

TEACHING AND TESTING PERCEPTION OF WORD STRESS: MANY SHADES OF PERCEPTION

Graeme Couper, Auckland University of Technology

The teaching and measurement of L2 learners' speech perception has generally focused on high variability phonetic training (HVPT) and phoneme identification tasks. However, it is also necessary to consider how the classroom teacher can assist with perception development. It is argued here that fundamentally, accurate perception requires understanding of the underlying phonological concepts and that once these concepts have been learned they can drive behaviour in both speech production and perception. Because there are many shades of understanding, it is of interest to go beyond a binary assessment of learners' perception. The results presented here are part of a larger project researching the teaching and evaluation of word-stress pronunciation that has been designed for replication. The first rendition found significant quantitative gains were made (Couper, 2022). However, this article discusses the qualitative aspects of the study, attempting to describe and interpret the various shades of conceptual understanding shown by the participants (N=18) and how this changed over time and after instruction. The implications for both researchers and teachers of a new way of testing perception, and the related classroom practices that this testing is designed to reflect, are explored.

Cite as: Couper, G. (2022). Teaching and testing perception of word stress: many shades of perception. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. <https://doi.org/10.31274/psllt.13266>

INTRODUCTION

There has been a long history of laboratory-based L2 speech research into the development of L2 phonological categories (mainly in reference to being able to correctly categorise different phonemes such as /r/ versus /l/ in English) amongst adult learners (e.g. Lively, Logan & Pisoni, 1993). This has built on knowledge from phonetics and phonology that pronunciation is extremely variable, with the acoustic form changing with the same speaker, and of course with different speakers. It also varies according to the phonetic context and rate of speech, and sounds are not produced one after the other, rather they overlap and are coarticulated. These facts mean that we need other cues such as stress and intonation to help us recognise words (Lively, Pisoni & Goldinger, 1994). Flege's (1995) Speech Learning Model also recognises this and hypothesises that it is possible for adults to learn new categories. L2 speech research led to the development of perception training, known as high variability phonetic training (HVPT) which relies on exposure to multiple exemplars from multiple speakers with feedback (Pisoni & Lively, 1995). Evaluation of the impact of instruction has tended to rely on identification tasks in which participants are forced to choose between which of two productions is correct (Lee & Lyster, 2016). While laboratory or computer-based research on HVPT has been shown to have a positive effect on perception (Thomson, 2018), there has been little classroom-based research. One exception is Lee and Lyster (2016), who found a positive effect for teaching speech perception when accompanied by corrective feedback. Their study, like most others, focused on phonemes.

The conceptual approach taken here avoids the debate about the relationship between perception and production, and whether or not improvements in perception lead to improvements in production (Huensch, 2016). The approach focuses on the formation of phonological concepts, not just of phonemes but all aspects of pronunciation including syllables, word stress, prominence, and intonation. It is argued that we should start with concepts because they drive or control behaviour, in this case the perception and production of speech (Fraser, 2006). The approach itself is learner-centred, beginning with learners' knowledge and perception and working with cognitive capacities to compare and contrast and to discuss the nature of similarities and differences between concepts in different languages (Couper, 2011). Couper (2006, 2011, 2013) isolated and tested two key variables in effective pronunciation teaching and learning, namely Critical Listening and Socially Constructed Metalanguage (SCM). Critical Listening involves listening to two versions of the same production, one acceptable and one not, and being guided by a teacher to explore how they are different. This requires effective communication and leads to the development of classroom owned metalanguage, SCM, which helps in this process. The approach was developed in working with the concept of English syllable codas and this study investigates how those findings might be transferred to the teaching of a different phonological concept, namely, word stress.

Word stress, like all phonological concepts, is language specific. In English, it requires an understanding of the English concept of the syllable in the first instance. Word stress is realized through a combination of pitch, volume, and length, but more importantly, it becomes apparent through the reduction and vowel changes that occur in the surrounding unstressed syllables. Indeed, Cutler (2015) reports that the reduced vowel is the key to word stress recognition, even more so than correct placement. Therefore, while learners need to learn where the stress is placed, they also need to understand the nature of word stress. The focus of the teaching and testing in this study is on that understanding.

The many shades of conceptual understanding encouraged Couper (2011) to move beyond the traditionally binary identification test to one that might reflect these shades of perception. He developed a fine-grained perception test for syllable codas, based on Critical Listening, in which learners listened to contrasting pairs of the same word, once said correctly and once with an additional schwa. They were then asked to say if the two items were the same or different. If they could hear a difference, they had to say where the difference was and describe the difference, or if they could not do that, they had to say which one was better. This led to a scale describing the degree to which they understood the English concept of syllable codas. Following face-to-face discussion, it was determined if they clearly did not understand the concept, partially understood it but could not pinpoint the problem or explain it, probably understood it but any explanation was vague, or clearly understood it. The test revealed a finer-grained report on the status of learners' concepts, which it is argued is valuable from both a teaching and research perspective. With the exception of Thomson's (2012) study that included a measure of participants' confidence in the accuracy of their answers, this is perhaps the only perception evaluation method that has attempted to go beyond discrimination/identification tasks. The study reported here transfers this approach to the testing of word-stress perception, and leads to the following questions:

RQ1. How do participants perceive word stress?

RQ2. How does teaching word stress using a conceptual approach change participants' perceptions?

METHODS

This section focuses on collection and analysis of qualitative data from participants' descriptions of their perceptions of word stress. Detailed descriptions of teaching and testing items, and of the teaching intervention and worksheets used, are available at <https://pronunciationteaching.wordpress.com/psllt-teacher-notes-and-worksheets-2/> .

The participants were 18 students from an intact class at upper intermediate level preparing for academic study. They were aged from 18-50, with a wide range of backgrounds and L1s, most frequently: Mandarin, Cantonese, Korean, Japanese, Arabic, and Farsi.

The instruction involved four fifty-minute lessons on two-syllable words from the academic word list (being appropriate vocabulary work for these students). See the summary chart of words (in link above) sorted according to NZ English stress patterns.

An AB-BA design was used, half the class received the instructional intervention in the first half of the semester and the other half received it in the second half of the semester. While I was giving them the additional instruction, the class teacher did pronunciation related textbook exercises covering phrasing and tonic syllables, pronunciation of final consonants and past tense endings, and weak forms and connected speech.

Data Collection and Analysis

All participants completed tests at Time 1 and Time 2 (comparison group eight weeks later; experimental group post-teaching) in the computer lab using Voice Thread on the Blackboard learning platform. Participants listened to a word said twice, once with an appropriate stress pattern and once without, and were asked if they could hear a difference. They listened again and were asked to say which version was better. These results were easily quantified, showed significant gains, and have been reported in Couper (2022). If they said they could hear a difference, they listened yet again and tried to describe how they were different. Their descriptions have been analyzed here. The items they listened to, and the stress patterns used are summarized in Table 1 below.

Table 1

Description of Perception Test Items: Expected versus Unexpected Stress Patterns

word*	unexpected stress pattern
'breakfast	equal stress, no reduction on 2 nd syllable (sounds like 'fast' - /'brek'fa:st/)
co'mmit	said like 'comet' - /'kɒmət/, stress on first syllable.
'constant	equal stress, no reduction /'kɒn'stænt/
con'strain	equal stress, 1 st syllable slightly more, neither syllable reduced /'kɒn'streɪn/
'credit	stress on 2 nd syllable and no reduction on 1 st syllable /kre'dɪt/
ob'tain	equal stress, vowel on 2 nd syllable more like /e/ than /eɪ/ /'ɒb'teɪn/
oc'cur	stress on 1 st syllable, no stress on 2 nd and no change to vowel quality /'ɒkɜ:/
per'cent	stress on 1 st syllable, sounds like <i>person</i> , more /e/ than /ə/ /'pɜ:sɛn/
'licence	reduction on 1st syllable, stress on 2 nd /lɪ'sɛns/

'sector	reduction on 1 st syllable, stress on 2 nd /sek'tɔ:/
'wages	said as one syllable, 2 nd syllable omitted /'weɪdʒs/
e'merged	said as three syllables - stress on 2 nd and 3 rd syllables /ə'mɜː'dʒed/

*expected stress pattern shown, reduction expected on all unstressed syllables.

For analysis, participants' responses were downloaded as .wav files and anonymised for name, group, and time. They were then transcribed by a research assistant and double-checked by a colleague, who was given instructions on how to categorise their utterances:

- 2 = Accurately perceive which word is pronounced correctly, but they are unable to describe the difference except in very vague terms such as *clearer* or *sounds wrong*. This suggests limited awareness of pronunciation in general and of stress in particular.
- 3 = Accurate perception of which word is better and awareness of differences in pronunciation, usually in reference to phonemic differences, but there is still limited evidence of awareness of the role that stress plays in making the pronunciation sound better.
- 4 = Accurately understands the English language concept of stress (Note: items *emerged* and *wages* describe additional and deleted syllables respectively).

The raters, the researcher and a colleague, first rated a small sample of the anonymised sound files independently before meeting to confer over uncertainties and differences. They then rated the remainder and agreed in most cases, and following further discussion reached full agreement on the others, making slight adjustments to ratings. Despite the subjectivity and difficulty in understanding exactly what participants meant in some cases, and the time involved, consensus was achievable. The analysis would have been easier if the data had been collected face-to-face rather than through a recording because it would have allowed for checking of what was meant. This would have helped where the participant did not clearly verbalise their conceptual understanding, which potentially led them to being placed at level 3 rather than 4. On occasion, the participant just repeated the word, mirroring the two versions with varying levels of accuracy. In these instances, an acoustic analysis was considered, but it was felt this would add layers of unnecessary complication and add little to our understanding of their perceptions. Therefore, they were coded as level 2.

RESULTS

Firstly, I will report on the descriptions participants used and how they were assigned to different levels of understanding (RQ1). Then I will report on how those understandings appeared to change following instruction for the experimental group and over time for the comparison group (RQ2).

Descriptions of Perception

Perception was rated at 0 if participants said they could not hear a difference, and at 1 if they could hear the difference but could not identify the best version. This report focuses on the next three levels. They were rated at 2 if they could hear the difference and identify the best version but did not know or could not say why it was best. They were rated at 3 if they had some awareness of the difference but may not have been focusing on the salient differences, and at 4 if they were aware of the salient differences.

Level 2

At this level, participants heard the words as different but could not say why. They said, "The first one is better", or added a general comment such as "the second one is much better because the second one is much clearer than the first one". Sometimes they repeated the words to try and indicate the difference, for example "the first one sound is occur the second one sound is occur the sound is different". Also, if they said "I like the second one is better because the first one has pronounce another word" this was rated at level 2 because there is no evidence of awareness of a possible cause for the difference.

Level 3

Here, there was some indication of awareness, although it was not clear if it was awareness of the different stress pattern. They may have used words like *emphasis*, *accent*, *pressure*, but not clearly enough to be sure they were referring to word stress [comments in square brackets]. Examples of this were:

- The second one has one emphasis point. I think it's good [not clear where they heard the "emphasis point"]
- 'Occur' seems exact one which is correct, the second one mostly put some pressure on "o" and "u" which says 'occur' which is not correct one. The correct one is 'occur; like 'e' in the first 'occur' [not clear because there was no stress on the second syllable, so they may be aware of stress but they are still not hearing it correctly]
- the first one is much clearer than the second one and the accent is better [difficult to analyse because of the different meanings of accent, but the raters thought the reference may have been to accent rather than stress.]

They also pointed to things like it sounding like two words, changes in intonation, the length of the words, differences in individual sounds. Examples of this were:

- Second type is like.....depend on two words
- the first one tone is going up but second one tone is going down
- one is intonation so natural but second one is not natural intonation
- The second one shorter than first one. It's easy to listen
- The first one she didn't pronounce the /aɪ/ and in the second one she pronounce /aɪ/
- The second is correct because 'r' is silent in this word

Level 4

At this level, there was clear recognition that the difficulty was with word stress. Sometimes they used the term directly, for example "The first one has stress on the right place, the second one has on the wrong place". Sometimes, as in "I think the second is clear because is easy to hear and not really stress on the word", the raters needed some discussion before concluding that the participant understood the word stress was incorrect.

Participants also described stress in other ways, such as:

- The second one is correct because the first syllable comes short the second one little bit long
- There is different between those two is two different pronouncing. First one different accent, first one is 'obtain', second one is 'obtin'
- I think the second one is right pronunciation, I think maybe the accent is last place of the word so I choose last one

Changes in Perception

No claims are being made about the generalisability of results showing movement between levels 2, 3, and 4, but they can nevertheless be quantified. Table 1 shows the number of items, out of 12, rated at each level (2, 3, and 4) and in total (2/3/4, i.e. the number where they both heard a difference and chose the best option) for each participant at Time 1 (pre-instruction) and Time 2 (post-instruction). It can be seen that all participants made progress in total, with seven also making progress between levels 2, 3, and 4, while three did not change.

Table 2

Individual Changes in Level for the Experimental Group Following Instruction

Student	AE1		BE1		CE1		DE1		EE1		FE1		GE1		HE1		IE1		JE1	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Level 2	4	4	1	3	2	3	5	9	9	7	8	9		1	11	6	5	1	4	6
Level 3			5	7		5				1	1	1	1			1	1	8	4	4
Level 4	3	5							3				5	7		5				
Total2/3/4	7	9	6	10	2	8	5	9	9	11	9	10	6	8	11	12	6	9	8	10
Change		+		+		+		=		+		=		+		+		+		=

Those in the comparison group also appeared to make some progress, in terms of progressing from levels 2 to 3 and 4, even though as a group they did not improve in the Total2/3/4. However, unlike the experimental group, four of them made progress but the other four got worse. Of the four who made progress, three had scored highly (in Total2/3/4) on the initial pre-test while the fourth one had scored particularly poorly.

Table 3

Individual Changes in Level for the Comparison Group over Time

Student	AC1		BC1		CC1		DC1		EC1		FC1		GC1		HC1	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Level 2	2	3	9	5	9	7	7	3	1		1	1	6	6	4	4
Level 3	2	1	1	3		1	2	3		3	2	2	4	3	2	
Level 4				2	1	2		3		1	4	3				
Total2/3/4	4	4	10	10	10	10	9	9	1	4	7	6	10	9	6	4
Change		-		+		+		+		+		-		-		-

Finally, five participants in the comparison group later attended the intervention classes and re-sat the tests. They all made progress in total and between levels 2, 3, and 4.

Table 4*Individual Changes in Level Post-intervention for the Comparison Group*

Student	AC		EC		FC		GC		HC	
Time	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3
Level 2	3	3		2	1	3	6	4	4	3
Level 3	1	7	3	9	2	1	3	6		2
Level 4			1	1	3	4		1		
Total2/3/4	4	10	4	12	6	8	9	11	4	5
Change		+		+		+		+		+

DISCUSSION

In response to the first research question, this approach to teaching and testing provided insights into participants' perceptions of word stress. During in-class discussions SCM developed in relation to critical listening, with participants describing stress using words such as *stronger, longer, pressure, emphasis, tone is going up/down, accent, and stress*. In talking about the unstressed syllable, they used words such as *smaller, quieter, and shorter*. The words used both in class and in the tests indicated how they needed help to notice the difference and to also notice what was happening with the unstressed syllable. Even though it is difficult to analyze, especially when relying on comments the students have recorded, rather than face-to-face, the fine-grained perception test revealed a great deal about participants' perceptions of word stress. In particular, even though participants could often choose the best pronunciation, there was little recognition that the main reason behind this was word stress. This supports the findings from Couper (2011, 2013) and shows that they can be transferred to the teaching and testing of word stress. It underscores the value of moving beyond a binary test of identification when trying to understand how learners perceive the target language speech. It is argued that this reveals more of the process learners are going through and that this has important implications for teachers. It also shows the value of critical listening and SCM, as found in Couper (2006, 2011, 2013).

To answer the second research question, the fine-grained perception test found qualitative changes in perception following teaching. Although the gains made in being able to hear the difference and choose the best option were clear cut (that is Total2/3/4, see Couper, 2022, for statistical analysis), there were also small but steady gains in the way participants reported their perceptions. This suggests that although they have made gains, there is still more work to do to fully understand the concept. Four of those in the experimental group also made some gains over the eight weeks between their two tests, but four regressed. It may be that they learned this from their lessons, or perhaps it was an impact of familiarity with the test. Of the four who made progress, three had scored highly (in Total2/3/4) on the initial pre-test while the fourth one had scored particularly poorly, and perhaps this was a factor. Nevertheless, the participants improved their perception following conceptual instruction, which supports Couper's earlier findings (2006, 2011, 2013) that this approach to instruction leads to improvements in both perception and production. It also supports Lee and Lyster's (2016) finding that instruction can lead to improved perception when accompanied by corrective feedback. The ability to form new phonological concepts and therefore to accurately categorise L2 speech supports Flege's

(1995) theory that adults can re-set category boundaries and findings that HVPT improves perception (Thomson, 2018).

There are implications for both teachers and researchers. Researchers might use more fine-grained perception tests to reveal more of the concept formation process (Fraser, 2006). Both teachers and researchers need to recognise that learning L2 phonological concepts is a process. This process starts with only vaguely noticing the characteristics of the category and slowly working out category boundaries through multiple exposures (as noted in HVPT). In the case of word stress, learners need to understand the English concept of the syllable first. Then, they need to understand both how English stress is realised and the role of unstressed syllables (Cutler, 2015). This learning process, where it has not occurred naturally, requires contrasting and comparing L2 to L1 concepts. Finally, stress placement needs to be learned along with the learning of new vocabulary, but first the underlying concept must be understood. Because there are many stages in the formation of concepts, and thus, many shades of perception, teachers need insights into how their learners think about the target L2 category. These can be gained through Critical Listening activities and accompanying SCM and they inform teachers as they provide on-going explanations and practice activities.

Teachers and researchers who might be interested in replicating this study should consult the full description provided at <https://pronunciationteaching.wordpress.com/psllt-teacher-notes-and-worksheets-2/> They will also want to note the limitations in being able to be sure that participants' responses were being interpreted correctly. As was noted above, a face-to-face discussion would have given greater certainty in those cases where participants did not clearly verbalise their conceptual understanding. A further limitation was that responses depended on participants' motivation and time in supplying explanations. These limitations suggest face-to-face collection of data should be considered.

In conclusion, this report has focused on the descriptive aspect of a fine-grained perception test that is part of a study into using a conceptual approach to teaching pronunciation. It argues that teaching and testing needs to reflect the many shades of perception and that by focusing on concepts, gains will be made in both production and perception.

ACKNOWLEDGMENTS

I would like to thank my colleague, Annette Sachtleben, for her help with data analysis and colleagues in France, Lada Achilova, Dan Frost, and Alice Henderson for their reflections and their willingness to be involved in future rounds of this project.

ABOUT THE AUTHOR

Graeme Couper, an Associate Professor at Auckland University of Technology, has many years teaching experience in a wide range of countries and contexts which he applies to his research into L2 pronunciation teaching and learning. His research has produced theoretically and empirically supported explanations of what makes pronunciation teaching effective. He may be contacted at graeme.couper@aut.ac.nz and welcomes enquiries from teachers and researchers interested in replicating this study.

REFERENCES

- 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.
- Couper, G. (2013). Talking about pronunciation: Socially constructing metalanguage. *English Australia*, 29(1), 3-18.
- Couper, G. (2022). A conceptual approach to teaching L2 pronunciation: Perception of word stress. *TESL-EJ* 26(1), 1-20. <https://doi.org/10.55593/ej.26101a6>
- Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reid & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 106-124). Malden, MA: Wiley Blackwell.
- Flege, J. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore, MD: York Press.
- Fraser, H. (2006). Helping teachers help students with pronunciation: a cognitive approach. *Prospect*, 21(1), 80-96.
- Huensch, A. (2016). Perceptual phonetic training improves production in larger discourse contexts. *Journal of Second Language Pronunciation*, 2(2), 183-207. <https://doi:10.1075/jslp.2.2.03hue>
- Lee, A., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, 38, 35-64.
- Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/ 2: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of The Acoustical Society of America*, 94, 1242-1255.
- Lively, S., Pisoni, D., & Goldinger, S. (1994). Spoken word recognition: Research and theory. In M.A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 265-301). San Diego, CA: Academic 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). Timonium, MD: York Press.
- Thomson, R. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, 62(4), 1231-1258. <https://doi: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>