A Conceptual Approach to Teaching L2 Pronunciation: Perception of Word Stress

Graeme Couper
Auckland University of Technology
<graeme.couper@aut.ac.nz>

Abstract

Research has demonstrated that pronunciation teaching can be effective, but there have been very few classroom-based studies that have focused on the perception aspect of pronunciation. This article explains the theory and practical application of a conceptual approach and reports on its impact on perception of English word stress. The quasi-experimental study (N=18) involved four fifty-minute lessons on stress in two-syllable words in a high-intermediate ESL class. The experimental (n=10) group was tested pre, post (mid-semester) and delayed (end-semester) and the comparison (n=8) group was tested twice before receiving the instruction in the second half of the semester and being tested again. In line with a conceptual approach, the innovative testing method used moves beyond traditional identification tasks in an attempt to measure changes in participants' conceptual understandings. The tests showed large and significant gains in perception, which were retained eight weeks later. The comparison group made no gains, but after instruction improved on a par with the experimental group. This study presents an effective way of teaching pronunciation concepts, supported by theory, that also leads to improvements in perception. It informs and provides a template for both teachers and researchers who may want to replicate the study.

Keywords: perception, word stress, concepts, L2 pronunciation, pronunciation teaching, conceptual approach, critical listening, socially constructed metalanguage.

There is now a large amount of empirical evidence to show that explicit teaching can improve pronunciation (Lee et al., 2015). It is argued here that the precise nature of that explicit instruction plays a key role in its effectiveness. Underpinning such success is accurate L2 speech perception, and more recently there have been studies that have focused on learners' perception and the degree to which perception can be changed through instruction and corrective feedback (Gomez Lacabex & Gallardo del Puerto, 2014; Lee & Lyster, 2016). Studies have also tested whether or not changes in perception are accompanied by changes in production (Huensch, 2016; Lee & Lyster, 2017). Rather than continuing with the discussion on how perception and production are related, this report takes a conceptual approach, a term

coined by Fraser (reported in Couper, 2021), a Cognitive Linguistic lens, which views both perception and production as dependent on understanding the concepts underlying L2 phonological categories (Fraser, 2010). The aim here is to provide teachers with an empirically supported answer to the question of how they can help learners with perception.

The study takes the case of English word stress. It proposes and tests a tightly defined and empirically and theoretically driven type of instruction, with a view to replication. The empirical and theoretical basis for instruction is explained in the literature review that follows, where I begin with what is known about learning L2 speech categories and their underlying concepts. This is followed by a review of what is already known about how teachers can help learners with speech perception as well as a consideration of what teachers actually do in the classroom. Finally, I bring the focus to the word, the role that stress plays in word recognition and the implications for teaching English word stress. This leads to the rationale for the proposed method for teaching and testing the perception of word stress.

Review of Literature

Learning How to Perceive L2 Speech Categories

Speech categories here are taken to include the whole range of phonological categories such as phonemes, syllables, stress, and intonation. However, the vast majority of the research into learning L2 speech categories has focused on phonemic contrasts (Huensch & Tremblay, 2015). Speech perception of phonemes can be visualised as a perceptual space in which we allocate different speech sounds, or phones, to particular phonemes. As Brown (2000, pp. 6-7) observes, "successful acquisition of phonological representations requires accurate perception of phonemic contrasts in the input." This is made particularly difficult because, as is well established in phonetics (For example, see Ladefoged & Disner, 2012; Port, 2007), the nature of speech is such that it is not simply a sequence of sounds that are combined to produce words and phrases. Lively et al. (1994) argue that a large amount of information is needed to learn L2 speech categories because of a lack of acoustic invariance, linearity and segmentation. There is a lack of acoustic invariance because "each time a speaker produces a phoneme, a different acoustic form is generated" (p. 267), and this varies according to speakers, rate of speech, and phonetic context. There is a lack of linearity because "In natural speech, phonemes overlap and are coarticulated in order to achieve transmission rates of up to 10 phonemes per second" (p. 266). Finally, lack of segmentation means "context-sensitive cues for phonemes, stress, and intonation contours must be used to aid in segmentation" (p. 268).

Exposure to a wide range of exemplars is required to provide the information needed to establish or adapt categories. That is, we draw on episodic events, memories of particular examples, rather than generalisations in assigning sounds to categories (Pisoni & Lively, 1995). This understanding spurred many laboratory-based studies in which exposure to multiple speakers in multiple contexts, known as High Variability Phonetic Training (HVPT), was found to support L2 speech perception learning (Thomson, 2018). These studies have focused almost exclusively on phoneme categories, but a few have investigated perception of suprasegmental categories such as tone (Sereno & Wang, 2007), prosody (Aoyama & Guion, 2007) and word recognition (Walley, 2007). More recently Huensch and Tremblay (2015) have investigated syllable structure.

HVPT in laboratory or computer-based settings helps learners with L2 speech perception. However, a classroom-based setting offers opportunities to move beyond identification tasks

with yes/no feedback. Unfortunately, there has been little classroom-based research into teaching perception. Gomez Lacabex and Gallardo del Puerto (2014) used computer-based training in the classroom with young learners and found both "auditory discrimination/identification practice and listen-and-repeat practice" (p. 500) improved perception. Lee and Lyster (2016, p. 40) is perhaps the only study to investigate adult "classroom-based perception training including L2 instruction and oral feedback provided during instructor-student interaction." As they point out, there is a significant difference between laboratory-based perception training and a classroom context where instructor-student interaction may be used to work on perception. They used form-focused instruction aimed at getting students to notice target forms, which involved explicit instruction, input enhancement and awareness tasks. An important feature of their success was interaction with the teacher.

Concepts and Implications for Teaching

Assignment to a particular category relies on a mental representation, or concept, of the category (Murphy, 2002). In some cases, the concept is easily grasped, for example, chess pieces are either black or white and can be categorised accordingly. However, phonological categories are conceptually more difficult and forming an accurate understanding of those concepts requires multiple exposures in multiple contexts. Further, learners often need more than exposure to be able to conceptualise sounds in the same way that target language (TL) speakers do (Fraser, 2000). Cognitive Linguistics offers insights into how we can use our cognitive skills, such as the ability to notice, and compare and contrast, to learn these concepts (Mompean, 2014). Fraser (2000, p. 26) suggests that teachers need to "appreciate and imaginatively explore what the sounds seem like to the learners, gradually leading them to more appropriate ways of thinking about English pronunciation" and to explain pronunciation in relation to the way learners think about sounds rather than from the perspective of the English speaker. Beginning with the learner's perspective and their descriptions of how they perceive the TL, leads to the co-construction of commonly agreed terminology that can also be used later for feedback. The informal metalanguage that arises from this communication has been called Socially Constructed Metalanguage (SCM) (Couper, 2011). Support for this approach can be found in educational psychology where it is understood that student-teacher and student-student interaction assists in the construction of meaning (Williams & Burden, 1997) and mediation by significant others enhances learning (Vygotsky, 1978). This interaction leads to co-constructed discourse (Gibbons, 2006) and underlies Swain's (2006) concept of languaging.

In addition to the importance of effective communication between teacher and student, Fraser (2001) proposes Critical Listening exercises. In discussing how contrast facilitates concept formation, Fraser (2006, p. 91) says "The key contrast is not the contrast between members of a minimal pair, but the contrast between what the speaker intended to say and what a listener understood them to say." While there are parallels between HVPT and Critical Listening, there are also differences in that Critical Listening relies on the teacher to help learners to hear differences. Couper (2006, 2011, 2013) provided empirical evidence that Critical Listening and effective communication lead to significant improvements in learners' pronunciation of syllable codas. Couper (2011) found that the use of SCM without Critical Listening practice led to significant gains in production and smaller gains in perception, while Critical Listening without SCM led to significant gains in perception and lower gains in production. The two combined led to significant gains in both production and perception. This not only does support the value of these two methods but also emphasises the point that concepts underlie both production and perception.

Teachers

While there is evidence that noticing, explicit instruction, awareness tasks, computer-based identification tasks with corrective feedback, student-teacher interaction, and Critical Listening can all help with perception, this is often not reflected in classroom practice. Research into what teachers know and do, suggests that teachers make some attempt to teach pronunciation, but they tend to focus on production rather than perception (Couper, 2021). This teaching also tends to be teacher-centred, with little guided or free practice (Baker, 2014). A longitudinal study by Burri and Baker (2021) found that teachers recognised the value of raising awareness, but their pronunciation teaching was also mainly teacher-centred. Teachers have reported activities that might foster perception, such as minimal pairs, listening discrimination, students recording themselves, dictation, and other listening tasks (Baker, 2014; Burri & Baker, 2021; Buss, 2016; Couper 2016, 2021). Couper's (2016) report on teachers in Uruguay found they tended to use traditional listening comprehension tasks for awareness raising, which both Cauldwell (2013) and Field (2008) suggest do not lead to pronunciation improvements. Instead, they propose a bottom-up approach to bring the focus to the actual nature of sounds in context. Couper (2021) found teachers were aware that students "often can't hear the difference" (p. 17) and needed help with perception. However, only two of the 19 participants knew how they could help learners with perception and the formation of phonological concepts. This lack of teacher knowledge is a further impetus for this study. It is hoped that by finding more empirical evidence to show what works in the classroom, teachers will be able to be better prepared to help learners with their speech perception.

The Word

Word stress is the process by which we make one part, or syllable, of the word sound more prominent than others. In the case of English, stress is created through a combination of length, pitch change, and loudness. In addition, unstressed syllables also undergo changes in vowel quality, that is the vowel itself changes and it becomes shorter (Cutler, 2015; Lewis & Deterding, 2018). Stress is thought of differently in different languages with different combinations of length, pitch, and loudness (Gussenhoven, 2005). There are also different ways of directing the listener's attention to the meaning. In English, the presence of something (the stressed syllable) is highlighted through the absence of something else, in this case the full vowel sound on the unstressed syllable. All these differences underscore that the concept of stress is language specific (Gussenhoven, 2005). In terms of achieving intelligibility, failing to reduce unstressed syllables is a much more important factor in the failure to recognize words than is putting the stress on the wrong syllable (Cutler, 2015). Field (2005) found that when non-native speakers use accurate stress they are more easily understood and Trofimovich and Isaacs (2012) found inaccurate word stress was worse than inaccurate grammar in terms of how it affected intelligibility. Although Jenkins (2000) suggested that word stress might not be important for speakers using English as a Lingua Franca (ELF), more recent findings (Lewis & Deterding, 2018; Deterding, 2013) suggest word stress is also important in ELF contexts.

Therefore, learners, teachers, and textbook writers, must move beyond the traditional focus on where the stress is. They need to consider the concept of stress in English and compare it with concepts in other languages. For example, a speaker of a language that does not reduce unstressed vowels may put the stress on the correct syllable but give all syllables their full length so that to the English listener it sounds like every syllable is being stressed. Jung et al. (2017) identified three patterns of inaccurate stress "misplaced, multiple prominent stresses, absence of prominent stress" (p. 332). They also identified the role the learner's first language

plays in difficulties with word stress, which can be traced back to differences in the concept of stress. Because perception and production of word stress patterns may be difficult for L2 learners (Field, 2005; Munro & Derwing, 1995), we need to help them fully understand the precise nature of stress, not simply where it is placed. While some learners acquire this concept easily, others do not, which is where the proposed teaching approach focusing on the formation of new concepts is particularly powerful.

While we are interested in how learners produce stress, we also want to know how they perceive it, which provides insights into their understanding of the concept and helps provide a basis for teaching pronunciation. In order to understand learners' perceptions, researchers have traditionally used discrimination or identification tasks. However, Thomson (2012) asked his participants how confident they were, and Couper (2006, 2013) had lengthy discussions with participants exploring exactly how they heard different words. The current study also moves beyond a binary approach in both teaching and testing to gain better insights into perception. This informs both teachers and researchers by helping to understand learners and the stage that they are at in developing their concepts of syllables and word stress, which is explained in the next section.

Background and Aims of the Study

This study aims to see how well the conceptual approach that effectively used Critical Listening and Socially Constructed Metalanguage (SCM) in teaching syllable codas (Couper, 2011) can be transferred to teaching word stress. The key principles of this approach are:

> 1. Talk the same language (start with your learners' perceptions to make sure you are communicating effectively by using terms that you all understand in the same way - SCM).

> 2. Compare and contrast perceptions (use Critical Listening to help learners understand the difference between how they perceive the TL and how expert speakers perceive it).

> 3. Compare and contrast productions (note that good productions from students are just as good as native speaker productions)

> 4. Lots of practice and feedback are needed (using SCM helps to ensure feedback is understood).

The aim is to provide sufficient detail of the actual teaching to allow for replication of the study. The worksheets along with further explanations and teacher notes are provided in Appendix A. This study builds on an exploratory investigation by Couper (2012) and there are plans to replicate it as part of a larger project with colleagues in France, dubbed "Cognitive Phonology in Teaching and Evaluation of Perception" C-PTEP. Covid-19 has severely hampered opportunities to roll this out but I have been able to run one rendition with an intact class in New Zealand, which is being reported on here.

One of the results of taking a conceptual approach is the realisation that concept formation is not an all or nothing process. For example, Couper (2011) reported that learners' descriptions of their perceptions could be divided into more fine-grained groupings. This arose from the teaching, but clearly has implications for the most appropriate method of testing. A binary test that establishes whether or not learners have correctly identified the correct production

overlooks some of these nuances. The most fundamental one is that learners may not hear a difference, which renders their response a pure guess. Secondly, they may have some feeling for what is correct, but may not really understand what it is that is salient to the expert speaker about the correct version. Thomson (2012), also attempted to take some of this uncertainty into account by asking participants how confident they were in their responses. Couper (2011, p. 168) talked to participants as they made their responses to a Critical Listening task involving epenthesis in syllable codas and coded them according to whether or not epenthesis was heard and identified as an error:

- No = Difference not heard, no awareness of significance of extra vowel.

- Maybe = Correct answer chosen, place of error not clearly indicated, and lack of explanation to show understanding.

- Probably = Correct answer chosen, place of error pinpointed but explanation did not clearly show understanding.

- Yes = Correct answer chosen, place of error pinpointed, clear explanation provided.

In attempting to transfer this scale to perception of word stress in a way that could also be identified through computer-based responses to prompts, the following scale has been used in this study (although the distinctions between 2, 3, and 4 will not be reported on here):

0 = Cannot hear the difference, or say they find it difficult to hear it.

1 = Can hear the difference but cannot identify the best choice.

2 = Can hear the difference and identify the best choice, but no clear explanation.

3 = As for 2 but can identify part of the reason for the difference, although not the salient one.

4 = As for 2 but can also identify the salient difference (full understanding).

This leads to the following research questions:

1. How does a conceptual approach affect participants' perceptions of word stress?
2. What sorts of approaches to the measurement of perception does this teaching lend itself to?
3. How well can a conceptual approach be applied to the teaching of word stress?

Method

Research Design

The study collected pre- post- and delayed post data on speech production and perception, although only the perception data are presented here. An AB-BA design was adopted with an intact class. That is, half the class, the experimental group, received four lessons of concept focussed instruction rather than their regular instruction in the first half of the semester. The other half of the class, the comparison group, received the concept focussed instruction in the

second half of the semester. Table 1 provides an overview of the timing of tests and teaching interventions.

Table 1. Research Design.

Group	Week 2	Week 3	Weeks 4-7	Week 7	Week 10	Weeks 11-14	Week 15
Exp't *n*=10	Pre-test	Feedback[1]	Teaching[2]	Post-test *n*=10		Class	Delayed[3] post-test *n*=9
Comp *n*=8	Pre-test	Feedback	Class		Re-test[3] *n*=8	Teaching	Post-test *n*=5

Notes. [1] Feedback: All students were given feedback on pronunciation (not just on stress); [2] The intervention is referred to as "Teaching"; [3] There was an eight-week gap between post-test and delayed post-test for the experimental group and between the pre-test and the re-test for the control group.

Participants

The participants were attending a one-semester English language class in Auckland. The class was at a High Intermediate level and was designed for refugees and new migrants settling in New Zealand. Those who performed well on the course would also be in a position to attend a further one-semester pre-degree level course aimed at preparing students for undergraduate university study. There were 24 students in the class at the beginning of the course. Twenty-one students agreed to participate in the study. Their mean age at the beginning of the study was 31 years and mean time in the country was 2.5 years. They came from a wide range of countries and language backgrounds. Although no attempt is made to relate L1 to the impact of instruction, the reader might like to get a feel for the wide range of participant backgrounds. See Table 2 for an overview of age, time in the country, L1, and the record of class and test participation.

All students were given a test of both speaking and perception in the computer lab, in Week 2. They were given feedback on their pronunciation the following week, so that it served as a diagnostic for them, if they or the teacher wanted to follow up on any of the issues I identified. I did not focus on stress in this feedback, although I did not ignore it when there were problems. Participants were ordered according to their scores and then assigned alternately to one group or the other with the aim of having two groups of equal proficiency. This did not work perfectly as there was sometimes a mismatch between perception and production and because on the first day of teaching, some students initially assigned to the experimental group were not present, requiring a rapid reallocation of students who were there to the experimental group. Fortunately, the groups' mean scores on perception were quite similar: 6.9/12 for the experimental and 7.1/12 for the control group.

After four weeks of teaching, all ten students completed the post-test, and nine completed the delayed test, 8 weeks later. Numbers for the comparison group were a little lower. Of the eleven students initially assigned to that group, only eight completed both the pre-test and the second test eight weeks later. Of these only five availed themselves of the concept-focused teaching later in the semester and completed the post-test.

Table 2. Overview of participants and participation.

Student Code	Age	Time	Language	Classes attended
Experimental Group				
AE	44	4	Cantonese	3
BE	19	0.5	Japanese	4
CE	40	3.5	Arabic	3
DE	31	3	Arabic	4
EE	21	2.5	Tamil	4
FE	41	2	Tamil	4
GE	18	0.2	Japanese	4
HE	53	2	Dari	4
IE	30	3	Farsi	4
JE[1]	19	3	Mandarin	4
Comparison Group				
AC	26	1	Khmer	4
BC[2]	21	2	Japanese	0
CC[2]	34	5	Mandarin	0
DC[2]	38	6	Farsi	2
EC	19	1	Tigrinya	4
FC	40	4	Spanish	4
GC	24	2	Indonesian	4
HC	43	1	Mandarin	3

Notes. [1] missed delayed test; [2] missed post-test

The Teaching

There were four fifty-minute lessons, held in a computer lab so that participants could easily make and listen to recordings. However, they were also asked to move away from the computers for various classroom activities and interactions.

The underlying theory and rationale for this conceptual approach to teaching have already been explained, as has the approach to word stress.

The items for teaching, and testing, were two syllable words taken from the most frequently occurring members of the word families in the Academic Word List. They have been ordered according to New Zealand English stress patterns, as shown in Appendix A. The most common are: Stress on the first syllable, with a schwa or /I/ on the second syllable (Type 1); and schwa or /I/ on the first syllable, and stress on the second syllable (Type 2). I have collapsed the remaining variants under the category of stress on the first syllable and the full vowel retained in the second syllable (Type 3). A few additional words which are commonly mis-stressed, such as breakfast, were also added.

Lesson one (see worksheets and description of teaching procedure in Appendix A). The teaching focus was on the introduction of the concept of the syllable in English and comparing this with the concept in other languages, followed by introducing the concept of word stress in English, then in other languages.

The teacher began with the introduction of a friend called Graeme and wrote the name on the board. They then asked how many parts the students heard. Answers ranged from one to five. The next step was to explain that Graeme is Kereama in Maori and go through the same procedure leading to the conclusion that Maori heard four parts in Graeme and that syllables are different in different languages. The third step was to focus on students' names and talk about the differences in perception of the number of syllables confirming that the idea of the syllable is different in different languages, that there are different ways of hearing syllables. Students recorded their answers and did further practice in identifying syllables (completed for homework) on Week 1 Worksheet 1 (See Appendix A).

The second half of the lesson introduced the concept of stress, returning to the name Graeme and asking for the difference between the two parts. Answers included descriptions such as stronger, longer, pressure, emphasis, tone is going up/down, accent and stress. Students' attention was also drawn to how the role of destressing and changing the vowel in the unstressed syllable. They used words such as smaller, quieter, and shorter. These were of course all words that could be used later in instruction and feedback. English was then compared with Maori stress and then with other languages. The teacher used circles on the board (as shown in Week 1 worksheet 2) and asked learners to work out stress patterns for their names and the names of others in the class, exploring different perceptions of stress amongst people with different L1s.

Lesson Two. This lesson began with Critical Listening, based on the pre-test (See Appendix A, Week 2, worksheet A). As a single group, students listened to the examples of students' speech, wrote down what they heard, and then discussed the different ways in which they heard the words. They then ordered the words according to the stress pattern. This was followed by students recording words, first Type One (stress on first syllable) and then Type Two (stress on second syllable), on their own and in sentences, and listening back to self-evaluate. (See Week 2. Worksheet B and Worksheet C). These formed the basis of the Critical Listening practice for week 3.

Lesson Three. The focus of the teaching in week three was on Critical Listening based on the participants' recordings. (See Week 3. Worksheet A and Worksheet B)

Lesson Four. The focus of this lesson was revision and further practice, using a card game that required accurate perception and production to achieve communication. (See Week 4. Worksheet A – with comments).

Teaching Received by the Comparison Group.

These activities from their course book, Unlock: Listening and speaking skills 3 (Ostrowska, 2014), were used by the class teacher while the experimental group was with me.
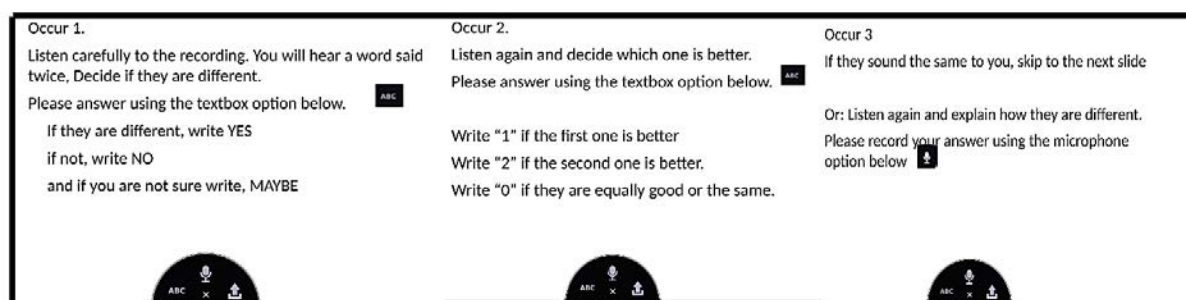
   1. Unit 1, p. 29: focus on "signposting phrases" being pronounced as a complete phrase and underling the stressed syllable in each phrase. Also on p. 72 on stress changes according to word class.

   2. Unit 2, p. 37: focus on pronunciation of final consonants. Also pronunciation of regular past tense endings on p. 63.

   3. Unit 3, p. 55: Weak forms and connected speech.

   4. Unit 4, p. 77: "Underline the syllable which has the most stress in each phase".

In the second half of the semester, while students in the comparison group attended the concept-focused instruction the others from the experimental group also covered the exercises below. In talking to the class teacher after the semester, she did comment that when she gave the experimental group this instruction, they seemed to understand it and carry out the activities much more easily.

Data Collection and Analysis

The data was collected in the computer lab using Voice Thread (n.d.). The items used in the perception test were: occur, obtain, credit, constrain, constant, commit, breakfast, percent, emerged, wages, licence, sector. The original intention was to only include 6 of the 12 perception test items in the teaching (occur, constrain, constant, commit, breakfast, emerged). Unfortunately, as the teaching progressed and minor tweaks were made, two of the other test items were inadvertently added (obtain and percent).

For each of the 12 words, chosen from difficulties that students had experienced two sound files were created from student recordings, one where the stress was incorrect - either in the wrong place or on both syllables, and one where it was correct. There were three slides for each word pairing (Voice Thread, (n.d.) works on slides and moves from one to the next. One can select record or type in a text box). So, the participant would see, for example, these slides for the word "occur", shown in Figure 1.

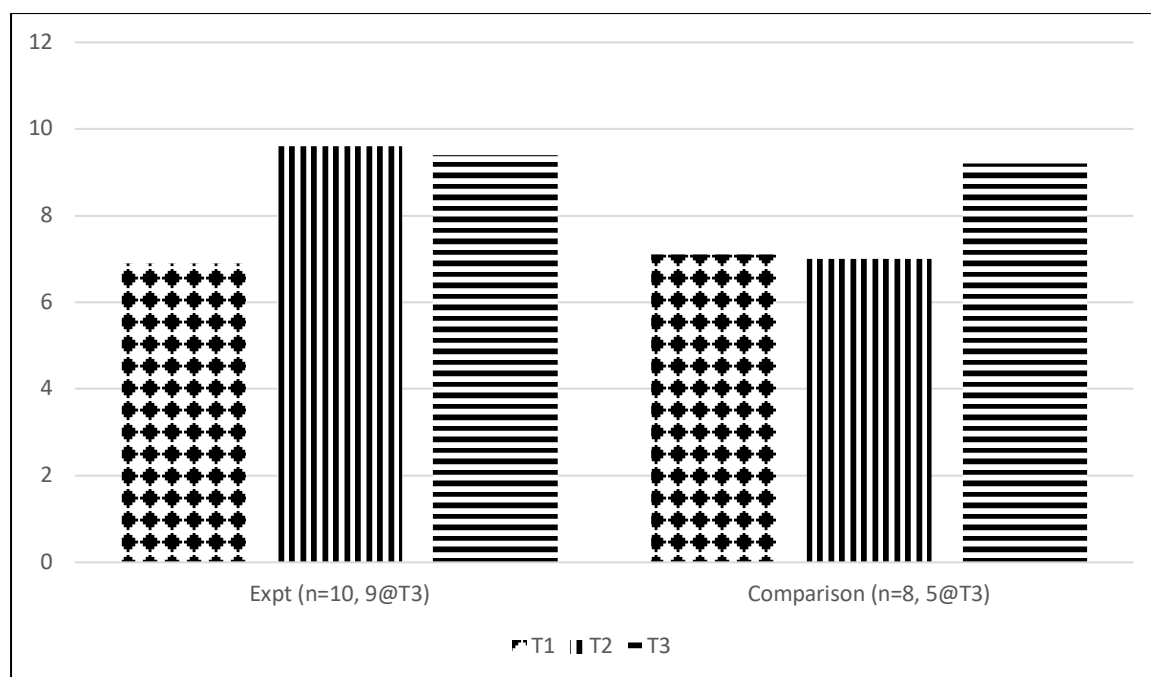**Figure 1. Screenshots of Voice Thread (n.d.) slides.**

The first slide attempts to determine whether or not the participants can hear a difference. This was coded as Yes, No, or Maybe. In the analysis, No and Maybe were conflated into one category. The second slide attempts to establish if participants can identify the most accurate use of word stress. If they said both were the same, this was conflated with choosing the wrong option. The third slide asked them to try and explain the difference orally. The data from the third slide becomes relevant for those who can hear a difference and can accurately choose the most accurate version. It provides a more qualitative description of their understanding of the concept of word stress and how this changed and will be the subject of a follow-up report. Therefore, the main focus here is on whether or not participants could both hear a difference and accurately choose the most accurate version.

The perception results from the first two slides described in the previous paragraph and shown in figure 1, were transferred to an SPSS data file. A colleague double-checked that this was done accurately. This led to three columns for each time: hears diff, correct, hears diff + correct. The various statistical procedures used are described in the results section.

Results

As has already been argued, this study aims to move beyond discrimination and identification tests and focus more on the degree to which learners understand the target concept. This report is limited to whether learners hear the difference, and where they do, if they identify the most accurate form. Therefore, the analysis focuses on these results.

The results across a number of statistical tests show large and significant gains in perception for the experimental group and that these gains were largely maintained over time (See Appendix B for more detailed reporting of the statistical method and results). The comparison group made no gains over time, but those who received the instruction following their repeat test made gains similar to the experimental group. Firstly, the Mean scores for being able to both hear the difference and choose correctly are depicted in Figure 2. These scores are out of 12 and show the ten learners in the experimental group improved from a mean of 6.9 at time one to 9.6 at time two, following instruction. Nine of these learners re-sat the test eight weeks later, achieving a mean score of 9.4. By way of contrast, the eight learners in the comparison group, which had their normal lessons with some pronunciation exercises, made no gains over the eight weeks between the test at time one (M = 7.1) and at time two (M = 7.0). This group was given the opportunity to attend the instruction, with the five who did so scoring a mean of 9.2 on the test post-instruction.

Note. This shows the mean score out of 12, "hear the difference and choose correctly".

Figure 2. Perception test results for both groups at three times.

Descriptive statistics, presented in Table 3, indicate large gains from concept-focussed teaching. The lack of any overlap between the confidence intervals for T1 and T2 for the experimental group gives strong support to the claim that this is a clear difference in results. Both parametric and non-parametric tests were also run, all providing evidence that the gains made following instruction were both large and significant. Table 4 shows the individual results for participants. Here it is clear that they all made at least some gains following instruction. This is reflected in a Wilcoxon signed rank test confirming the significance of the findings, with 10 positive differences, $W = 55.0$, and $p = .005$. A paired samples *t*-test supports these findings with $t = -5.449$. $SD = 1.6$. df 9, $p = .000$ for the experimental group between time one and time two. In addition, a Bayesian factor analysis strongly supports the conclusion that the instruction was effective with a B_{10} value of 86.91 and a large median effect size of -1.485.

Table 3. Descriptive statistics for "hear the difference and provide correct answer".

Group and Time	Mean/12	Confidence Interval	Standard Deviation
Expt Group T1 (n=10)	6.9	[5.5-8.2]	2.5
Expt Group T2 (n=10)	9.6	[9.0-10.2]	1.3
Expt Group T3 (n=9)	9.4	[8.2-10.7]	1.8
Comp Group T1 (n=8)	7.1	[5.0-9.0]	3.3
Comp Group T2 (n=8)	7.0	[5.4-8.6]	2.8
Comp Group T3 (n=5)	Post teaching so now combined with Expt Group T2		
Exp+C n=15) pre-teach	6.4	[5.3-7.5]	2.4
Exp+C (n=15) post-teach	9.5	[8.6-10.3]	1.8

Table 4. Individual results for Expt and Comp groups, perception (hear diff + correct) across three times.

Expt	AE	BE	CE	DE	EE	FE	GE	HE	IE	JE
T1	7	6	2	5	9	9	6	11	6	8
T2	9	10	8	9	11	10	8	12	9	10
T3	9	10	7	10	11	11	7	12	8	--

Comp	AC	BC	CC	DC	EC	FC	GC	HC
T1	4	10	10	9	1	7	10	6
T2	4	10	10	9	4	6	9	4
T3	10	--	--	--	12	8	11	5

Further tests were also run on the results for identifying the correct version, regardless of whether they could hear a difference. Here the experimental group improved from a mean of 7.7 at time one to 9.7 at time two. A paired samples *t*-test finds this significant (*t* = - 4.243) at *p* = 0.002, with a large effect size of *d* = 2. A Wilcoxon signed rank test also finds a significant effect for instruction (W = 45, *p* = .007, and 9 positive differences, 1 tie). In summary, even though the starting point was higher (at 7.7/12) when excluding participants' ability to hear a difference, evidence of clear progress was still found.

Finally, there is the consideration as to whether or not this learning can be extended to other contexts. As already noted, it was intended to only teach half of the tested items but in the event eight were taught, making an analysis of an impact difficult to quantify. However, a comparison of immediate pre-test with immediate post-test scores on non-taught items shows an average of 56.9% (57.7% overall) compared with 71.7% (79.2% overall). On the surface, it seems that some learning may have extended to other contexts, but it is not possible to say how much.

Discussion

This section considers the findings in relation to the research questions before addressing some of the limitations and implications for teaching and further research.

RQ 1. How does a conceptual approach affect participants' perceptions of word stress?

The study found a significant and large effect for the approach used, providing further evidence to support the inclusion of an explicit focus on perception when teaching pronunciation. It also found that the majority of these gains were retained over time. In addition to showing that the findings for syllable codas (Couper, 2011) could be extended to word stress, it also supports Lee and Lyster's (2016) findings that teaching can improve perception. This complements other findings that HVPT can help learners in a self-study situation (Thomson, 2018) and answers the question as to what the teacher can do to help. It also extends the teaching of perception beyond the common phoneme target (Thomson, 2018) to aspects of pronunciation that have so far received little research attention. The majority of studies have tended to focus on phonemes, with just a few looking at other aspects of pronunciation such as syllable codas (Couper, 2011; Huensch & Tremblay, 2015), tone (Sereno & Wang, 2007) and prosody (Aoyama & Guion, 2007). This was a relatively short intervention focusing on word stress in two-syllable words, but it found that by helping learners to understand the underlying concept of word stress they

were able to improve their perception and largely retain those improvements. By focusing on the underlying concept, it is expected that the chances of long-term retention are increased.

RQ2. What sorts of approaches to the measurement of perception does this teaching lend itself to?

The approach to measuring changes in perception is a key consideration in this study. Unlike most studies that have used identification or discrimination tasks, this study started from the standpoint that perception is not an all or nothing issue. In using identification tasks, the student's response is either right or wrong. However, as we form phonological concepts in the new language we start with a very approximate idea and slowly sharpen up the boundaries between those concepts. Rather than simply saying which one is better or identifying the odd one out, the aim is to find a measure that better reflects the status of the learner's concepts. Thomson (2012) did ask participants how confident they were with their answers, which reveals useful additional information. Couper (2011) was dissatisfied with what was revealed in the discrimination tasks he had previously used and walked his participants through exactly how they interpreted what they were hearing, with regards to syllable codas. This enabled the development of a scale to describe the extent to which the participants understood the concept of English syllables (no, maybe, probably, yes). Here, in the case of word stress, an attempt was made to adapt this to computer-based data collection, to allow for greater efficiencies. The levels of conceptual understanding identified here were: 0) cannot hear the difference (very limited understanding), 1) can hear the difference but cannot identify the best choice, 2) can hear the difference and can identify the best choice, 3) as for 2 but can identify part of the reason for the difference, although not the salient one, and 4) as for 2, but can also identify the salient difference (full understanding). Results to level 2 have been presented here, namely, whether or not they can hear the difference and if they can identify the best choice. Results for the distinction between levels 2, 3, and 4 require a more qualitative treatment for which there is no space here. Nevertheless, even taking into account levels 0-2 is a move away from the standard binary assessment and helps to shed light on concept formation. This is seen as particularly important as concept formation is central to the aim of the teaching approach being promoted.

RQ 3. How well can a conceptual approach be applied to the teaching of word stress?

The teaching here is based on the notion that learners often need more than exposure to conceptualise sounds in the TL and that they can be helped through drawing on their cognitive skills. These might include the ability to notice, and compare and contrast (Mompean, 2014). In this teaching, I attempted to understand how the participants heard word stress and used this as a basis for explanations and feedback, leading to the development of socially constructed metalanguage (SCM) (Couper, 2011). They used words such as stronger, longer, louder, pressure, emphasis, tone is going up/down, accent and stress. In talking about the unstressed syllable, they used words such as smaller and shorter. It is argued that the success of such communication is pivotal in effective teaching. The other related factor is the use of Critical Listening practice which allows learners to explore and understand the boundaries between different phonological categories (Fraser, 2001, 2006a). Through this teaching, it was quite easy to apply the conceptual approach developed in relation to syllable codas to the teaching of word stress.

These results detail the importance of focusing on concept formation, through the use of Critical Listening and Socially Constructed Metalanguage, in helping learners to improve

pronunciation. While the argument may continue as to the relationship between perception and production, this approach focuses instead on the role of understanding the concept behind phonological features such as word stress. It was seen that learners need to understand the concept of both the syllable and word stress in order to be able to adequately perceive and produce it. At one level, it may be noted that with English one really needs to learn the stress pattern along with each word, unlike in other languages where patterns are more predictable. However, it is argued that there is a more important underlying issue which is the conceptual nature of word stress, that it involves much more than just knowing where to place the stress. This means that teachers need to make sure they focus not just on placement but on all the other aspects of stress, in particular, what is happening to the unstressed syllables.

Limitations

Although the teaching did not focus specifically on the tested items, eight of the 12 items were included in Critical Listening work, which was not initially planned. The original intention was to include only six in the teaching but as the items for Critical Listening practice were collected from student recordings throughout the study, I took two additional words from the perception test without noticing until after data collection. In the future, it would be better to evenly split the number of test items taught. One might also consider increasing the number of test items, but of course that is always a delicate balancing act, especially when working within the constraints of a classroom context.

Further data was also collected regarding responses to the open-ended question asking participants to explain why one version was better than the other, which was designed to distinguish between levels 2, 3, and 4 as described above. This has not been presented here, as there were a number of technical difficulties that arose as a result of doing this via computer rather than face-to-face. While there appeared to be more understanding expressed after instruction, in the delayed post-test several participants did not provide full explanations. The presentation of this analysis requires a more detailed qualitative approach that there is no room for in the space of this article.

Implications

There are a number of implications for teachers, and researchers who may attempt to replicate this study. Teachers should incorporate a focus on both perception and production in their teaching. The best way to do this is by focusing on development of underlying concepts. To be able to teach word stress, learners need to understand the concept of the syllable, so consider ways first of all to increase awareness of the differences in this concept in English and their L1s, before moving on to a focus on word stress. Ideally, a diagnostic test would be helpful in setting priorities for learners and in helping them to become more aware of areas of difficulty. The recordings collected during diagnostic testing can be used to create Critical Listening activities. These activities can be used to further raise awareness and to help teachers understand how their students hear the TL. Students can listen multiple times on their own to try and hear these differences, but they also benefit from teacher guidance in helping them to interpret these concepts. The teacher should bring the focus to the reduction and vowel change occurring in the unstressed syllable so that when they produce these words, they avoid the common mistake of stressing every syllable. Get students to make further recordings so that they can check their own and each other's production and get appropriate feedback as they continue to adjust their conceptual understanding. Remember that this also needs to be meaningful and needs to lead to freer practice. In this set of lessons, we finished with a card

game, but with a little more time, further free practice such as a role play would be recommended. Most importantly, teachers need to start with learners' understandings and get them to explain how they perceive speech. Teacher explanations risk being confusing if they are couched in TL terms rather than in terms the students understand. Researchers would ideally also be the teachers, or at the very least they need to work cooperatively to ensure teachers understand the rationale behind the approach being taken so that they can adjust their teaching in the real-life interactions of the classroom.

Conclusion

This study has provided a detailed description of a particular type of instruction, explained further through the notes and worksheets in Appendix A (also available on IRIS: https://www.iris-database.org/iris/app/home/detail?id=york:940132). This report has focused on the impact on perception of word stress because there is a clear need for more empirical evidence and practical explanation of how teachers can help with perception. While teachers use minimal pairs and discrimination type activities these are limited unless teachers also work with students to increase awareness of the underlying concepts behind this perception. The approach presented here includes work on production, but this report has focused on perception because it tends to be neglected, and there is little guidance for teachers on how to teach it. The argument is that pronunciation requires both perception and production, and that the best way to develop both is to foster the formation of L2 phonological concepts. I would like to encourage other classroom researchers to consider replicating this study. To this end, I will work on further guideline instructions and would encourage anyone who is interested to get in touch. There will be a need to adjust the individual worksheets according to local conditions, and of course, if you have more time, you could extend the focus beyond two-syllable words. The data for this study was collected just before lockdown for Covid-19, and ongoing disruptions have hampered the collection of further data.

About the Author

Graeme Couper is an Associate Professor in the School of Language and Culture at Auckland University of Technology. He has many years' teaching experience in a wide range of countries and contexts that he brings to bear in his research into L2 pronunciation teaching and learning. ORCID ID: 0000-0003-4561-0446

Acknowledgements

To cite this article

References

Aoyama, K. & Guion, S. (2007). Prosody in second language acquisition. In O.-S. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (281-297). John Benjamins.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly, 48*, 136–163. https://doi.org/10.1002/tesq.99

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4-63). Blackwell.

Burri, M., & Baker, A. (2021). I feel...slightly out of touch: A longitudinal study of teachers learning to teach English pronunciation over a six-year period. *Applied Linguistics, [early view],* 1-19. https://doi.org/10.1093/applin/amab009

Buss, L. (2016). Beliefs and practices of Brazilian EFL teachers regarding pronunciation. *Language Teaching Research, 20*(5), 619-637. https://doi.org/10.1177/1362168815574145

Cauldwell, R. (2013). Phonology for listening: Teaching the stream of speech. Speech in Action.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*(1), 46-66.

Couper, G. (2011). What makes pronunciation teaching work? Testing for the effect of two variables: Socially Constructed Metalanguage and Critical Listening. *Language Awareness*, *20*(3), 159-182. https://doi.org/10.1080/09658416.2011.570347

Couper, G. (2012). Teaching word stress: Learning from learners' perceptions. *TESOL in Context Special Edition S3: November 2012.* Downloaded from http://www.tesol.org.au/Publications/Special-Editions

Couper, G. (2013). Talking about pronunciation: Socially constructing metalanguage. *English Australia, 29*(1), 3-18.

Couper, G. (2016). Teacher cognition of pronunciation teaching amongst English language teachers in Uruguay. *Journal of Second Language Pronunciation*, *2*(1), 29-55. https://doi.org/10.1075/jslp.2.2.02Cou

Couper, G. (2021). Teacher cognition of pronunciation teaching: The techniques teachers use and why. *Journal of Second Language Pronunciation*, *7*(2), 212-239 https://doi.org/10.1075/jslp.20004.cou

Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reid & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106-124). Wiley Blackwell.

Deterding, D. (2013). Misunderstandings in English as a lingua franca: An analysis of ELF interactions in South-East Asia. De Gruyter. https://doi.org/10.1515/9783110288599

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399–423. https://doi.org/10.2307/3588487

Field, J. (2008). *Listening in the language classroom.* Cambridge University Press.

https://doi.org/10.1017/CBO9780511575945

Fraser, H. (2000). Coordinating improvements in pronunciation teaching for adult *learners of English as a second language.* Australian National Training Authority (ANTA Innovative Project).

Fraser, H. (2001). *Teaching pronunciation: A handbook for teachers and trainers.* Sydney, Australia: TAFE NSW Access Division. Retrieved February 11, 2015, from http://helenfraser.com.au/publications/

Fraser, H. (2006). Helping teachers help students with pronunciation: a cognitive approach. *Prospect, 21*(1), 80-96.

Fraser, H. (2010). Cognitive theory as a tool for teaching second language pronunciation. In S. de Knop, F. Boers & T. de Rycker (Eds.), *Fostering language teaching efficiency through cognitive linguistics* (pp. 357-379). Mouton de Gruyter.

Gibbons, P. (2006). Bridging discourses in the ESL classroom. Continuum.

Gomez Lacabex, E., & Gallardo del Puerto, F. (2014). Two phonetic-training procedures for young learners: Investigating instructional effects on perceptual awareness. *The Canadian Modern Language Review, 70*(4), 500-531. https://doi.org/10.3138/cmlr.2324

Gussenhoven, C. (2005). *Understanding phonology*. Hodder Arnold.

Huensch, A. (2016). Perceptual phonetic training improves production in larger discourse contexts. *Journal of Second Language Pronunciation, 2*(2), 183-207. https://doi.org/10.1075/jslp.2.2.03hue

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics, 52*, 105-120. https://doi.org/10.1016/j.wocn.2015.06.007

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving, 7*, 2-9. https://doi.org/10.7771/1932-6246.1167

Jenkins, J. (2000). The phonology of English as an international language: New models, new norms, new goals. Oxford University Press.

Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-stress patterns through auditory priming tasks. Studies in Second Language Acquisition, 39, 319-346. https://doi.org/10.1017/S0272263117000031

Ladefoged, P., & Disner, S. F. (2012). Vowels and consonants: An introduction to the sounds of language (3rd ed.). Wiley-Blackwell.

Larson-Hall, J. (2015). A guide to doing statistics in second language research using SPSS and R. Taylor Francis.

Lee, A., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition, 38*, 35-64. https://doi.org/10.1017/S0272263115000194

Lee, A., & Lyster, R. (2017). Can corrective feedback on second language speech perception errors affect production accuracy? *Applied Psycholinguistics, 38*(2), 371-393. https://doi.org/10.1017/S0142716416000254

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*, 345–366. https://doi.org/10.1093/applin/amu040

Lewis, C., & Deterding, D. (2018). Word stress and pronunciation teaching in English as a Lingua Franca contexts. *The Catesol Journal, 30*(1), 161-176.

Lively, S., Pisoni, D., & Goldinger, S. (1994). Spoken word recognition: Research and theory. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 265-301). Academic Press.

Mompean, J. (2014). Cognitive Linguistics and phonology. In J. Littlemore & J. Taylor (Eds.), *Bloomsbury companion to Cognitive Linguistics* (pp. 253-276). Bloomsbury.

https://www.bloomsbury.com/us/bloomsbury-companion-to-cognitive-linguistics-9781474237321/

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73-97.

Murphy, G. (2002). *The big book of concepts.* MIT (Bradford)

Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning, 68*(4), 1032-1075. https://doi.org/10.1111/lang.12310

Ostrowska, S. (2014). Unlock Level 3: Listening and speaking skills. Cambridge University Press.

Pisoni, D., & Lively, S. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross language research* (pp. 434-459). York Press.

Port, R. F. (2007). The graphical basis of phones and phonemes. In O.-S. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 349–365). John Benjamins.

Sereno, J., & Wong, Y. (2007). Behavioural and cortical effects of learning a second language: The acquisition of tone. In O.-S. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 239-258). John Benjamins.

Swain, M. (2006). Languaging, agency and collaboration on advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). Continuum.

Thomson, R. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62*(4), 1231-1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation, 4*(2), 208-231. https://doi.org/10.1075/jslp.17038.tho

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*, 905-916.

Voice Thread. (n.d.) *Voice Thread*. https://voicethread.com

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Walley, A. (2007). Speech learning, lexical reorganization, and the development of word recognition by native and non-native English speakers. In O.-S. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 315-330). John Benjamins.

Williams, M., & Burden, R. (1997). Psychology for language teachers: A social constructivist approach. Cambridge University Press.

Appendix A. Teacher notes and worksheets

Overview of a study for replication.

There are two main aims for this project: "Cognitive Phonology in Teaching and Evaluation of Perception" C-PTEP.

1. to test the effectiveness of a particular method of teaching pronunciation, using the case of word stress (the method has been developed and tested by me and is based on the classroom application of Cognitive Phonology), and

2. to test the value of using a scale to describe and measure learners' perception and improvement in perception following instruction both immediately and over time (This measurement scale is novel and has been developed by me to provide a more finely grained measure of speech perception than is typically used by other researchers). I also have colleagues in France who are planning to replicate this study, and I would invite others to replicate it too.

The project involves a series of four forty-minute lessons and pre- post- and delayed testing of speech perception and production.

Items for teaching and testing:

I have taken these from the most frequently occurring members of the word families in the Academic Word List. Available at:

https://www.victoria.ac.nz/lals/resources/academicwordlist

I then ordered them according to stress pattern, as shown in table 1. Obviously the most common are: Stress on the first syllable with a schwa or /I/ on the second syllable (Type 1) and schwa or /I/ on the first syllable and stress on the second syllable (Type 2). I have collapsed the remaining variants under the category of stress on the first syllable and the full vowel retained in the second syllable (Type 3).

Table 1. Stress patterns (In NZE) on most frequent two-syllable words from AWL

Schwa or /I/ + stress			Stress + schwa or /I/			Stress +full	V=/ə/+Str N=Str+full
Achieve	Enhanced	Removed	Alter	Method	**N=V**	Adults	Abstract
Affect	Ensure	Required	Ceases	Neutral	Author	Annual	Compounds
Assigned	Estate	Response	Chapter	Normal	Challenge	Aspects	Conduct
Assume	Exceed	Restore	Cited	Notion	Channel	Complex	Conflict
Attached	Ignored	Restraints	Civil	Option	Credit	Concept	Contract
Attained	Implies	Retained	Constant	Panel	Cycle	Context	Contrast
Aware	Imposed	Revealed	Crucial	Passive	Factors	Decades	Export
Behalf	Induced	Select	Data	Portion	Features	Expert	Extract
Commenced	Inferred	Survive	Error	Prior	Focus	Finite	Impact
Commit	Invoked	**N=V**	Ethnic	Published	Functions	Framework	Insert
Compiled	Involved	Approach	Final	Random	Label	Income	Project
Conceived	Mature	Collapse	Founded	Region	Labour	Insights	Survey
Confined	Obtained	Consent	Gender	Rigid	Layer	Manual	**Stress+full**
Constraints	Occur	Debate	Global	Section	Lecture	Likewise	**N=V**
Convinced	Perceived	Decline	Granted	Sector	Licence	Media	Access
Denote	Percent	Design	Image	Status	Major	Nuclear	Comments
Deny	Precise	Display	Instance	Tension	Purchase	Offset	Format
Derived	Promote	Release	Items	Theory	Sequence	Outcomes	Index
Despite	Pursue	Reverse	Journal	Thesis	Target	Series	Input
Device	Refine		Legal	Topic	Trigger	Solely	Issues
Domain	Relaxed		Logic	Version		Somewhat	Levy
Emerged			Mental	Vision		Visual	Network
						Volume	Output
						Welfare	Process
						Widespread	Prospect
							Schedule

There are a few additional words from an earlier exploratory study that have been added as I have already established that they caused problems and have recordings.

Tests: Pre- Post- and Delayed-Post

Pre-test 1: reading words alone and in sentences (10 Type 1, 10 type 2), and extemporaneous speech (based on a picture story)

Words plus sentences:

1. **Breakfast**: I ate fruit for **breakfast**.
2. **Channel:** It is good to **channel** your efforts into one project.
3. **Collect:** I **collect** the mail every morning.
4. **Commit:** You should **commit** yourself to your studies.
5. **Constant:** Inflation is a **constant** threat to the economy.
6. **Constrain**: A lack of money will **constrain** what I spend.
7. **Credit**: Give the lecturer **credit** for all the work he does.
8. **Emerged:** He **emerged** from the cave.
9. **Enhanced:** She **enhanced** her chances of winning the race.
10. **Functions**: I cannot use many of the **functions** on my phone.
11. **Labe**l: The writing was so small he could not read the **label**.
12. **Licence**: Have you got a **licence**?
13. **Obtain**: It is difficult to **obtain** those things.
14. **Occur**: It did not **occur** to him.
15. **Percent**: The bank offered three **percent** interest.
16. **Precise**: Please be **precise** in your measurements.
17. **Removed:** The name was **removed** from the list.
18. **Sector:** The tertiary **sector** needs more money.
19. **Valid:** Your complaint is **valid**.
20. **Wages:** Teachers need higher **wages**.

There were also errors amongst other words in the carrying sentences. These will be noted separately and checked again in the post-test and delayed post-test. The words were: efforts, project, threat, economy, lecturer, chances, interest, measurements, tertiary, complaint.

Picture story: Students have to tell a story based on a picture. I will evaluate it by counting the number of syllable and word stress mistakes. Will need to measure this against the total number of multi-syllable words used. Keep separate measurements for 2-syllable words as these are the focus of the current instruction. Reflection: This part of the test was not very effective at gathering data in a computer-based setting so would need to be collected face-to-face or reframed.

Pre-test 2: speech perception. Twelve of the most challenging words from Pre-test 1 are chosen, sound tracks of the words are cut and pasted into separate files so that there is one example with accurate word stress and number of syllables (number of syllables on the basis

that syllables underlie the production of stress) and one example where the stress is not so good. In two cases these difficulties revolve around syllables as well (emerged and wages).

The items were (the number in brackets refers to the speech sample that was best, either 1 or 2): occur (1), obtain (1), credit (1), constrain (2), constant (2), commit (2), breakfast (2), percent (1), emerged (2), wages (1), licence (2), sector (2).

Outline of teaching with worksheets.

Note: This was originally designed with eight twenty-minute weekly sessions, but I conflated it into four forty-minute weekly sessions. Classes held in the computer lab.

Week 1 – Part A: 20 minutes. Introduction to concept of syllable in English and comparison with other languages.

Syllables: Start with introduction, this is my friend Graeme (with picture) T writes name on board. How many parts do you hear in his name "Graeme"? Elicits 2/3/4. His name in Maori is Kereama. How many parts? 4. Tells me that Maori heard 4 parts in Graeme. So it's different in different languages. What about your names? Going around the class we see how different speakers perceive the number of syllables in each other's names. Conclusion that the idea of the syllable is different in different languages, that there are different ways of hearing syllables.

Practice: Listen to these words (see worksheet). How many syllables do you hear? (Check the meanings of the words for homework).

Week 1 – Part B: 20 minutes. Introduction to concept of word stress in English and comparison with other languages.

Stress: Look at the name again: Gra / eme. What's the difference between the 2 parts? Elicit descriptions longer stronger etc. Compare English with Maori stress and then with other languages. Use circles on board (as shown in worksheet 2). Learners work out stress patterns for their names and the names of others in the class, exploring different perceptions of stress amongst people with different L1s: Group discussion (as indicated in worksheet 2).

Focus on how stress is formed, especially through destressing and changing the vowel in the unstressed syllable (At this stage examples will be limited to this pattern).

Week 1 Worksheet 1

A. In English, we would think of "Graeme" as having two parts, or syllables. How many syllables would it have in your language or other languages you know?

1. English: Graeme — two syllables

2. Maori: Kereama — four syllables

3. Other Language .... Number of syllables ....

4. Other Language .... Number of syllables ....

5. Other Language .... Number of syllables ....

6. Other Language .... Number of syllables ....

B. Look at these words from the academic wordlist. Write down the number of syllables. The first one is done as an example:

1. approach: Two syllables a — pproach Meaning: (N) A way of doing things. (V) to come closer.

2. analysis: _____ Meaning: detailed study or examination of sth to understand it more.

3. definition: _____ Meaning: describing the meaning of sth.

4. established: _____ Meaning: respected and existing for a long time

5. involved: _____ Meaning: taking part in something.

6. issues: _____ Meaning: problems or difficulties

7. required: _____ Meaning: necessary

8. structure: _____ Meaning: the way on which the parts are connected to form a whole

9. categories: _____ Meaning: groups of similar things

10. constraints: _____ Meaning: limitations or restrictions

11. scheme: _____ Meaning: a plan or a system for organising sth.

12. mechanisms: _____ Meaning: a set of moving parts that perform a task

13: challenge: _____ Meaning: sth that requires work, focus or skills

14: acknowledged: _____ Meaning: recognised

- English

Gra eme

- Maori

Ke re a ma

Now, think about how you form stress in your first language. Use different sized circles to show the syllables and stress for your name, or if your name is a very short one, choose another name from your language which is longer. Now do the same for at least two other languages.

First Language:………………………………………..

(Draw circles here):

Now talk to people with different first languages. Ask how your name sounds to them and ask them to draw circles. Compare with the circles you have drawn. Then draw circles for their language below, and compare them with their circles.

Other languages: 1)

(Draw circles here):

Other languages: 2)

(Draw circles here):

Discuss how stress is different in different languages. Make notes.

Week 2 - Critical Listening 1; stress, exploring difference:

[In preparation, prepare examples from pre-test 1 of words that caused difficulties. These will include the ones used in pre-test 2, but if possible, add others as we want to include half of the words used in pre-test 2 in the teaching materials].

Whole class, get students to listen to the examples, write down what they hear, and then discuss the different ways in which they heard the words (see Week 2 Worksheet A). (If possible, use students' own recordings from pre-test). This is followed up with participants recording examples of Type 1 and Type 2 patterns on Voice Thread (See Week 2 Worksheets B and C). I got them to make both lots of recordings so I could prepare some critical listening practice based on them for the following week

Week 2. Worksheet A: Listening for stress.

Listen to these words. You will hear each word twice. Write down if the word sounds the same or different each time. If it is different, describe the difference. Does one sound better? Listen and make notes.

1. occur: ....
2. commit: ....
3. constrain: ....
4. constant: ....
5. corporate: ....
6. emerged: ....
7. breakfast: ....
8. precise: ....

Discuss how the unstressed syllable sounds, and differences in different languages.

So there are two things: where the stress is and how it is produced.

Put the words above into columns below according to where the stress is.

Words with stress on the first syllable	Words with stress on the second syllable

Now make a note of whether the words are: nouns, verbs, adjectives, (or more than one).

 Do you think there might be a pattern here?

Now work with a partner and practise saying the words. Give your partner feedback. Get the teacher to help.

Week 2: Worksheet B: Practice in context: Type One; stress on first syllable

Listen to the words and sentences. Now record the word on Voice thread, followed by the sentence. Listen again and decide if you have pronounced it correctly. Did you remember to reduce the second syllable? Make notes of any difficulties you had.

1. ceases: If something ceases, it stops.

   notes...

2. chapter: Read the second chapter in the book.

    notes...

3. cited: I cited one hundred sources in my essay.

   notes...

4. civil: In a disaster, civil defence tries to help.

   notes...

5. constant: I am sick of his constant complaints.

   notes...

6. crucial: If something is crucial it is important and essential.

   notes...

7. error: I made an error when I said that.

   notes...

8. ethnic: There are many ethnic groups in Auckland.

   notes...

When you have finished, get your classmates to check.

The teacher will check too.

Week 2 Worksheet C: Practice in context: Type Two: Stress on second syllable.

Listen to the words and sentences. Now record the word on Voice thread, followed by the sentence. Listen again and decide if you have pronounced it correctly. Did you remember to reduce the first syllable? Make notes of any difficulties you had.

1. achieve: Try to achieve the best grade you can.

    notes………

2. affect: The weather is going to affect the game.

    notes………

3. assigned: When the tasks were assigned, I was given this job.

    notes………

4. assume: I assume you know the answers.

    notes………

5. attached: The file is attached to this email.

    notes………

6. attained: Most of the students attained 'A' grades.

    notes………

7. aware: I am aware that you do not want to do your homework.

    notes………

8. behalf: When someone does something in your name, they do it on your behalf.

    notes………

9. commenced: He commenced his studies last year.

    notes………

10. commit: If you want to do well you must commit yourself completely.

    notes………

When you have finished, get your classmates to check.

The teacher will check too.

Week 3. Start with Listening Type One (Stress on first syllable). After group and class discussion of the differences, students then work on computers first listening on Voice Thread, then doing extra practice, recording themselves and comparing. They do this first at the level of the word, then the sentence (See Week 3 Worksheet A). They then go through the same process with Type Two (Stress on the second syllable). See Week 3 Worksheet B.

Week 3: Worksheet A , Type One Critical Listening

Listen to these words. You will hear each word two or three times. Write down if the word sounds the same or different each time. If it is different, describe the difference. Does one sound better? Listen and make notes.

1. breakfast: ………..
2. civil: ………
3. constant: ………
4. crucial: ……….
5. efforts: ………
6. error: ……….
7. ethnic: ………
8. sources and essay: ……………………………………………………………………………………………………

    ………

Listen again on Voice Thread, where you will also hear my own recording.

Go to: Blackboard – Pronunciation Practice – 1st syllable extra practice

Once you are happy you can hear the differences, and understand the importance of reducing the

vowel, you can record the words yourself, listen and compare your recording with mine.

Then go back to your recordings from week 2 and listen.

Go to: Blackboard – Pronunciation Practice – Stress on first syllable

I have also recorded a model for you so you can hear the whole sentence. You may practice by

recording your sentences again and comparing them with mine.

Week 3 Worksheet B. Critical Listening of Type Two: Stress on second syllable

Listen to these words. You will hear each word twice. Write down if the word sounds the same or different each time. If it is different, describe the difference. Does one sound better? Listen and make notes.

1. assigned: ....
2. assume: ....
3. attached: ....
4. aware: ....
5. commenced: ....
6. commit: ....
7. obtain: ....
8. occur: ....
9. percent: ....
10. precise: ....

Listen again on Voice Thread, where you will also hear my own recording.

Go to: Blackboard – Pronunciation Practice – 2nd syllable extra practice

Once you are happy you can hear the differences, and understand the importance of reducing the vowel, you can record the words yourself, listen and compare your recording with mine.

Then go back to your recordings from week 2 and listen.

Go to: Blackboard – Pronunciation Practice – Stress on second syllable AE2/3 2019

I have also recorded a model for you so you can hear the whole sentence. You may practice by recording your sentences again and comparing them with mine.

Week 4 – Part A: Revision and introduction of Type Three

Revision of two types of word stress pattern and introduction of the third, less common one where there is no reduction. Listening and placing words in correct column as in Week 4 Worksheet A. Ideally one would get students to record the three patterns, and if there were time do further Critical Listening based on them.

Week 4 Part B: Practice with card game

Although the real focus is on Types 1 and 2, Type 3 introduces a bit more of a challenge for the better students as well.

Overview

40 cards: each one with a word and a short sentence or phrase. The aim then is to get rid of all your cards. 3 suits: Type 1 (16 cards), Type 2 (16 cards), Type 3 (8 cards). See cards below.

Game instructions and rules:

Remember the different stress patterns:

Type One: Stress + reduced/very short vowel. e.g. constant

Type Two: Reduced/very short vowel + stress. e.g. commenced

Type Three: Stress + full vowel. e.g. access

To play this game you have to pronounce and hear these patterns correctly.

You have five cards. Each one has a word, and the word in a sentence on it.

The first one with no cards is the winner.

Rules

Groups of four or five players.

Player a: deal 5 cards to each player.

Player b, to the left of the dealer: read out what is on your card.

Player c, to their left: If you have a card with the same stress pattern, read out the word and sentence. (If you don't, pick up a card. That is the end of your turn).

Players b and c: Show each other your cards to check.

- The number in the top left-hand corner can be used to confirm the answer.

- If they are correct, the two cards are put to one side.

- If they are not correct, player c keeps both cards (or whoever used the wrong stress pattern)

Player c: Read out what is on one of your other cards.

Player d, to their left: If you have a card with the same stress pattern, read out the word and sentence…

And so on, until someone has no cards left.

Cards used in card game (reduced in size)

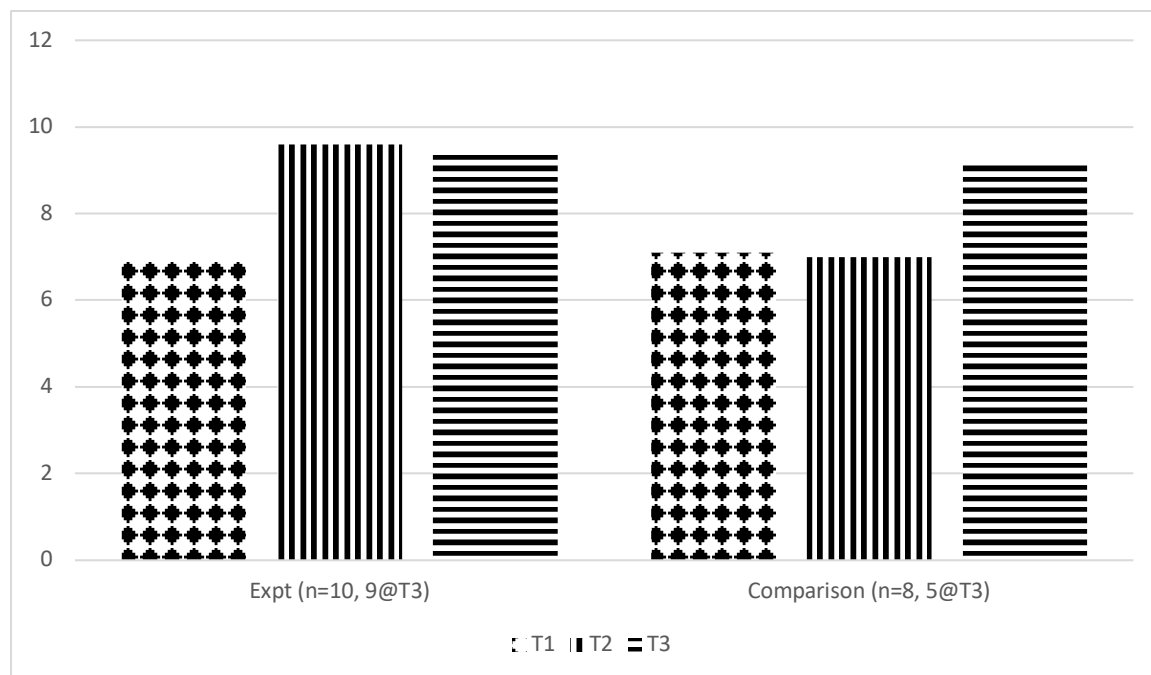|	
1. **Items** I bought 12 items from the shop.	**1.** **Constant** Some things are constant, others change.
1. **Lecture** I fell asleep during the lecture.	**1.** **Instance** You are often right, but in this instance you are wrong.
1. **Legal** It is not legal to drive so fast.	**1.** **Published** The book was published last year.
1. **Features** This computer has many new features.	**1.** **Passive** He did not complain, he was very passive.
1. **Granted** The bank granted him a loan.	**1.** **Image** I can still see the image in my mind.
1. **Region** It is an important wine-growing region.	**1.** **Normal** It is normal to like food.
1. **Major** Violence is a major problem.	**1.** **Sequence** Write the names in alphabetical sequence.
1. **Purchase** I could not afford to purchase it.	**1.** **Random** Choose a card at random.
2. **Conceived** I conceived the idea on the train.	**2.** **Convinced** She convinced me that she was right.
2. **Deny** I cannot deny that you are right.	**2.** **Domain** What is the domain name for your website?
2. **Exceed** Do not exceed the speed limit.	**2.** **Implies** He doesn't say it directly but he implies it.
2. **Obtained** Somehow, he obtained the money.	**2.** **Mature** Older students are sometimes called mature students.

2. **Promote** There are many ways to promote products.	2. **Pursue** Pursue your goals.
2. **Percent** The bank gave me 23 percent interest!	2. **Select** When choosing, be careful to select the right one.
2. **Collapse** Without support, it will collapse.	2. **Consent** Do you consent to being in this research?
2. **Survive** It was too cold to survive.	2. **Perceived** I perceived the difference between the two words.
3. **Comments** Write your comments in the book.	3. **Concept** It is a difficult concept to understand.
3. **Decades** It has been a problem for decades, maybe forty years.	3. **Finite** I have a finite amount of money.
3. **Outcomes** The outcomes were unexpected.	3. **Impact** The new tax had a large impact.
3. **Access** There is wheelchair access to the building.	3. **Format** Make sure it is in the correct format.

Appendix B. Details of statistical method and results

Following suggestions from Larson-Hall (2015), the following descriptive statistics are presented showing means (M), using BCa bootstrapping for 95% confidence intervals (CI) and standard deviations (SD) in an endeavour to make the statistics as robust as possible. See Figure 2 for an overview of results on perception tests, Table 3 for descriptive statistics, and Table 4 for individual results. Following this, paired samples *t*-tests were run, using 95% bootstrapping (Larson-Hall, 2015) to further increase the accuracy of the CI, which helps to overcome any failure to meet the conditions for parametric tests, such as the assumptions of normality and equal variances which may not be met here. Finally, a Bayesian paired t-test was run before considering the results of non-parametric tests.

Firstly, descriptive statistics for the ten participants in the experimental group were analysed for means, confidence intervals and standard deviations. Scores, out of 12, increased between time 1 (M = 6.9, 95% BCa bootstrap CI [5.5-8.2], SD = 2.5) and time 2, where there was also a reduction in the SD (M = 9.6, CI [9.0-10.2], SD = 1.3). Results for the nine participants who re-sat the test at time 3 (n = 9), eight weeks later, were compared with their results from T2 (n = 9), showing a small reduction in scores from time 2 (M = 9.6, CI [8.8-10.4], SD = 1.3) to time 3 (M = 9.4, CI [8.2-10.7], SD = 1.8). In percentage terms this represents an increase from 57.5% on the pre-test to 80% on the immediate post-test, dropping back slightly to 78% in the delayed post-test.

Between time 1 and time 2, the comparison group had their normal lessons including some pronunciation exercises. Their test results for time one (M = 7.1, 95% BCa bootstrap CI [5.0-9.0], SD = 3.3) and time two (M = 7.0, CI [5.4-8.6], SD = 2.8) showed little change.

Note. This shows the mean score out of 12, "hear the difference and choose correctly".

Figure 2. Perception test results for both groups at three times.

Appendix B. Details of statistical method and results

Table 3. Descriptive statistics for "hear the difference and provide correct answer".

Group and Time	Mean/12	Confidence Interval*	Standard Deviation
Expt Group T1 (n=10)	6.9	[5.5-8.2]	2.5
Expt Group T2 (n=10)	9.6	[9.0-10.2]	1.3
Expt Group T3 (n=9)	9.4	[8.2-10.7]	1.8
Comp Group T1 (n=8)	7.1	[5.0-9.0]	3.3
Comp Group T2 (n=8)	7.0	[5.4-8.6]	2.8
Comp Group T3 (n=5)	Post teaching so now combined with Expt Group T2		
Exp+C n=15) pre-teach	6.4	[5.3-7.5]	2.4
Exp+C (n=15) post-teach	9.5	[8.6-10.3]	1.8

*95% BCa bootstrapping.

The descriptive statistics alone already indicate large gains from concept-focussed teaching. The lack of any overlap between the confidence intervals for T1 and T2 for the experimental group gives strong support to the claim that this is a clear difference in results.

A paired samples *t*-test for the experimental group between time 1 and time 2 finds $t = -5.449$. SD = 1.6. df 9, $p = .000$. Using bootstrapping to analyse the mean difference (Larson-Freeman, 2015) also finds a statistical result: Mean = -2.7, Std error = .495, CI [-3.7-1.9], $p = .006$.

There is some variation in the way in which measures of effect size are calculated. I have followed Larson-Freeman's suggestion to use Volker's (2006 as cited in Larson-Freeman 2015, p. 148) *d* measure of effect size for paired samples (2015, p. 148) using *t*-tests. The effect size is large at $d = 1.74$. Here is the calculation: Mean at time 1 (6.9) minus Mean at time 2, (9.6), divided by the square root of 2 times the difference between the standard deviations at times 1 (2.5) and 2 (1.3).

$$d = \frac{6.9 - 9.6}{\sqrt{2}\,(2.5 - 1.3)} = 1.74$$

A more recently proposed alternative to the above-described classical frequentist analysis is a Bayes Factor analysis (Norouzian et al., 2018). This analysis has the advantage of comparing how likely the null hypothesis is with the likelihood of the alternative analysis (Jarosz & Wiley, 2014). The scores for T1 and T2 have been analysed through JASP (jasp-stats.org) and reveal a Bayes Factor of $B_{10} = 86.91$, which suggests the alternative hypothesis is 86.91 times more likely than the null hypothesis, a very strong finding in favour of the alternative (Jarosz & Wiley, 2014). Bayesian statistics also provide a nuanced measure of the effect size (Norouzian et al., 2018). In this case, a large median effect size, 95% CI [-2.546, -0.528], of -1.485 is found.

There seems to be little doubt that the teaching intervention had a large and significant effect, but just to be sure, a non-parametric related samples Wilcoxon signed rank test was also carried out. There were 10 positive differences, that is all participants made at least some improvement, as can be seen in Table 4. The results are: W = 55.0, std error = 9.7, Standardised test statistic = 2.85, $p = .005$. So, both parametric and non-parametric tests confirm that there was a significant change from T1 to T2.

Appendix B. Details of statistical method and results

Table 4. Individual results for Expt and Comp groups, perception (hear diff + correct) across three times.

Expt	AE	BE	CE	DE	EE	FE	GE	HE	IE	JE
T1	7	6	2	5	9	9	6	11	6	8
T2	9	10	8	9	11	10	8	12	9	10
T3	9	10	7	10	11	11	7	12	8	--

Comp	AC	BC	CC	DC	EC	FC	GC	HC
T1	4	10	10	9	1	7	10	6
T2	4	10	10	9	4	6	9	4
T3	10	--	--	--	12	8	11	5

One might also ask for the results for identifying the correct version, regardless of whether they could hear a difference. Here the experimental group (n=10) improved from M = 7.7 (CI [6.8-8.6], SD = 1.8) to M = 9.7 (CI [9.1-10.4], SD = 1.3) at time 2. A *t*-test shows $t = -4.243$, SD = 1.5, df 9, $p = .002$. Bootstrapping for paired samples tests finds M = -2.0, std error = .452, CI [-2.7-1.3], $p = .007$. Volker's measure for paired samples finds a large effect size of d = 2. A non-parametric related samples Wilcoxon signed rank test also finds a significant effect of instruction; W = 45, std error = 8.3, std test stat = 2.70, $p = .007$ (9 positive differences, 1 tie). Even though the starting point was higher when excluding participants' ability to hear a difference, evidence of clear progress was still found.

Participants in the original comparison group were also able to receive the instruction. Only five of the eight in this group, actually came to the lessons and completed the post-test. This is understandable as it was getting towards the end of semester, and they were starting to focus on their assessments. It is worth noting that the five who did complete, tended to have greater problems with perception of word stress (M = 5.4 at Time 2, prior to instruction). They made good progress reaching M = 9.2 on the post-test. Because there were so few participants, no attempt has been made to run statistical tests on the performance of this five.

Finally, there is the consideration as to whether or not this learning can be extended to other contexts. As already noted, it was intended to only teach half of the tested items but in the event eight were taught, making an analysis of an impact difficult to quantify. However, a comparison of immediate pre-test with immediate post-test scores shows an average of 56.9% (57.7% overall) compared with 71.7% (79.2% overall). On the surface, it seems that some learning may have extended to other contexts, but it is not possible to say how much.