Effects of attention to segmental vs. suprasegmental features on the speech intelligibility and comprehensibility of the EFL learners targeting the perception or production-focused practice

Yenkimaleki, M.; Heuven, V.J.J.P. van

System 100 (2021) 102557

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Effects of attention to segmental *vs.* suprasegmental features on the speech intelligibility and comprehensibility of the EFL learners targeting the perception or production-focused practice

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

[a] *University of Nahavand, Iran*
[b] *University of Pannonia, Hungary*
[c] *Leiden University, the Netherlands*

ARTICLE INFO

Keywords:
Segmentals
Suprasegmentals
Intelligibility
Comprehensibility
Speaking skills

ABSTRACT

The present study investigated the training benefits of segmental *vs.* suprasegmental aspects for the intelligibility and comprehensibility of spoken English as a Foreign Language (EFL). Five groups of lower intermediate EFL learners were formed by random assignment. A pretest of the speaking skills, intelligibility, and comprehensibility of the learners' speech confirmed that the five groups were homogeneous before starting the training program. The control group listened to authentic audio tracks in English, and discussed their contents, watched authentic English movies, and did exercises to improve speaking skills without receiving explicit segmental and suprasegmental explanations and exercises. The experimental groups received an explanation of segmental or suprasegmental features (during one-sixth of the teaching time) followed by production-focused or perception-focused practice (during another one-sixth of the teaching time). The total instruction time was the same for all five groups, i.e., 15 h. Students then took a posttest in speaking skills targeting their speech intelligibility and comprehensibility. The findings revealed that the speech intelligibility of learners who received segmental training followed by production-focused practice was better than that of all other groups. Learners who received suprasegmental instruction followed by production-focused practice outperformed all other groups in terms of comprehensibility.

1. Introduction

Pronunciation instruction is in a state of resurgence now, charged extensively by recent studies in the arena of segmental and suprasegmental studies on English as a Foreign (or Second) Language (EFL and ESL, respectively) in different parts of the globe (Levis et al., 2016; Yenkimaleki & Van Heuven, 2019c, 2020). Different approaches to pronunciation instruction were employed. For example, the purpose of Saito (2011) was to identify and prioritize problematic pronunciation areas for particular EFL learners, while Yenkimaleki and Van Heuven (2019a) investigated the relative benefits of the implicit *vs.* explicit approach to pronunciation teaching.

Couper (2006) concluded that the effective pronunciation teaching involves: a. making learners aware that there is a difference between what they say and what native speakers say. b. helping learners to hear the difference and practice it. c. helping learners to

* Corresponding author. University of Nahavand, Nahavand, Iran.
E-mail addresses: myenkimaleki@gmail.com (M. Yenkimaleki), v.j.j.p.van.heuven@hum.leidenuniv.nl (V.J. van Heuven).

https://doi.org/10.1016/j.system.2021.102557
Received 13 August 2020; Received in revised form 16 May 2021; Accepted 19 May 2021
Available online 14 June 2021

discover useful patterns and rules. d. giving feedback and providing opportunities for further practice.

It has been shown that the oldest methods of teaching pronunciation, involving exercises in elocution, imitation drills, and reading aloud, failed to work in actual communicative language use (Cohen et al., 1991), and that accuracy of pronunciation varies according to the type of practice learners are engaged in (Dickerson, 1975; Kissling, 2018). Effective practice has to move beyond the simple identification and mimicking of decontextualized sound contrasts to the perception of more communicative characteristics of the input (Wong, 1987; Yenkimaleki, 2017, 2018), and the ability to move beyond accurate production of discrete sounds to integrating those sounds into effective communication. We know that, as students become more skilled in using segmental and suprasegmental features in the target language, their orientation to spoken input may shift towards noticing how particular characteristics of English speech carry particular meanings (Kennedy & Trofimovich, 2010). Pronunciation instruction attunes listeners to the sound system of the target language, but it can also strengthen their bottom-up processing skills, in particular speech segmentation and word identification, which leads to better word recognition and speech understanding in the L2 (Kissling, 2018).

Research shows that students' speech intelligibility and comprehensibility in EFL/ESL context may be enhanced through pronunciation instruction (e.g., Gordon & Darcy, 2016; Levis et al., 2016; Munro & Derwing, 1995; Yenkimaleki & Van Heuven, 2019b) but little is known about whether segmental versus suprasegmental instruction, or, for instance, a perception-focused versus a production-focused type of instruction makes a difference in effecting changes in students' pronunciation.

Therefore, it behooves us to explore the impact of segmental *vs.* suprasegmental instruction on EFL learners' speech intelligibility and comprehensibility, targeting the perception-focused or production-focused practice. The results of the study will help the instructors in EFL/ESL settings where only limited curricular time is available for instruction (whether on segmentals or on prosody) and practice, to use the instruction time in the classroom more constructively in teaching segmentals or prosody targeting either on a perception-focused or a production-focused instruction.

1.1. Review of related literature

Students learning L2 pronunciation benefit from being explicitly taught phonological form to help them notice the difference between their own productions and those of proficient speakers in the L2 community (Derwing & Munro, 2005; Yenkimaleki & Van Heuven, 2018, 2020). Explicit instruction of phonological rules makes learners conscious of segmental and suprasegmental rules that might play a key role in L2 speech intelligibility (Venkatagiri & Levis, 2007). Derwing and Rossiter (2003) also found that learners who received explicit feedback on vowel and consonant errors, improved their accuracy over time, but their overall comprehensibility did not improve.

Phonological awareness is an individual's awareness of the phonological structure, or sound structure, of words (Yenkimaleki, 2017, 2019). Phonological awareness is a construct that is measured by how well learners can perceive the structure of the L2 system. Conscious knowledge of the sounds, syllable structure, and prosody of the L2 (e.g., metalinguistic knowledge about pronunciation) is actualized by the learners as a consequence of explicit instruction.

Learners may be conscious of how a feature of L2 speech conveys meaning but, in practice, still cannot proficiently use it. For instance, an L2 learner might have conscious knowledge that in all polysyllabic English words one syllable carries the heaviest stress, as opposed to each syllable being equally prominent. Increasing learners' knowledge of the sound system of the L2 might help them segment the speech stream and identify words (Kissling, 2018). Field (2005) stated that the learner may also be aware that proficient English listeners use word stress as an important cue for word identification. Despite having knowledge of word stress in English, the learner might still fail to consistently stress the appropriate syllable. We believe that developing linguistic awareness is not enough; learners through practice (e.g., segmental and suprasegmental based exercises targeting perception or production-focused exercises) should also turn the acquired rules to automaticity (e.g., EFL learners will be able to pronounce foreign language words without occupying their mind by focusing on the learned rules of the target language sound system; this is usually the result of learning, repetition, and practice). Learners need to develop automaticity of L2 phonological and phonetic processing, and classroom practice needs to generalize to pronunciation behavior in spontaneous speech outside the classroom (Segalowitz & Hulstijn, 2005).

Some researchers have emphasized suprasegmental rather than segmental aspects in enhancing speaking skills (Avery & Ehrlich, 1992). Celce-Murcia et al. (1996) stated that if L2 pronunciation teachers give priority to the suprasegmental aspects of English, the learners' comprehensibility will improve, and also greater change may be affected in the learners' performance of the target language sound system. More recently, Hahn (2004) and Isaacs (2008) argued for the superiority of suprasegmental instruction in EFL/ESL contexts. Knowing the degree to which the various suprasegmental features affect listeners' ratings of L2 speech would substantially strengthen the foundation for pronunciation pedagogy because suprasegmental may be considered as one guiding principle for pronunciation teaching (Morley, 1991).

1.2. Intelligibility and comprehensibility

The intelligibility of a speaker, or of a speech utterance, is the degree to which a listener is able to recognize the linguistic units in the stream of sounds, and to establish the order in which they were spoken (e.g., Gooskens & Van Heuven, 2021; O'Brien-Wilson & Spaulding, 2010). If the listener does not know the language the speaker uses, the speaker's intelligibility is (close to) zero, even if the utterance(s) would be perfectly intelligible to native listeners of the language. The comprehensibility of a speaker (or of a spoken text) is the degree to which a listener is able to understand the speaker's meaning and intentions O'Brien-Wilson & Spaulding, 2010). Intelligibility, then, is the result of speech recognition, comprehensibility of speech understanding (Rogers et al., 2004). In this view, speech understanding is a process of a higher order than speech recognition (e.g., Denes & Pinson, 1963; Gooskens et al., 2010;

Gooskens & Van Heuven, 2021). In a similar vein, Smith and Nelson (1985) defined intelligibility as word/utterance recognition and comprehensibility as word/utterance meaning. A speaker's utterance can be intelligible for a listener but still hard to comprehend. In fact, comprehension is impossible, even for perfectly intelligible speech, if the linguistic units are nonsense words, as in Carroll's (1871) Jabberwocky. Munro and Derwing (1995) define intelligibility as the extent to which a speaker's message is actually understood by the listener, whereas comprehensibility is the listener's judgement of how difficult it is to understand the speaker. Smith and Nelson limit intelligibility to recognizing at the lexical level, that is, word recognition. For Munro and Derwing, the lexical level is only one aspect of how a listener understands a speaker's message. In addition to lexical intelligibility, a speaker can also be understood at the level of the message, and at the level of the intention. For example, Jenkins (2000) primarily looked at lexical intelligibility in the Smith and Nelson sense of the word. Levis (2018) describes intelligibility, in the traditional sense, as the speaker's ability to produce, for a listener, accurately decodable speech; but then he defines comprehensibility along the lines of Munro and Derwing as the subjective ease with which speech is perceived by a listener. Research on L2 intelligibility has primarily been done at the word level, and most of this research has found that segmental errors and incorrect word stress are strongly implicated in the loss of intelligibility.

Derwing et al. (1998) suggested that pronunciation instruction might enhance the intelligibility (as well as accentedness) of L2 learners' speech, as rated by native speakers of the target language. Training perception skills in pronunciation instruction may result in better production skills, especially when learners are at lower proficiency levels (Yenkimaleki & Van Heuven, 2016, 2019b), and when the training is accompanied by phonetic instruction (Sakai & Moorman, 2018).

However, Levis (2018) holds that recently the importance of intelligibility in the EFL context has been increasing, while the importance of accentedness often seems to be decreasing. Accentedness, or a foreign accent, is the "extent to which an individual's L2 speech differs from a particular variety of English" (Grant & Brinton, 2014) and is perceived to distinguish speakers from different native-like communities (Derwing & Munro, 2005, 2015). Levis emphasizes that these are independent constructs that should remain complementary in research (e.g., Munro & Derwing, 1995) while in teaching they are competing constructs for the instructors.

1.3. Current study

Because of the importance of EFL/ESL pronunciation instruction with research-based evidence, we pursue a systematic investigation of how pronunciation teaching can improve EFL/ESL speaking skills. The present study is set up to find out if instruction on segmental vs. suprasegmental properties makes EFL speech produced by Iranian lower-intermediate students more intelligible and/or comprehensible. We also aimed to find out what type of practice condition (perception focused vs. production focused) works better. Concretely we asked the following research questions:

1. Which one of two types of explicit pronunciation instruction, i.e., focusing on segmental properties *vs.* prosodic properties, enhances the speaking skills of Iranian EFL learners in terms of their (i) intelligibility, and (ii) comprehensibility?
2. Which type of practice, i.e., focusing on speech perception versus focusing on speech production, enhances the EFL learning, in terms of (i), and (ii) above?

The independent variable underlying question (1) is what we will call the linguistic *Domain* factor, i.e., segmental phonology versus prosody (also called suprasegmental phonology). We will refer to the variable underlying question (2) as the *Practice* factor, i.e., with a focus on perception versus on production. The two factors define four experimental treatments, the benefits of which can be compared against a control group condition. The difference between five groups is the (partial) type of treatment they receive which is different for each group (see Table 1 for a more detailed training program).

We propose that training explicitly targeting segmental properties will be beneficial especially for the EFL learners' intelligibility, while the prosody-oriented instruction will rather improve the learners' comprehensibility. Similarly, when the practice following the explanations focuses directly on speech production, the learner's overall speaking skills will improve more than when the practice emphasizes speech perception (Yenkimaleki, 2017; Sakai & Moorman, 2018; Yenkimaleki & Van Heuven, 2020).

Table 1
Summary of activities and time spent (minutes) by five groups of participants in the experiment.

Activity	Group				
	CON.	EXP. Segments Perception	EXP. Segments Production	EXP. Prosody Perception	EXP. Prosody Production
Audio tracks/movies & discussion	675	375	375	375	375
Listening to instructor explanations/ guidance	225	225	225	225	225
Segmental teaching/perception-focused		300			
Segmental teaching/production-focused			300		
Prosody teaching/perception-focused				300	
Prosody teaching/production-focused					300
Total time spent	900	900	900	900	900

2. Method

2.1. Participants

The participants of this study were 80 students of industrial engineering at the University of Nahavand, Iran. The participants were chosen from among 240 industrial engineering students by systematic random sampling. Since the University of Nahavand just accepts female students, all of the participants were female freshmen students (e.g., the first year of their study at the BA level). Their age range was between 18 and 20 years; their first language was Persian. They were randomly assigned to one of five groups (see below).

2.2. Ethics

We received approval for the present study from the ethics committee of the Department of Modern Languages of the University of Nahavand. All the participants agreed to take part in the research project based on informed consent and received a small amount of money for their services.

2.3. Procedure

Five groups of 16 students were formed. The first was a control group, which did not receive explicit segmental or suprasegmental instruction but instead did different types of exercises to develop their English speaking skills. They also watched authentic English materials (e.g., audios, videos), and discussed their contents in different sessions (fifteen sessions of 1 h). This is equivalent to the routine English language curriculum.

The four experimental groups spent 20 min less time per session on the routine curriculum and instead received 10-min of explicit teaching of either segmental or prosodic properties of English, followed by 10-min of practice with focus on either the perception or on the production of the phenomena that had just been explained. The materials for the treatment were American English speech fragments. The phonological phenomena to be demonstrated and explained were chosen on the basis of the authors' earlier experience and studies. The types of training also were based on the authors' previous studies (e.g., Yenkimaleki, 2016, 2017; Yenkimaleki & Van Heuven, 2018, 2020). The authors knew some features would be important in the speech intelligibility and comprehensibility of Persian learners of English as a Foreign Language (see below for examples). The students already knew that they were going to receive pronunciation training for the program. The specific type of the contents for each session was not divulged to the students before. The reason was that the instructor did not want students to prepare any materials outside of the classroom.

We will refer to the four experimental groups as Segments-perception, Segments-production, Prosody-perception, and Prosody-production. The first author was the instructor for all five groups. He has done his PhD in this area; he is specialized in pronunciation training for EFL/ESL learners, and has some ten years' experience in teaching and research in this area.

2.3.1. Segmental training

The mother tongue of the students was used sometimes in the analysis for the elaboration of the issues so that students would understand the points completely. Contrastive analysis was used in some cases for teaching the differences between the English and Persian sound systems. The book, which was used as a complementary resource, was *English Pronunciation in Use*, 10th edition (Hancock, 2003).

The segmental training proceeded in five stages, as follows. Stage 1: description and analysis. At this stage, the instructor presented the articulatory differences between two sounds. Then students were asked to sort a random list of words containing the key sounds into the two categories. They also, as a whole class activity, conducted a treasure hunt to locate classroom items containing the two key sounds. Stage 2: listening discrimination. Learners were given the opportunity to tune their ears to sound differences. Techniques applied at this stage include word- and sentence-level practice with minimal pairs, i.e., words that differ by only one phoneme such as *back* and *pack* (consonants) or *pick* and *peck* (vowels). Stage 3: monitored practice. The learners' attention was focused on correctly articulating the target sounds, with little or no use of their mother tongue. Correct articulation here refers to comprehensible and intelligible pronunciation (output) (Gooskens & Van Heuven, 2021; O'Brien-Wilson & Spaulding, 2010; Rogers et al., 2004). Examples of monitored practice include the reading aloud of sentences, tongue twisters, or short dialogues. Stage 4: guided practice. Guided practice combined a focus on the target sounds with the need for learners to use English to accomplish the task. Typical activities in this stage of practice are cued dialogues. Moreover, in this stage, the instructor also listened to how students identified a number of words that constituted pronunciation challenges for the students. The instructor wrote some words on the board and explained the contrast that was missed, e.g., *scene ~ sin, feel ~ fill.* The instructor monitored the students' performance and helped them in case of any question. The instructor provided feedback whenever it was necessary to facilitate student learning. In some cases, the students gave feedback as well. The peer feedback was welcome but monitored (corrected) by the instructor. Stage 5: communicative practice. The practice was less structured and required students to attend to both the target pronunciation features and the content of the message being communicated. Typical activities were those that require the natural production of the target language such as storytelling, role play, and problem-solving.

2.3.2. Suprasegmental training

The students received explicit teaching of prosodic features of English during each training session. Stage 1: students acquired procedural knowledge about the English language prosodic features, i.e., intuitive knowledge that cannot be verbalized (e.g., Piske,

2008; Yenkimaleki & Van Heuven, 2018) (e.g., phonetic/phonological sensitivity). Stage 2: students were involved in speech shadowing. Authentic audio files (i.e., materials not produced for teaching purposes in EFL classrooms but produced for movies, news cast, and for other communication purposes in the media) were selected with a maximum duration of 5 min; they were based on the speech of a single native English speaker (e.g., zappenglish.com). Students chose a topic that they were already familiar with and which would be of interest to them. The students listened to the audio files once quickly just to get used to the speaker. They paid attention to the speaker's rhythm, accentuation, and pace of speaking. The students were asked to shadow the input speech, i.e., to repeat the words immediately after the speaker in an online fashion, for about 30 s at a time. Students paused, tried again, and even recorded and listened back to their own versions (e.g., when somebody sings the words of a song s/he already knew well, s/he tries to imitate the speaker's pronunciation and pace as best as s/he could). This helped the students focus on how the English speakers modulated speed, used intonation, and blended words together. By mimicking the speaker, students could begin to improve their own intonation, connected speech, and overall fluency. Stage 3: students targeted problematic sounds. Throughout the lesson, the instructor listened to how students spoke, identified a number of words that reflected pronunciation challenges for the students. The instructor wrote the words on the board and explained the contrast that was missed, e.g., *steam/esteem, prayed/parade* (incorrect vowel epenthesis) or *foreign/for rain* (incorrect stress pattern). In this stage, the instructor also asked the students to transcribe and drill new vocabulary (e.g., by introducing new vocabulary, the instructor will be sure that the students will not routinely do the exercises by repeating what they had already practiced). Here, the instructor focused on aspects of pronunciation such as word stress, sentence stress (e.g., the practice of sentence stress was based on listening and producing the same stress patterns that the native speakers/instructor had used. The new vocabulary items were in new sentences, not in the sentences the students had already practiced), and intonation. New vocabulary that came up during the training program was written down. Stage 4: the instructor asked the students to contextualize their tasks (e.g., repetition of key words in a listening passage). Stage 5: having done the required practical tasks on contextualization, the students performed meaningful, authentic tasks (e.g., choice of correct word in a sentence or a sentence in a paragraph). Stage 6: students were asked to do realistic tasks (e.g., a role-play of a situation similar to one that one may face in real life or a discussion of the students' real-life situation or concerns; for details see Yenkimaleki, 2017: 50–88).

2.3.3. Assessment

Both pretest and posttest consisted of interviews that were run systematically to evaluate the participants' speech intelligibility and comprehensibility. The raters were two English language lecturers at the University of Nahavand who were native speakers of Persian, and one American native speaker in Hamadan, Iran, who held an MA in linguistics from the University of Arizona, USA. Two of the raters are specialized in pronunciation training and did their PhD in this area. Both of them have lived outside of Iran when working on their PhD dissertations, one in The Netherlands, and one in the UK. The third rater was an American native speaker who studied linguistics and was employed as a teacher of English as a foreign or second language. We wanted to have at least one native speaker as a rater so as to add to the quality of the assessment, since in some cases the nuances of pronunciation can be better recognized by native speakers. Judgements by native and non-native raters need to be consistent with each other. Ekmekçi (2016) and Zhang and Elder (2014) concluded that native and non-native English language teachers display almost identical rating behavior in assessing EFL students' oral proficiency. The first step in our data analysis will be to check whether this is also the case in our study.

The pretest and posttest consisted of six questions. The questions were descriptive in nature so that world knowledge of the students was not involved in answering the questions, e.g., *describe the historical sites of the city*. All the questions were open ended, and students had 1 min to answer each question. The same prerecorded questions were asked to all the participants (students were in different rooms, and they could not see each other after answering the questions). The questions in the pretest and posttest were different so that students would not remember any point from the pretest (to control the test effect). Raters, by consulting with each other, wrote twelve questions for the pretest and posttest of the study before the program started. Six questions (out of twelve questions) were randomly chosen for the pretest, and six for the posttest. Recordings were made directly onto a digital computer through individual, table-mounted microphones.

The interviewers used an assessment sheet that addressed two general components: intelligibility and comprehensibility.

1. *Intelligibility*. The experts were asked to imagine how easy it would be for a native English listener to write down the sentences spoken by the learner by way of dictation (speech recognition task). They indicated the degree of intelligibility on an 11-point scale from 0 (impossible to recognize even a single word) to 10 (all words can easily be recognized in the correct order).
2. *Comprehensibility*. The experts were instructed to imagine how difficult it would be for a native English listener to reconstruct the meaning of the utterances and the speaker's intentions, on an 11-point scale from 0 (speaker makes no sense at all) to 10 (speaker's meaning and intentions will be perfectly understood).

The third component that had to be judged was of a composite nature. Raters were instructed to judge the following three aspects of the participant's speaking skill:

3a. *Accentedness*. Raters indicated the perceived strength of the non-native accent in the EFL speech of the participant on a 6-point scale from 0 (very strong non-native accent) to 5 (authentic native English accent).[1]

[1] We assume that the raters interpreted scale 3a *Accentedness* as predominantly targeting the segmental quality of the speaker's EFL, since the nonsegmental properties were targeted separately and explicitly by the scales (3 b-c).

3b. *Word stress.* Raters judged how often the correct syllable was stressed in each polysyllabic word in the participants' speech production, on a 6-point scale from 0 (words very often have the stress on the wrong syllable) to 5 (all words have the stress on the correct syllable).

3c. *Sentence prosody.* Raters estimated the correctness of the sentence prosody (sentence stress and intonation) on a 6-point scale from 0 (sentences prosody deviates very strongly from the English norm) to 5 (sentence prosody is as in native English speech).

The lower granularity of the scales used in (3) was meant to speed up the rating process. In the analysis of the results, the scores on these scales were multiplied by 2, so that all judgement scales ran between 0 and 10.

3. Results

3.1. Agreement of raters

The first stage of the data analysis consisted of a test of the reliability of the three raters. We computed Cronbach's alpha for the three raters for each of the five rating scales separately for the pretest and the posttest. Cronbach's alpha coefficient is a measure of between-rater agreement. For all scales computed (both for pretest and for posttest scales), alpha was in excess of the commonly used minimum threshold for reliability, i.e., alpha >.7 (alphas ranged between 0.730 and 0.957).[2] The results are shown in Table 2. In this table, we also list the mean score given by each rater for each of the scales.

Table 2 shows that, generally, the three raters do not deviate much from one another in terms of the mean ratings per judgement scale. In the pretest there is no difference in the overall ratings among the three raters, although small discrepancies between raters are observed when we consider the scales separately. In the posttest, there is a general effect for rater 1 (R1) to be stricter than the other two, while R2 (the native listener) is rather more lenient than the two nonnatives on the Comprehensibility scale and in his assessment of Word stress. Given the satisfactory overall agreement among the raters, we decided to carry out subsequent analyses on the ratings averaged over the three raters.

3.2. Analysis of effects and interactions

The results of the experiment are summarized in Table 3, which presents the mean scores obtained by the five groups of learners, on the pretest and on the posttest, for each of the five rating scales. Moreover, the difference between the pretest and posttest scores, i.e., the gain or delta, is specified in the next two columns of the table. We checked the normalcy of the distributions post-hoc. The distribution of scores on all scales is approximately (and sufficiently) normal for parametric testing. All distributions were unimodal with skew and kurtosis values that do not invalidate the assumptions made by ANOVA. All statistical tests were performed with an alpha level of 0.05.

The differences between the means were tested with a separate one-way Analysis of Variance (ANOVA) for each rating scale, with learner group as a fixed between-participants factor, assuming a normal distribution of scores and equal variances across learner groups. As expected, there are no significant differences between the five learner groups for any of the rating scales before the treatment started, with mean scores between 5.19 and 5.56 for Intelligibility, between 5.46 and 5.63 for Comprehensibility, between 5.71 and 6.21 for Accentedness, between 4.54 and 5.21 for Word stress, and between 4.46 and 5.04 for Sentence prosody, $F(4, 75) < 1$ for all three scales ($.465 < p < .989$).

As shown in the gain columns of Table 2, the posttest scores were always better than the corresponding pretest scores, with a range between .04 and 2.75 on a scale from 0 to 10. However, two out of the 25 differences between pretest and posttest failed to reach significance, as estimated by t-tests for matched samples. Non-significant gains were observed only for the control group, on two scales, i.e., Comprehensibility and Accentedness.

Now that we have established that the gain is significant in the large majority of the conditions, we will from here on concentrate the analysis on differences in gain, as a function of the rating scale and learner group. This was done by a separate one-way ANOVA for each rating scale as well as for the overall gain, i.e., the unweighted mean of the five scales, with the five learner groups as a single fixed factor. Differences between pairs of means were determined by a post-hoc analysis using Duncan's correction for multiple comparisons ($\alpha = .05$). When means in the same row of Table 4 share the same superscript letter, they do not differ significantly from one another. Effect size is expressed as partial eta squared, $p\eta^2$.[2] The results are shown in Table 4. A graphic representation of the results is seen in Fig. 1A–F.

The ANOVAs bear out that there are significant effects of instruction type on the gain obtained for all five performance scales, as well as for the overall gain. The gain in the control condition is small for most of the criteria that were judged, and smaller than the gain found for any experimental condition, with the exception of the Sentence-prosody scale, where the control group did (insignificantly) better than the second poorest group, i.e., the group that received the Prosody-perception type of instruction. The control condition was always in the lowest gain category (as determined by a post-hoc comparison of means, and denoted by the superscript 'a' in Table 4). In terms of the overall gain, i.e., the unweighted mean of the five criteria judged, the control condition is significantly poorer

[2] Given that the between-rater agreement is high, it would be odd for within-rater agreement to be poor; it would mean that each rater would give inconsistent judgements to different speakers in the same way as the other two raters did. However, if raters are internally inconsistent in their judgements the between-rater agreement must be low.

Table 2
Mean scores per rater (R1, R2, R3) on five scales in pretest and posttest. *F* ratio, probability (*p*), and effect size (partial η^2) obtained in a repeated measures ANOVA are specified. Degrees of freedom (nominally 2, 158) were Greenhouse-Geiser adjusted before computing *p*, when the assumption of sphericity was not met. Means with the same superscript do not differ from one another (post-hoc test with Bonferroni correction for multiple comparisons). Cronbach's alpha is a measure of agreement observed among the raters.

	Pretest							Posttest						
	R1	R2	R3	*F*	*p*	pη^2	alpha	R1	R2	R3	*F*	*p*	pη^2	alpha
Intelligibility	5.3^{a}	5.4^{b}	5.4^{a}	4.4	.017	.053	.957	6.0^{a}	6.3^{b}	6.3^{ab}	12.5	<.001	.135	.933
Comprehensibility	5.6^{b}	5.6^{b}	5.4^{a}	4.3	.017	.051	.930	6.2^{a}	6.8^{b}	6.3^{a}	22.5	<.001	.222	.872
Accentedness	5.7^{a}	6.0^{a}	6.2^{b}	6.2	.003	.073	.783	6.7^{a}	6.7^{a}	6.8^{a}	.3	.707	.004	.784
Word stress	5.1^{b}	4.8^{ab}	4.7^{a}	5.5	.005	.065	.875	6.3^{a}	7.5^{c}	7.0^{b}	30.9	<.001	.281	.730
Sentence prosody	4.9^{a}	4.8^{a}	4.7^{a}	.9	.418	.011	.867	6.1^{a}	6.0^{a}	6.6^{b}	11.6	<.001	.128	.738
Overall	5.3^{a}	5.3^{a}	5.3^{a}	.8	.464	.010		6.3^{a}	6.6^{b}	6.6^{b}	22.3	<.001	.220	

Table 3
Mean and Standard Deviation of scores on five rating scales (Intelligibility, Comprehensibility, Accentedness, Word stress, Sentence prosody) obtained by five groups of learners on pretest and posttest. For each scale, the difference between posttest and pretest score is indicated under Gain. $N = 16$ for each cell. The significance of the gain was determined by a paired *t*-test and two-tailed p-values.

Rating scale	Learner group	Pretest		Posttest		Gain		Sign.	
		Mean	SD	Mean	SD	Mean	SD	*t* (15)	*p*
Intelligibility	1. Control	5.27	.85	5.71	.96	.44	.43	4.0	.001
	2. Segments, perception	5.35	.85	6.08	.76	.73	.39	7.5	<.001
	3. Segments, production	5.33	.86	7.31	.58	1.98	.51	15.6	<.001
	4. Prosody, perception	5.19	.83	5.81	.88	.63	.47	5.3	<.001
	5. Prosody, production	5.56	.92	6.04	.65	.48	.44	4.4	.001
Comprehensibility	1. Control	5.63	.93	5.67	.81	.04	.48	.3	.736
	2. Segments, perception	5.52	.83	6.79	.57	1.27	.43	11.9	<.001
	3. Segments, production	5.56	1.04	6.54	.88	.98	.64	6.1	<.001
	4. Prosody, perception	5.50	.92	5.90	.67	.40	.52	3.0	.008
	5. Prosody, production	5.46	.84	7.35	.59	1.90	.62	12.3	<.001
Accentedness	1. Control	6.29	1.09	6.42	.91	.13	.98	.5	.617
	2. Segments, perception	5.96	1.34	6.54	.82	.58	1.03	2.3	.039
	3. Segments, production	6.04	1.07	6.88	1.13	.83	.86	3.9	.002
	4. Prosody, perception	5.88	1.09	6.58	1.37	.71	.67	4.3	.001
	5. Prosody, production	5.71	1.09	7.04	.81	1.33	.77	6.9	<.001
Word stress	1. Control	4.79	.74	5.50	1.07	.71	.71	4.0	.001
	2. Segments, perception	4.54	1.32	7.29	.62	2.75	1.09	1.1	<.001
	3. Segments, production	5.08	1.11	7.46	.95	2.38	.48	19.6	<.001
	4. Prosody, perception	5.21	1.32	6.75	.68	1.54	1.19	5.2	<.001
	5. Prosody, production	4.75	1.06	7.46	.56	2.71	.75	14.5	<.001
Sentence prosody	1. Control	4.79	.74	5.75	.64	.96	.54	7.1	<.001
	2. Segments, perception	4.46	1.19	6.25	.94	1.79	.90	7.9	<.001
	3. Segments, production	5.00	.97	6.79	.85	1.79	.68	1.6	<.001
	4. Prosody, perception	5.04	1.09	5.75	.97	.71	.51	5.5	<.001
	5. Prosody, production	4.71	.99	6.63	.99	1.92	.73	1.6	<.001

Table 4
One-way ANOVA for five gain parameters as well as the overall gain, with five learner groups as a factor ($df_1 = 4$, $df_2 = 75$). Means in the same row that share a superscript letter do not differ from one another (Duncan post-hoc test with $\alpha = 0.05$).

Gain on dependent	Control	Segmentals		Prosody		*F*	*p*	pη^2
		Perc	Prod	Perc	Prod			
A. Intelligibility	$.44^{a}$	$.73^{a}$	1.98^{b}	$.63^{a}$	$.48^{a}$	32.6	<.001	.635
B. Comprehensibility	$.04^{a}$	1.27^{b}	$.98^{b}$	$.40^{a}$	1.90^{c}	28.8	<.001	.606
C. Accentedness	$.13^{a}$	$.58^{ab}$	$.83^{bc}$	$.71^{abc}$	1.33^{c}	4.0	.005	.176
D. Word stress	$.71^{a}$	2.75^{c}	2.38^{c}	1.54^{b}	2.71^{c}	15.9	<.001	.459
E. Sentence prosody	$.96^{a}$	1.79^{b}	1.79^{b}	$.71^{a}$	1.92^{b}	10.5	<.001	.360
F. Overall	$.45^{a}$	1.43^{c}	1.59^{c}	$.80^{b}$	1.67^{c}	23.3	<.001	.554

than any experimental group. We will, therefore, draw the provisional conclusion that all four experimental treatments were significantly more beneficial to our learners than the control condition. In the remainder of the analysis we exclude the control group, thereby trimming the analysis to a 2×2 design, so that we can analyze the effects of Linguistic domain (segments, prosody) and type of Training (focus on perception, focus on production) as well as the interaction between these factors for each scale. The results are

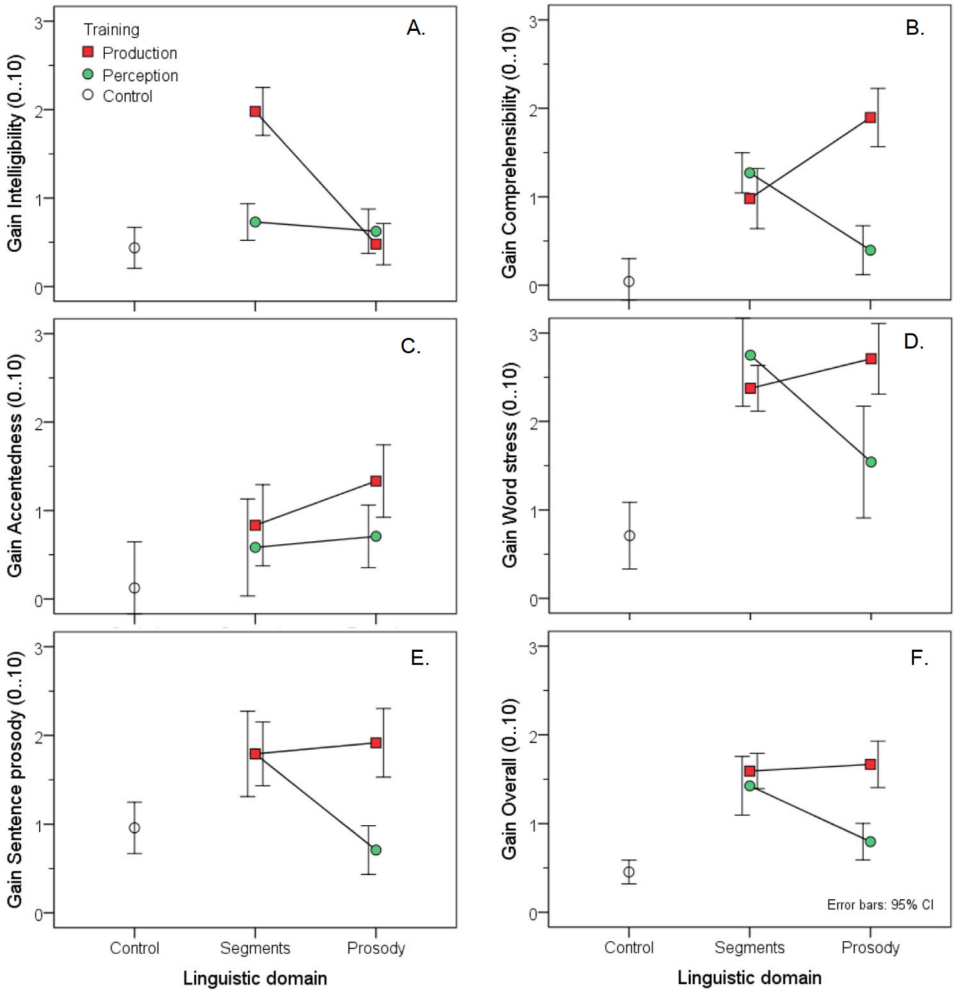

Fig. 1. Gain (posttest score minus pretest score) on five judgement scales (A: Intelligibility, B: Comprehensibility, C: Accentedness, D: Word stress, E: Sentence prosody) and the Overall gain (mean of A-B-C-D-E), broken down by Linguistic domain (segmental phenomena, prosody) and by Training practice (focus on perception, focus on production). The gain for the control condition is shown as well. Error bars are the 95% confidence interval of the mean.

shown in Table 5.

What is shared by the panels B through F in Fig. 1, and the corresponding rows in Table 5, is that there is little difference between the two types of Training (i.e., with exercises focusing on speech production versus on perception) when the explicit instruction was about segmental properties. For all performance scales B-D-E, as well as for the overall gain, there is a sizable gain due to segmental instruction relative to the control condition. In these panels, the effect of explicit explanation of prosodic features yields roughly the same gain on average, but there is a strong interaction here with the type of exercises done after the explanation: the effect of prosody is very large when the exercises focus on speech production, but much smaller when the focus is on speech perception; in panel E for the

Table 5
Two-way ANOVAs for main effects and interaction of Linguistic domain and Type of training on the gain from pretest to posttest for five performance scales and the overall performance. Letters A to F refer to panels in Fig. 1.

Gain in	Effect of Domain			Effect of Training			Interaction		
	F	*p*	pη^2	*F*	*p*	pη^2	*F*	*p*	pη^2
A. Intelligibility	50.0	<.001	.455	23.7	<.001	.283	37.9	<.001	.387
B. Comprehensibility	.0	.881	.000	18.9	<.001	.239	41.5	<.001	.409
C. Accentedness	2.2	.143	.035	4.3	.042	.067	.8	.377	.013
D. Word stress	3.6	.062	.057	3.0	.090	.047	11.3	.001	.158
E. Sentence prosody	7.1	.010	.106	11.3	.001	.159	11.3	.001	.159
F. Overall	5.4	.024	.082	18.8	<.001	.239	8.7	.005	.126

judged quality of the Sentence prosody, the gain is, in fact, less (but not significantly so) than in the control condition.

These results differ markedly from what can be seen in panel A, for the judgement of Intelligibility. Not much gain is apparent in any combination of explanation and exercises with the exception on one: the explicit explanation of segmental properties of English followed by exercises focusing on speech production (but not perception), yields a large gain in intelligibility.

We performed a final three-way ANOVA, with Judgement scale as a within-participant factor, and Linguistic domain and Training as between-participant factors. Degrees of freedom were Greenhouse-Geiser corrected when p-values were determined, but the nominal degrees of freedom are specified below. The analysis shows, first of all, that the magnitude of the gain varies significantly depending on the performance aspect judged, $F\,(4, 240) = 65.1$ ($p < .001$, $p\eta^2 = .520$). The gain is largest for the judged correctness of Word stress (2.3 points), followed by the correctness of Sentence stress (1.6 points), and around 1 point on the three remaining scales: 1.1, 1.0 and .9 for Comprehensibility, Intelligibility and Accentedness, respectively, which do not differ from one another in the Bonferroni post-hoc test with $\alpha = .05$. Importantly, the three-way interaction shows that the discrepant behavior of the Intelligibility scale relative to the other scales in Fig. 1 is highly significant, $F\,(4, 240) = 18.5$ ($p < .001$, $p\eta^2 = .235$).

4. Discussion

The findings revealed that teaching segmental contrasts followed by production-focused practice strongly improved the intelligibility of EFL learners' speech. Teaching suprasegmental features followed by production-focused practice, in contradistinction, improved the comprehensibility of EFL learners. Given that intelligibility is mainly a matter of accuracy of segmental pronunciation (i.e., the ensemble of acoustic properties of a sound in context that allow a native listener to identify it correctly) and comprehensibility is additionally determined by phrasing, accentuation and clause marking (i.e., all aspects of sentence prosody), intelligibility is a prerequisite for comprehensibility: the listener must first recognize a sufficient number of words before s/he can try to make sense of a string of words making up sentences and texts.

The findings of this study converge with Venkatagiri and Levis (2007), who stated that suprasegmental instruction is an important variable in the comprehensibility of EFL learners' speech. The greater amounts of explicit knowledge of phonological patterns and rules may have engendered greater speech comprehensibility. EFL learners with better explicit knowledge of phonological structures and patterns of English were generally rated as more intelligible speakers.

Similarly, Gordon and Darcy (2016) found that explicit pronunciation instruction focused on suprasegmental features (stress, rhythm, reductions, and linking) was more effective for increasing the comprehensibility of non-native English than explicit instruction on segmental features (vowels, consonants) and implicit instruction.

Examination of all the individual participants in the experimental groups revealed that participants exhibited general improvement in speech intelligibility and comprehensibility. As implications for instructors in second language classrooms, we suggest that instructors when teaching segmental and suprasegmental, provide students with production-focused exercises in the EFL context.

It may be pointed out that increased conscious attention in pronunciation materials to training students to monitor their production through the teaching of formal rules, noticing the differences, providing constructive feedback, and reflective activities result in the enhancement of speaking skills (Yenkimaleki & Van Heuven, 2019c). The assumption is that pronunciation improves through gradual monitoring of the acquired system based on conscious knowledge of the facts learned about the language (Crawford, 1987; Levis, 2018).

The contribution of the present study to the growing body of literature on pronunciation teaching is that we examined the separate effects of teaching segmental and suprasegmental features of English followed by different types of activities on EFL students' speech comprehensibility and intelligibility. We suggest that in teaching segmental and suprasegmental the authentic tasks would be provided for the students to make them reflect on the accuracy of their productions after the fact (e.g., post-hoc monitoring).

5. Summary and conclusions

In this investigation, we studied the contribution made by replacing a small part (one-sixth) of the routine curriculum in the teaching of English and a Foreign Language of Iranian students with Persian as the L1 by the explicit explanation of either the segmental or the prosodic structure of English and pointing out differences in the respective domains with the mother tongue of the learners for the experimental groups. The explicit explanation was followed for another one-sixth of the teaching time by exercises focusing on either the perception or on the production of the phenomena that had just been explained, yielding four groups of participants in a 2×2 factorial design, which could be compared with a control group that was exposed to the routine curriculum only. The students' performance in EFL was judged before and after the treatment on three phonological scales, i.e. authenticity of the segmental pronunciation (Accentedness), correct placement of Word stress, and appropriate use of Sentence prosody, as well as on two global performance scales of Intelligibility (ease of EFL speech recognition) and Comprehensibility (ease of EFL speech understanding).

In the pretest, none of the five participant groups obtained any ratings on any of the performance scales that differed from those obtained by any other group, so that we can conclude that the five groups were equal prior to the treatment.

The effect of the treatment was expressed in terms of the gain, i.e., the posttest score minus the pretest score for each of the five rating scales. All five groups of participants gained by their respective treatment, be it that the control group generally gained less than

each of the four experimental groups.

In terms of the phonological assessment scales, the four experimental groups benefit in approximately the same way from the various treatments. Explanation of segmental properties yields a moderate to considerable improvement on all scales, relative to the gain obtained by the control group, and it makes little difference whether the subsequent exercises focus on speech perception or production of EFL. When the explanation concentrates on prosodic phenomena, the general gain is comparable to that of the segmental explanation but now it matters what kind of exercises follow. If the prosodic exercises emphasized speech perception, the gain is very small, but when the focus was on speech production the gain is generally larger than what is seen with segmental explanation.

Intelligibility, however, is served best by the combination of explanation of segmental differences between Persian and English followed by exercises involving the production of these phenomena. Moreover, it would seem that, generally, asking students to do exercises focusing on speech perception is less beneficial than production-oriented exercises – although the difference is small when the preceding explanations were about segmental differences.

The greatest improvement on any of the performance scales judged in the present experiment was observed in the correctness of word stress, with 2.3 points on average on a scale from 0 to 10. Better word prosody will have an effect on both the intelligibility and on the comprehensibility of a non-native speaker. When the vowels and consonants are not accurately pronounced, as is generally the case in non-native English, the correct position of the stress in polysyllabic words is crucial for word recognition (e.g., Cutler, 1984; Cutler & McQueen, 2014; Van Heuven, 2008). However, correct phrasing, intonation, and realization of sentence stresses, which should be marked by prominent pitch changes on the stressed syllable of only the communicatively important words in the sentence (e.g., Van Heuven, 2018), will at the same time enhance the comprehensibility of the EFL speaker (Hahn, 2004).

Recommendations based on the present study should not be based on the overall gain observed in this study (as shown in Fig. 1F) because the averaging of all effects observed obscures what is really going on. The pedagogical implications of this study would be:

- If instructors in EFL settings want to improve the student's intelligibility, they should explain the segmental features of English and then engage the student in exercises that focus on speech production. This scenario will quite probably ensure better production of the vowels and consonants (especially in clusters) of English so that word forms will be easier to recognize.
- If it is the student's comprehensibility instructors want to improve, explanations should primarily target prosodic phenomena, and, again, subsequent exercises should focus on the production of the L2 phenomena rather than on their perception. Presumably, the correct realization of phrasing, speech melody, and the marking of sentence stress on (only) the communicatively important words are conducive to better comprehensibility.

This study has some limitations. The participants of this study were exclusively female, since the University of Nahavand is designated for educating women. Although there would be no reason, *a priori*, to assume that men would respond differently to the various treatments in our study, an additional study should be set up with both genders to verify the generalizability of this study across genders, while at the same time increasing the number of participants.

Future studies may investigate the effects of attention to segmental *vs.* suprasegmental features on developing EFL students' listening comprehension skills. Future studies also can be set up following the design of this experiment with students from different L1 backgrounds to confirm the generalizability of the results in this study.

Acknowledgement

We thank all the students and colleagues who collaborated with us in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2021.102557.

References

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford University Press.

Carroll, L. (1871). *Through the looking-glass, and what Alice found there*. London: Macmillan.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York: Cambridge University Press.

Cohen, A. D., Larson-Freeman, D., & Tarone, E. (1991). The contribution of SLA theories and research to teaching language. In *Paper presented at the Regional language Centre Seminar on language Acquisition and the second/foreign language classroom. Singapore, 22–26 April 1991*.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*(1), 46–66. Retrieved from http://hdl.handle.net/1959.14/329829. (Accessed 15 May 2021).

Crawford, W. W. (1987). The pronunciation monitor: L2 acquisition considerations and pedagogical priorities. In J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 103–121). Washington, DC: TESOL.

Cutler, A. (1984). Stress and accent in language production and understanding. In D. Gibbon, & H. Richter (Eds.), *Intonation, accent and rhythm: Studies in discourse phonology* (pp. 77–90). Berlin: De Gruyter.

Cutler, A., & McQueen, J. M. (2014). How prosody is both mandatory and optional. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, & E. van Zanten (Eds.), *Above and beyond the segments: Experimental linguistics and phonetics* (pp. 71–82). Amsterdam: John Benjamins. https://doi.org/10.1075/z.189.06cut.

Denes, P. B., & Pinson, E. N. (1963). *The speech chain. The physics and biology of spoken language*. Murray Hill, NJ: Bell Telephone Laboratories.

Derwing, T. M., & Munro, M. J. (2005). second language accent and pronunciation teaching: A research-based approach. *Tesol Quarterly, 39*, 379–397. https://doi.org/10.2307/3588486
Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.42
Derwing, T., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393–410. https://doi.org/10.1111/0023-8333.00047
Derwing, T., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*(1), 1–18.
Dickerson, L. (1975). The learner's interlanguage as a set of variable rules. *Tesol Quarterly, 9*, 401–408.
Ekmekçi, E. (2016). Comparison of native and non-native English language teachers' evaluation of EFL learners' speaking skills: Conflicting or identical rating behaviour? *English Language Teaching, 9*(5), 98–105. https://doi.org/10.5539/elt.v9n5p98
Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *Tesol Quarterly, 39*, 399–423. https://doi.org/10.2307/3588487
Gooskens, C., & Van Heuven, V. J. (2021). Mutual intelligibility. In M. Zampieri, & P. Nakov (Eds.), *Similar languages, varieties, and dialects: A computational perspective*. Cambridge University Press. https://doi.org/10.1017/9781108565080 (in press).
Gooskens, C., Van Heuven, V. J., van Bezooijen, R., & Pacilly, J. J. A. (2010). Is spoken Danish less intelligible than Swedish? *Speech Communication, 52*, 1022–1037. https://doi.org/10.1016/j.specom.2010.06.005
Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners. *Journal of Second Language Pronunciation, 2*, 56–92. https://doi.org/10.1075/jslp.2.1.03gor
Grant, L. J., & Brinton, D. M. (2014). *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press.
Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *Tesol Quarterly, 38*, 201–223.
Hancock, M. (2003). *English pronunciation in use. Intermediate self study and classroom use*. Cambridge University Press.
Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review, 64*, 555–558.
Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness, 19*, 171–185. https://doi.org/10.1080/09658416.2010.486439
Kissling, E. M. (2018). Pronunciation instruction can improve L2 learners' bottom-up processing for listening. *The Modern Language Journal, 102*(4), 653–675. https://doi.org/10.1111/modl.12512
Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation (cambridge applied linguistics series, 27)*. Cambridge: Cambridge University Press.
Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *Tesol Quarterly, 50*, 1–38.
Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *Tesol Quarterly, 25*(3), 481–520.
Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x
O'Brien-Wilson, E., & Spaulding, T. J. (2010). Effects of noise and speech intelligibility on listener comprehension and processing time of Korean-accented English. *Journal of Speech, Language, and Hearing Research, 53*(6), 1543–1554. https://doi.org/10.1044/1092-4388(2010/09-0100
Piske, T. (2008). Phonetic awareness, phonetic sensitivity and the second language learner. In J. Cenoz, & N. H. Hornberger (Eds.), *Knowledge about language: vol. 6. Encyclopedia of language and education* (pp. 155–166). New York: Springer Science.
Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech, 47*, 139–154. https://doi.org/10.1177/00238309040470020201
Saito, K. (2011). Identifying problematic segmental features to acquire comprehensible pronunciation in EFL settings: The case of Japanese learners of English. *RELC Journal, 42*(3), 363–378. https://doi.org/10.1177/0033688211420275
Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied PsychoLinguistics, 39*, 187–224.
Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll, & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford University Press.
Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes, 4*, 333–342. https://doi.org/10.1111/j.1467-971X.1985.tb00423.x
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*(1–2), 39–62. https://doi.org/10.3366/E1753854809000305
Van Heuven, V. J. (2018). Notes on the phonetics of word and sentence stress: A cross-linguistic (re-) view. In H. van der Hulst, J. Heinz, & R. Goedemans (Eds.), *The study of word stress and accent: Theories, methods and data* (pp. 13–59). Cambridge University Press. https://doi.org/10.1017/9781316683101.002.
Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness, 16*, 263–277. https://doi.org/10.2167/la417.0
Wong, R. (1987). Learner variables and prepronunciation considerations in teaching pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 13–28). Washington, DC: TESOL.
Yenkimaleki, M. (2016). Why prosody awareness training is necessary for training future interpreters. *Journal of Education and Human Development, 5*, 256–261. https://doi.org/10.15640/jehd.v5n1a26
Yenkimaleki, M. (2017). *Effect of prosody awareness training on the quality of consecutive interpreting between English and Farsi*. Utrecht: LOT. https://www.lotpublications.nl/Documents/459_fulletext.pdf. (Accessed 15 June 2021).
Yenkimaleki, M. (2018). Implicit vs. explicit prosody teaching in developing listening comprehension skills by interpreter trainees: An experimental study. *International Journal of English Language and Linguistics Research, 6*, 11–21.
Yenkimaleki, M. (2019). Explicit teaching of segmentals versus suprasegmentals in developing speaking skills by interpreter trainees: An experimental study. *Forum: International Journal of Interpretation and Translation, 17*(2), 249–268. https://doi.org/10.1075/forum.18007.yen
Yenkimaleki, M., & Van Heuven, V. J. (2016). Explicit teaching of segmentals versus suprasegmentals: Which would yield better listening comprehension skills for interpreter trainees? An experimental study. *British Journal of English Linguistics, 4*, 11–22.
Yenkimaleki, M., & Van Heuven, V. J. (2018). The effect of teaching prosody awareness on interpreting performance: An experimental study of consecutive interpreting from English into Farsi. *Perspectives. Studies in Translation, 26*, 84–99. https://doi.org/10.1080/0907676X.2017.1315824
Yenkimaleki, M., & Van Heuven, V. J. (2019a). Prosody instruction for interpreter trainees: Does methodology make a difference? An experimental study. *Across Languages and cultures. A Multidisciplinary Journal for Translation and Interpreting Studies, 20*, 117–133. https://doi.org/10.1556/084.2019.20.1.6
Yenkimaleki, M., & Van Heuven, V. J. (2019b). Effects of prosody awareness training on the intelligibility of Iranian interpreter trainees in English. *Dutch Journal of Applied Linguistics, 8*, 291–309. https://doi.org/10.1075/dujal.17023.yen
Yenkimaleki, M., & Van Heuven, V. J. (2019c). The relative contribution of computer assisted prosody training vs. instructor based prosody teaching in developing speaking skills by interpreter trainees: An experimental study. *Speech Communication, 107*, 48–57. https://doi.org/10.1016/j.specom.2019.01.006
Yenkimaleki, M., & Van Heuven, V. J. (2020). Relative contribution of explicit teaching of segmentals vs. prosody to the quality of consecutive interpreting by Farsi-to-English interpreting trainees. *Interactive Learning Environments, 1–12*. https://doi.org/10.1080/10494820.2020.1789673. In press.
Zhang, Y., & Elder, C. (2014). Investigating native and non-native English-speaking teacher raters' judgements of oral proficiency in the college English test-spoken English test (CET-SET). *Assessment in education: Principles. Policy & Practice, 21*(3), 306–325. https://doi.org/10.1080/0969594X.2013.845547

Mahmood Yenkimaleki (PhD) is an assistant professor of Interpreting Studies and Applied Linguistics at the University of Nahavand, Iran. His area of interest is Interpreting Studies and Applied Linguistics.

Vincent J. van Heuven (BA, MA, and PhD from Utrecht University) is emeritus professor of Experimental Linguistics and Phonetics and former director of the Leiden University Centre for Linguistics. He is also emeritus professor at the University of Pannonia in Veszprém, Hungary. He served on the editorial boards (and as associate editor) of *Journal of Phonetics* (Academic Press/Elsevier) and *Phonetica* (Karger) and was editor of the series *Speech Research* (Mouton de Gruyter). Over 45 doctoral dissertations were written under his (co-) supervision. He is a life member of the Royal Netherlands Academy of Arts and Sciences.