamental precondition to learning, and even claim that learning is impossible without conscious awareness (Brewer, 1974; Dawson & Schell, 1987; Lewis & Anderson, 1985). In the field of foreign-language learning these views are echoed by, for instance, Bialystok (1978), who proposed a theoretical framework in which conscious knowledge plays a key role. In a similar vein, Rutherford and Sharwood Smith (1985) asserted that drawing the learner's conscious attention to the formal properties of the foreign language can be advantageous to second-language learning.

Correct pronunciation and use of prosody is important for non-native speakers addressing a native audience. Native listening is extremely robust and can deal with highly deficient speech input. However, when both the segmental information deviates by more than a critical amount from the native norm and the word and sentence prosody is defective, speech recognition and understanding are bound to break down (Cutler, 2012). The importance of pronunciation in foreign-language learning has been acknowledged by researchers and language instructors (e.g. Derwing, 2003; van Heuven, 1986; Hişmanoğlu, 2006; Wang & Lu, 2011).

The present study addresses the importance of awareness of the stress system of English for Iranian interpreter trainees. In the experiment, English is the non-native source language while Farsi (modern Persian) is the native target language. The experiment involves default direction interpreting (also called *recto* or straight interpreting). We will experimentally test the claim that prosody awareness training, at the word and sentence level, will lead to improved performance by Iranian interpreter trainees when asked to interpret spoken English into Farsi.

5. Method

5.1 Participants

Thirty second-year BA students (14 males) of English Translation and Interpreting at the State University of Arak (Iran) participated. They were divided into two classes of 15 (7 males in each group), one of which would serve as the experimental group and the other as the control group. All participants were aged between 20 and 22 and were native speakers of Farsi.

5.2 Procedures

In order to ascertain that the two groups had command of English at the beginning of the study, all participants took a pre-test of general English proficiency. The test battery used

was the standard Longman's Test of English as a Foreign Language (TOEFL) (paper-based version, http://www.ets.org/toefl/pbt/about/content/), which tested the learners' skills as follows:

(1) Listening comprehension: 30 questions about short conversations, 8 questions about longer conversations, 12 questions about lectures or talks (scores range between 31 and 68 points)
(2) Reading comprehension: 50 questions about reading passages (scores between 31 and 67 points)
(3) Structure and written expression (SWE): completing 15 sentences correctly and identifying 25 errors (scores between 0 and 6 points)
(4) Writing: one essay with a length of 250 to 300 words (scores between 31 and 68 points).

The final test score ranges between 310 and 677. We report the raw scores on the four components as well as the overall TOEFL score (after weighting and conversion).[1]

The control group received routine instruction in interpreting, i.e. the routine syllabus that has been used in the English Translating and Interpreting Department of Arak State University. Different techniques, aspects and types of interpreting are normally instructed and practiced. The experimental group spent 20 minutes less time per session on the routine curriculum and instead received awareness training on prosodic features of English (stress at word and sentence level). Altogether, each group took part in 18 sessions (two hours per session and one session every week) for a total of 36 hours of instruction (for a detailed survey of the contents of the lessons see Appendix 2 in Yenkimaleki, 2017, pp. 52−88). In both classes, authentic extracts from spoken English (news, political discussions, social interviews) were presented to the students, who then interpreted the extracts consecutively.

The same post-test was administered to the control and experimental groups alike to measure the quality of (consecutive) interpreting at the end of the treatment. The test included ten 30-second extracts (samples in Appendix) to be interpreted. In the choice of extracts, attention was paid to sentences in which stress at the word and/or sentence level affected meaning. For instance, in 'The market is an institution in which wealth acquires power. Wealth controls what gets produced and who gets it.' *wealth* was accented on first mention but was de-accented in the second sentence, to indicate that the listener should not look for a new referent.

The post-test took place in a language laboratory in the presence of a classroom instructor. Students were seated in sound-proofed half-open cubicles. Source texts were presented over loudspeakers at a comfortable listening level; note taking was allowed. After every fragment, participants were given two minutes to consult their notes and to record an interpretation in Farsi. Recordings were made directly onto a digital computer through individual, table-mounted microphones.

The same three out of the ten recorded texts per participant (see Appendix) were evaluated independently by three experts, who were senior colleagues in the Department of English in the Humanities Faculty of Tafresh University, Iran – i.e. a different university than the one that hosted the experiment. The judges were experienced instructors in interpreting between English and Farsi, and did not know the students they judged. Evaluation criteria (based on Sawyer, 2004, see Table 1) were explained beforehand. The order in

Table 1. Criteria used in the quality judgment of interpreting performance. The numbers are the maximum score that could be awarded per criterion. The overall maximum equals 100. After Sawyer (2004).

Meaning		Language use		Presentation	
Accuracy	20	Grammar	7	Pace	10
Omissions	15	Expression	7	Voice	10
Overall coherence	10	Word choice	7		
		Terminology	7		
		Accent	7		

which the 30 student interpreters were rated was the same for all judges; subjects in the control group were presented before any of the students in the experimental group. The three fragments selected for each subject were presented in immediate succession. The materials were played back over small loudspeakers without interruption or repetition; judges noted down their marks (one for each criterion) on paper evaluation sheets as the fragments progressed.

The ten evaluation criteria are defined and motivated as follows:

(1) *Accuracy:* interpreters should be faithful to the meaning of the source language. An optimal and complete message should be output such that the content and intent of the source language is preserved without omission or distortion.

(2) *Omission:* interpreters may intentionally omit part of the source language and concentrate on transferring the essence of the message (Jones, 2014). In our study, omissions were not counted against the interpreter as long as the interpretation preserved the content and intent of the source language; if not, they were scored as errors.

(3) *Overall coherence:* coherence is the extent to which the interpreter's output is meaningful and purposeful. Message coherence includes conceptual connectedness, evaluative and dialogical consistency and textual relatedness.

(4) *Grammar:* an attempt was made to evaluate the speech production of the participants, observing the standard structural rules of Farsi.

(5) *Expression:* utterances should be a manifestation of appropriate use of the target language given a specific target audience, e.g. in terms of formality and informality.

(6) *Word choice:* the choice of words in the target language should match the genre of the source language. The expectations of the audience (in relation to the social class they belong to) should be taken into account as well.

(7) *Terminology:* interpreters should be familiar with technical terms of the subject matter they are interpreting. An attempt was made to determine the extent to which the participants translated the technical terms when transferring the message.

(8) *Accent*: since the interpreter's intelligibility will depend on the quality of his/her pronunciation of the target language, the strength of the interpreter's accent was judged (in the case of *recto* interpreting, this criterion applies more or less vacuously, and will vary only insofar as a strong regional accent would compromise the interpreter's intelligibility).

(9) *Pace:* an intuitive judgment was made of how optimal the interpreter's rate of delivery was, i.e. neither very slow nor so fast that intelligibility would be compromised.

(10)   *Voice:* an interpreter with a pleasant and relaxed voice is more appreciated than one with a strained or nervous voice. A global judgment was made of the extent to which the interpreter's voice was appropriate for transferring the message.

6. Results

6.1 Effect of training program

Table 2 summarizes the raw component scores of the proficiency test of the control group and the experimental group.

A t-test for unrelated samples shows that none of the small differences on the pre-test and its components between the experimental and control group is significant, $t(28) = .415$ ($p = .682$) for listening comprehension, $t(28) = 1.087$ ($p = .288$) for SWE, $t(28) = 1.421$ ($p = .168$) for reading comprehension and $t(28) = -1.029$ ($p = .312$) for the overall TOEFL proficiency score.

Table 3 presents the summed ratings (averaged over the three fragments per partici-pant) given by the three experts, as well as the mean of the three raters (see Table 4 for a breakdown by rating scale).

Cronbach's alpha amounted to .997, which indicates excellent agreement among the raters. On the basis of this result, the mean rating score is considered a valid estimate of the students' performance.

An independent-samples t-test on the overall post-test scores shows that the 14-point advantage of the experimental group (71.3) over the control group (57.5) is highly signifi-cant, $t(28) = 2.4$ ($p = .001$, one-tailed).

Figure 1 plots the relationship between the pre-test (TOEFL) and post-test (mean rating) scores of the individual students in the experimental and the control group. The overall correlation between the pre-test TOEFL scores and post-test mean evaluation scores was $r = .930$ ($N = 30$, $p < .001$). Correlations computed for each group separately are even better, at $r = .964$ for the control group and $r = .943$ for the experimental group ($N = 15$, $p < .001$ in both cases). The within-group correlations are better than for the overall sample because there is a systematic difference between the groups. The regression line for the experimental group is 14 points higher than that of the control group. This illustrates that the two groups do not differ in terms of their pre-test perform-ance; the difference is in the post-test only, with better performance for the experimental group.

The sharpest test of the effect of the prosodic awareness training can be obtained by determining for each student the difference between his/her position (relative to the peer group) in the pre-test and the post-test. Accordingly, the scores on the TOEFL pre-test and the post-test scores were transformed to z-scores (i.e. by subtracting from

Table 2. Raw component and overall TOEFL scores (means and standard deviations) obtained by control and experimental groups. LC: listening comprehension, RC: reading comprehension.

	Control group						Experimental group				
	LC	SWE	RC	Writing	TOEFL		LC	SWE	RC	Writing	TOEFL
Mean	52.2	53.1	51.7	3.76	518.00	Mean	53.1	54.3	53.3	3.91	535.77
SD	4.2	3.4	3.2	0.73	46.52	SD	4.1	3.5	3.1	0.84	47.58

Table 3. Overall quality of interpreting in post-test (between 0 and 100). Ratings are listed per judge (RY, RA, RM) separately, as well as averaged over judges, for the control and experimental groups.

	Control group					Experimental group			
	RY	RA	RM	Mean		RY	RA	RM	Mean
Mean	57.6	57.4	57.4	57.5	Mean	70.9	71.3	71.8	71.3
SD	15.9	16.0	15.5	15.8	SD	15.1	15.2	15.8	15.3

each individual score the group mean and then dividing the difference by the standard deviation). The resulting z-scores (mean = 0 and SD = 1 for both the pre-test and the post-test) can be directly compared. The effect of the intervention (the training programs for the experimental and the control groups) can then be determined by computing the difference between the z-score on the post-test minus that of the pre-test. A positive z-difference will indicate that the prosody awareness training was beneficial. Since the individual pre-test and the post-test scores are strongly correlated, a t-test on the difference scores is less contaminated by error variance. The .45-z advantage of the experimental group over the control group is highly significant, $t(28) = 4.062$ ($p < .001$).

6.2 Specific effects of awareness training on prosodic components of tests

The TOEFL pre-test and the post-test scores were a (weighted) average of scores obtained by each student on specific groups of test items or components addressing different sub-skills. For instance, the TOEFL pre-test comprised separate test components to assess the students' listening comprehension, reading comprehension, grammatical knowledge and writing ability. The post-test scores were based on ratings for ten performance scales addressing different aspects of the students' interpreting performance. Some of these scales relate directly to prosody (e.g. accent, pace and voice) while others clearly do not (e.g. use of grammar, choice of words and terminology).

We will now evaluate the effect of the prosody awareness training on the students' interpreting performance more specifically by analyzing the component scores on the post-test separately. Prosody awareness training should have a beneficial effect on the

Table 4. Mean and SD of scores on ten aspects of consecutive interpreting by control and experimental groups ($N = 15$ per group). The three scales at the bottom address prosodic components. The difference between the means (experimental–control), and the associated t- and p-values (df = 28, two-tailed) are specified.

| Scale	Range	Control group		Exp. group		Difference		
		Mean	SD	Mean	SD	Mean	t	p
Accuracy	1–20	11.6	3.1	14.4	3.1	2.7	2.444	.021
Omissions	1–15	9.5	2.4	11.0	2.4	1.6	1.790	.084
Coherence	1–10	7.6	1.4	7.2	1.6	−.4	−.728	.473
Grammar	1–07	4.3	1.0	4.8	1.1	.6	1.455	.157
Expression	1–07	4.0	.9	4.7	1.1	.7	2.030	.052
Word choice	1–07	4.2	1.2	4.8	1.2	.6	1.428	.164
Terminology	1–07	4.0	1.2	4.9	1.2	.9	2.178	.038
Accent	1–07	3.4	1.5	6.3	.8	2.9	6.693	<.001
Pace	1–10	4.6	1.6	6.7	1.5	2.1	3.817	.001
Voice	1–10	4.4	1.9	6.6	1.7	2.1	3.165	.004
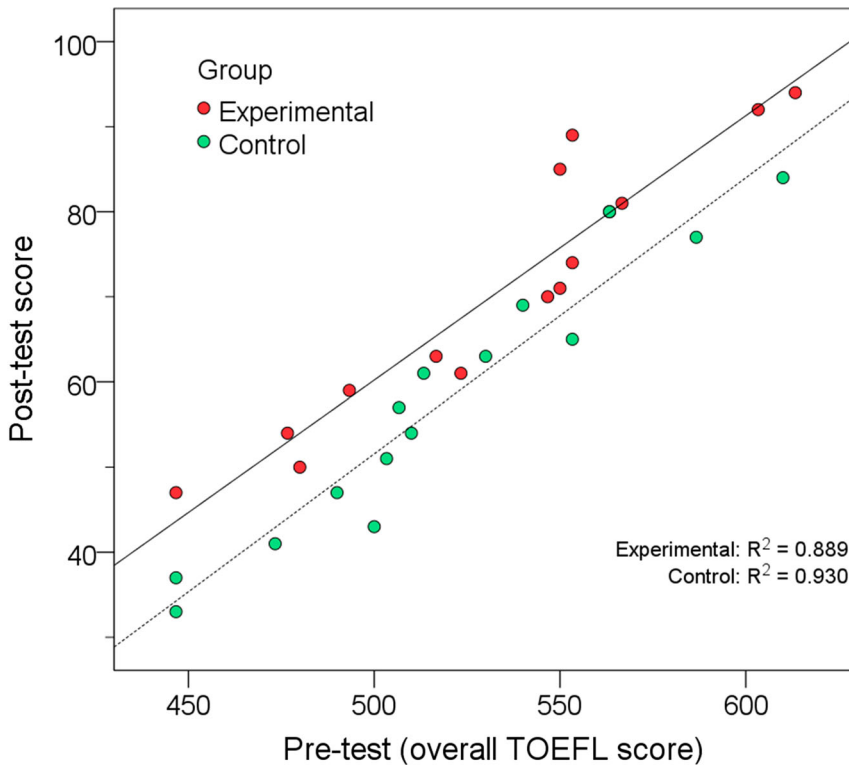
Figure 1. Post-test score plotted against the pre-test TOEFL score for 30 subjects. Linear regression lines have been drawn for each group.

rating scales that relate to prosody, rather than on the other, non-prosodic rating scales. Table 4 lists the mean ratings (and standard deviations) obtained by the 15 students in the control group (left) and the 15 in the experimental group (right).

Table 4 shows that, with the exception of one scale, i.e. coherence, the experimental group outperformed the control group on all aspects of consecutive interpreting. Moreover, the improvement obtained for specific prosody-sensitive aspects of interpreting performance is definitely larger, in terms of statistical separation (i.e. magnitude of t) than for the non-prosodic rating scales. The improvement on the three prosodic scales is indeed significantly larger than that on the other seven scales, $t(8) = 3.367$ ($p = .005$, one-tailed).

The overall conclusion of this section must therefore be that the experimental training program leads to a broad, general improvement in (nearly) all aspects of consecutive interpreting performance, but boosts prosodic skills most. This latter finding is what we would hope to find from a training program that addresses prosodic skills and awareness.

6.3 Predicting consecutive interpreting performance by multiple regression

Earlier correlation analyses showed that the student's post-test score can be predicted quite well from his/her overall TOEFL score, and that the correlation increases further when computed separately for the experimental and control groups. Using multiple

Table 5. Correlation matrix of five independent predictors (and overall TOEFL score) and post-test as the criterion variable. Cells contain the r-value and (in parentheses) the p-value of the coefficient ($N = 30$ per cell).

	Group	TOEFL	LC	SWE	RC	Writing
TOEFL	.191 (.312)					
Listening comprehension	.078 (.681)	.954 (<.001)				
SWE	.201 (.286)	.971 (<.001)	.901 (<.001)			
Reading comprehension	.259 (.166)	.930 (<.001)	.868 (<.001)	.896 (<.001)		
Writing	.102 (.592)	.941 (<.001)	.938 (<.001)	.927 (<.001)	.830 (<.001)	
Post-test	.419 (.021)	.930 (<.001)	.853 (<.001)	.936 (<.001)	.889 (<.001)	.887 (<.001)

linear regression analysis, we will now examine the extent to which the five predictors identified afford an even better prediction of the student's interpreting performance. The correlation matrix of the five predictors and the criterion is given in Table 5. The overall TOEFL score is also included in the table, but since it is fully predictable from the four component scores it cannot be entered as a predictor in the regression analysis.

The best single predictor of the post-test score is the student's score on the SWE component of the TOEFL test ($r = .936$). Table 5 also shows that the components of the TOEFL test are very strongly intercorrelated, so that little improvement remains when additional components are added. The student's group membership correlates moderately but significantly with the post-test score, indicating that belonging to the experimental group leads to a higher score on the post-test. Moreover, the intercorrelations between group membership and the components of the TOEFL test are small and insignificant. This makes group membership a good second predictor of the student's final score on the post-test.

When all five predictors are entered together, the model explains 95% of the variance in the post-test scores ($R = .974$). Using stepwise multiple linear regression yields an optimal model with three predictors. The predictors are, in descending order of importance, SWE, group and writing. The remaining predictors no longer make a significant contribution. The optimal model in Table 6 accounts for 94% of the variance in the post-test scores.

The conclusion from this subsection is that the student's performance as a consecutive interpreter can largely be predicted (i.e. with some 90% accuracy) from a combination of components of the TOEFL pre-test. Whether the individual student was a member of the experimental group or the control group is the second-most powerful predictor of the post-test score, raising the prediction accuracy to ca. 95%.

Table 6. Summary of stepwise multiple regression predicting each student's post-test score from group membership (experimental, control) and components of TOEFL pre-test.

Predictors	Beta	t	p	R^2	ΔR^2
SWE	.612	4.731	<.001	.877	
Group	.266	5.446	<.001	.932	.055
Writing	.293	2.298	.030	.944	.011

7. Conclusion and discussion

In this study, the effect of prosodic awareness training at the word and sentence level on the performance of interpreter trainees was examined. Our results showed that interpreter trainees perform better after having acquired conscious knowledge of word and sentence prosody, and of the differences in prosody between their working languages. A more detailed statistical analysis showed that our prosody awareness training yielded better interpreting performance on all quality criteria evaluated in this study, with the exception of textual coherence. Most of the effects of the awareness training were small and statistically insignificant (including the negative effect of the training on textual coherence). Crucially, the effects of the prosody training on prosody-related evaluation criteria (accentedness, pace and voice) were strong – and significantly stronger than the effects on the other seven criteria. This clearly indicates that the training is specific and that the students' improved performance is not due to some halo-effect caused by the novelty of this part of the curriculum. Rather, we would argue that the gain in performance is obtained because (as per Whalley & Hansen, 2006) better awareness of prosodic cues in the (non-native) input speech facilitates the listener's task of breaking up the incoming stream of sound into syllables, words and phrases, informs syntactic structure, and emphasizes salient content words.

It has been shown before that a closer approximation of the prosody of native English yields better intelligibility and comprehensibility of non-native speech (e.g. Derwing et al., 1998; Field, 2005). We recently showed that our prosody training program was successful in boosting the quality of the speech output in so-called inverse consecutive interpreting, i.e. from native Farsi into non-native English (Yenkimaleki, 2017; Yenkimaleki & van Heuven, 2016c). In inverse interpreting, of course, the increased awareness of the prosodic requirements of English is directly observable in the output of the interpreting process. In the present experiment, which targeted straight interpreting from foreign into native language, the effects of prosody are indirect, since they primarily affect the processing of the non-native input speech. Interestingly, the effects of the prosody training program we found are statistically stronger for the present study, i.e. the effect sizes are larger for interpreting into the native language than into the foreign language. Although we do not precisely know why this should be so, additional experiments with independent but similar groups of Iranian participants have shown that the same prosody training also yielded significantly better performance on English speech comprehension tests (Yenkimaleki & van Heuven, 2016a) and on English word recognition tests (Yenkimaleki & van Heuven, 2016b). We interpret these findings as unequivocal support for our claim that increased prosody awareness facilitates the processing of non-native input speech.

There used to a popular belief that knowing two languages very well is enough for a language user to produce a high-quality translation or (e.g. Lambert, 1978). More recently, however, it has been shown that perfect (early) bilinguals are not necessarily able translators or proficient interpreters (Grosjean, 2001). We would argue, therefore, that an excellent command of two working languages is a necessary but not a sufficient condition for high-quality translating and interpreting. That knowing how to process and understand the foreign input language is a precondition to interpreting is shown convincingly by our finding that the general level of proficiency in English is of overriding importance when it comes to predicting the quality of the student's interpreting performance.

The quality assessment of our student interpreters ranged between 33 and 94 points on a scale from 0 to 100. The exact location of the individual score within this range could be predicted with 88% accuracy from one single component score the student had obtained half a year earlier on the TOEFL test. Given this overwhelming effect of the TOEFL score, it is all the more surprising that the prosody awareness training still added 5% accuracy to the prediction of the students' performance. One might wonder, at this juncture, why the SWE test in the TOEFL battery should supply the best prediction of the student's interpreting performance rather than, for instance, the student's listening comprehension score. The latter component would relate more directly to processing the input speech as part of the interpreting process. The point is that the four component scores of the TOEFL test we administered are strongly intercorrelated, with r values ranging between 0.830 and 0.857. Differences within this narrow range are statistically insignificant and therefore not meaningful. All one can say is that all components of the TOEFL test are equally influential. That the SWE component is singled out as the first predictor in the optimal model (see Table 6) depends on the accidental circumstance that SWE correlates slightly (but insignificantly) better with the criterion than the other components, which in turn cannot add substantially to the prediction; this is a consequence of the 'winner takes all' principle underlying the stepwise regression model.

The four language skills tend to be strongly correlated in foreign-language proficiency tests (e.g. Liu & Costanzo, 2013; Poelmans, 2003). Given the roughly equal attention that the foreign-language curricula give to production and reception skills in both the visual and spoken language modalities, foreign-language learners are not likely to become highly proficient in just one skill rather than in all four. Moreover, there is substantial interaction between the skills: new words and structures that the student meets while reading will transfer to listening, and at least part of what is learnt through reading and speech perception will transfer to writing and speech production. The SWE module of the TOEFL test does not target the student's writing skill per se (this is done more directly by the writing module of the test). Rather, SWE tests the student's knowledge of English grammatical structure by asking the participant to detect ungrammaticalities and to complete unfinished structures. By doing so, the module tests the student's use of grammatical and lexical knowledge to predict how an English sentence develops over time. Being able to predict how a sentence will go on is an important subskill of word recognition and sentence comprehension (e.g. Rost, 2002).

Our results clearly show that prosody awareness training contributes substantially and significantly to the quality of interpreting from foreign into native language by native speakers of Farsi. The effects of prosody training will differ for other native–foreign language pairs, depending on the linguistic and phonetic similarity of the prosodic systems involved. The word and sentence prosody of English and Farsi would diverge more from one another than, for instance, German and English, but not as much as English and French. The pedagogical implication of this study would be to incorporate effective prosodic activities into the interpreter training curriculum, at least in Iran. This can be done by exposing students to authentic materials spoken in English with an abundance of word and sentence stresses that occur in unusual positions from the Farsi point of view. Whether the exposure should be supplemented with explicit explanations of the prosodic differences is an open question that should be addressed in

future research. Future research should also test the added value of computer-aided instruction in the area of prosody awareness training.

Gut, Trouvain, and Barry (2007) argued that instructors should employ different methodologies depending on the native and the target languages of the learners. One way to improve pronunciation and prosody would be for second-language learners to record their own speech, listen to it, compare their imitation with the native model, and repeat the exercise until they can no longer detect a difference between model and imitation (e.g. Bissiri & Pfitzinger, 2009; Sundström, 1998). There is, however, a limit to what discrepancies foreign-language learners can detect between a native model and their own imitation of it. The imitation process can be enhanced by providing auditory and/or visual feedback on the learner's attempts. For instance, prosodic errors can be auditorily detected by the learner by applying 'prosody conversion', i.e. replacing the native speech melody with the learner's imitation, or vice versa (e.g. Hirose, 2004; Nagano & Ozawa, 1990; Sundström, 1998). Providing visual feedback would be another powerful tool. Here, a (stylized) representation of the learner's imitation of the speech melody is superimposed onto the native model, so that the learner sees where the discrepancies are and what can be done to approximate the native model contour more closely (e.g. de Bot, 1983; Hardison, 2004; Su & Tseng, 2015).

Disclosure Statement

No potential conflict of interest was reported by the authors.

Funding

Notes on contributors

Mahmood Yenkimaleki is a guest scholar at Leiden University Centre for Linguistics, and former head of translation and interpreting studies department at Tafresh University, Iran. His area of interest is interpreting studies and applied linguistics.

Vincent J. van Heuven (BA, MA and PhD from Utrecht University) is an emeritus professor of experimental linguistics and phonetics and former director of the Leiden University Centre for Linguistics. He is now a professor at the University of Pannonia in Veszprém, Hungary. He served on the editorial boards (and as associate editor) of the *Journal of Phonetics* (Academic Press/Elsevier) and *Phonetica* (Karger) and was editor of the series *Speech Research* (Mouton de Gruyter). Over 40 doctoral dissertations have been written under his (co-) supervision. He is a life member of the Royal Netherlands Academy of Arts and Sciences.

References

Ahrens, B. (2004). *Prosody beim Simultandolmetchen*. Frankfurt am Main: Peter Lang.
Bansal, R. (1966). *The intelligibility of Indian English* (Ph.D. dissertation). University of London.
Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28*, 69–83.

Bissiri, M. P., & Pfitzinger, H. R. (2009). Italian speakers learn lexical stress of German morphologically complex words. *Speech Communication*, *51*, 933–947.

Bolinger, D. W. (1972). Accent is predictable (if you're a mind-reader). *Language*, *48*, 633–644.

de Bot, K. (1983). Visual feedback of intonation 1: Effectiveness and induced practice behavior. *Language and Speech*, *26*, 331–350.

Brewer, W. (1974). There is no convincing evidence for operant or classical conditioning in adult humans. In W. Weimer & D. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 1–42). Hillsdale, NJ: Erlbaum.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* Cambridge, MA: MIT Press.

Dawson, M., & Schell, A. (1987). Human autonomic and skeletal classical conditioning: The role of conscious cognitive factors. In G. Davey (Ed.), *Cognitive processes and Pavlovian conditioning in humans* (pp. 27–55). Chichester: Wiley.

Derwing, T. M. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, *59*, 547–567.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399–423.

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Malden, MA: Blackwell.

Gut, U., Trouvain, J., & Barry, W. J. (2007). Bridging research on phonetic descriptions with knowledge from teaching practice – The case of prosody in non-native speech. In J. Trouvain & U. Gut (Eds.), *Non-native prosody* (pp. 3–21). Berlin: Mouton de Gruyter.

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

van Heuven, V. J. (1986). Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In J. van Oosten & J. F. Snapper (Eds.), *Dutch linguistics at Berkeley, papers presented at the Dutch linguistics colloquium held at the university of California, Berkeley on November 9th, 1985* (pp. 67–84). Berkeley, CA: The Dutch Studies Program, U.C. Berkeley.

van Heuven, V. J. (2008). Making sense of strange sounds: (mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing*, *2*, 39–62.

van Heuven, V. J. (2014). Acoustic correlates and perceptual cues of word and sentence stress: Mainly English and Dutch. In C. Gussenhoven, Y. Chen, & D. Dediu (Eds.), *Proceedings of the fourth international symposium on tonal aspects of languages* (pp. 211–217). Baixas: International Speech Communication Association.

van Heuven, V. J. (2017). Prosody and sentence type in Dutch. *Nederlandse Taalkunde*, *22*, 3–29, 43–46.

Hirose, K. (2004). *Accent type recognition of Japanese using perceived mora pitch values and its use for pronunciation training system.* Proceedings of the international symposium on tonal aspects of languages, Beijing, 77–80.

Hişmanoğlu, M. (2006). Current perspectives on pronunciation learning and teaching. *Journal of Language and Linguistic Studies*, *2*, 101–110.

Huang, C., Lin, C., & Su, B. (2004). The effects of phonological awareness training on technological university students' phonics and vocabulary knowledge. *Journal of National Taipei Teachers College*, *17*, 59–90.

Huber, D. (2005). Phonetic aspects in translation studies. In A. Eriksson & J. Lindh (Eds.), *Proceedings of FONETIK* (pp. 49–50). Göteborg: Department of Linguistics, Göteborg University.

Jackson, C. N., & O'Brien, M. G. (2011). The interaction between prosody and meaning in second language speech production. *Die Unterrichtspraxis/Teaching German*, *44*, 1–11.

Jones, R. (2014). *Conference interpreting explained.* New York: Routledge.

Kager, R. W. J. (1989). *A metrical theory of stress and destressing in English and Dutch*. Dordrecht: Foris.

Lambert, W. E. (1978). Some cognitive and sociocultural consequences of being bilingual. In J. E. Alatis (Ed.), *International dimensions of bilingual education* (pp. 214–229). Washington, DC: Georgetown University Press.

Lewis, M. W., & Anderson, J. R. (1985). Discrimination of operator schemata in problem solving: Learning from examples. *Cognitive Psychology*, 17, 26–65.

Liu, J., & Costanzo, K. (2013). The relationship among *TOEIC*® listening, reading, speaking, and writing skills. In D. E. Powers (Ed.), *The research foundation for the TOEIC tests: A compendium of studies: Volume II* (pp. 1–25). Princeton, NJ: Educational Testing Service.

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, 98, 839–853.

Nagano, K., & Ozawa, K. (1990). *English speech training using voice conversion*. Proceedings of the first international conference on spoken language processing, Kobe, 1169–1172.

Nooteboom, S. G. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 640–673). Oxford: Blackwell.

Poelmans, P. (2003). *Developing second-language listening comprehension: Effects of training lower-order skills versus higher-order strategy*. Utrecht: LOT.

Rost, M. (2002). *Teaching and researching listening. Applied linguistics in action*. London: Longman.

Rutherford, W. E., & Sharwood Smith, M. (1985). Consciousness-raising and universal grammar. *Applied Linguistics*, 6, 274–282.

Sawyer, B. (2004). *Fundamental aspects of interpreter education. Curriculum and assessment*. Amsterdam: Benjamins.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition*, 24, 139–168.

Su, C.-y., & Tseng, C.-y. (2015). *A phonetics based computer aided prosody training system for L2 English learning*. Proceedings of the 18th international Congress of phonetic sciences, Glasgow. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0536.pdf

Sundström, A. (1998). *Automatic prosody modification as a means for foreign language pronunciation training*. Proceedings of an ISCA workshop on speech technology in language learning, Marholmen, 49–52.

Thompson, G. B., Tunmer, W. E., & Nicholson, T. (Eds.). (1993). *Reading acquisition processes*. Clevedon: Multilingual Matters.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, 23, 134–158.

Wang, R., & Lu, J. (2011). Investigation of golden speakers for second language learners from imitation preference perspective by voice modification. *Speech Communication*, 53, 175–184.

Wanner, E., & Gleitman, L. R. (1982). *Language acquisition: The state of the art*. New York: Cambridge University Press.

Whalley, K., & Hansen, J. (2006). The role of prosodic sensitivity in children's reading development. *Journal of Research in Reading*, 29, 288–303.

Wingfield, A. (1975). The intonation–syntax interaction: Prosodic features in perceptual processing of sentences. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 146–160). Berlin: Springer.

Yenkimaleki, M. (2017). *Effect of prosody awareness training on the quality of consecutive interpreting between English and Farsi*. Utrecht: LOT.

Yenkimaleki, M., & van Heuven, V. J. (2016a). Effect of explicit teaching of prosodic features on the development of listening comprehension by Farsi-English interpreter trainees. *International Journal of English Language Teaching*, 4(6), 32–41.

Yenkimaleki, M., & van Heuven, V. J. (2016b). The effect of prosody teaching on developing word recognition skills for interpreter trainees. *Journal of Advances in Linguistics*, 7, 1101–1107.

Yenkimaleki, M., & van Heuven, V. J. (2016c). The effect of prosody awareness training on the performance of consecutive interpretation by Farsi-English interpreter trainees: An experimental study. *Asia Pacific Translation and Intercultural Studies*, 3, 235–251.

Appendix 1. Audio extracts evaluated in the post-test

1. The market is an institution in which wealth acquires power. Wealth controls what gets produced and who gets it. So, for example, if a rich person wants a Mercedes Benz to be produced and shipped to him and you want a Mercedes Benz produced and shipped to you, guess what? Your wish has no power because you don't have any wealth.
2. In an all-out attack against a trapped and mostly defenceless civilian population, Israeli air raids have pounded Gaza for eight days. A thousand wounded and the Palestinian death toll at 139, more than half of these deaths civilian, including more than 22 children dead, another 180 wounded as well as three Palestinian journalists who have been killed by Israeli fire while in their vehicle. The Israeli death toll is at five.
3. Are the climate deniers right? Are some scientists colluding with the government to hide the truth about the climate change? Yes, according to top British scientist Kevin Anderson. But not the scandal you have heard about. Top scientists and government reports won't tell you, we are heading towards catastrophic climate change. Emissions are skidding out of control, leading us to a world perhaps six degrees hotter on average, much faster than anyone thought possible. Why doesn't the public know?