BETH SHEPPARD
BRIAN BUTLER
University of Oregon

Insights Into Student Listening From Paused Transcription

Listening comprehension is an essential and challenging skill for language learners, and listening instruction can also be a challenge for language instructors, since they have little access to the listening process inside students' minds. Greater knowledge about what learners perceive when they listen could help language teachers better tailor their instruction to student needs. In this mixed-methods study, students at 2 proficiency levels participated in a listening test based on Field's paused transcription method (2008a, 2008c, 2011). Results were analyzed quantitatively on the basis of student and text level, word class, and articulation rate. Transcription errors were analyzed qualitatively to identify patterns of mishearing. Paused transcription is recommended as a classroom activity to identify and raise awareness of student listening challenges.

Second language (L2) listening presents major challenges to learners, since the speed and lexical/syntactical choices of spoken discourse are out of the control of the listener. At the same time, listening is an essential skill for learners, since listening can provide many opportunities for continued language learning. For international university students in the US, listening also represents a primary way of accessing necessary information. It is important, therefore, to help incoming international students develop their listening skills as much as possible before they begin their university studies.

What Makes Listening Difficult

To help students develop listening skills in a second language, it is helpful know what makes listening difficult for them. Some studies have approached this question by asking learners why a text feels difficult. In response to these questions, learners have reported that sec-

ond language listening is hard for the following reasons (Goh, 2000; Liu 2002; Renandya & Farrell, 2011):

- The speaker is too fast.
- They do not know all the words.
- They cannot recognize known words in context.
- They cannot focus on the whole message.
- They feel anxious.

Other studies have approached this question by comparing language learner results on listening tests with specific differences in the audio texts. The following text factors have been found to increase the difficulty of L2 listening comprehension (Bloomfield et al., 2011; Brunfaut & Revesz, 2015; Revesz & Brunfaut, 2012):

- Greater lexical range and density;
- More formal, literate discourse structure (reduced redundancy, greater referential cohesion, greater information density);
- Indirectness (requiring listeners to infer implied meaning);
- Unfamiliar accent;
- Faster articulation rate and reduced pauses.

These are the challenges learners need to overcome as they develop into proficient L2 listeners.

Bottom-Up and Top-Down Listening Processes

Most discussions of second language listening development refer to top-down and bottom-up processes, both of which are essential for listening comprehension. Top-down (knowledge-based, concept-driven) processes involve using knowledge of the world, speech context, and recent co-text to predict or limit possible interpretations of the speaker's message. Bottom-up (text-based, stimulus-driven) processes involve recognizing phonemes, syllables, words, and relationships between words to decipher the speaker's message. Top-down and bottom-up processes are used simultaneously by all listeners, but skilled and novice listeners may use them in different ways. In particular, Field (2008d) emphasizes that skilled listeners use top-down processes to amplify and extend the speaker's message on the basis of automatic and very effective bottom-up processing, while novice listeners use top-down processes to compensate for incomplete bottom-up processing by making reasonable guesses about missed words and phrases.

In this study, we focus on the subset of bottom-up processes by which listeners identify words from the stream of sound. These include phoneme recognition, locating word boundaries, and lexical matching. We will refer to these processes as aural decoding.

Listening Instruction

A good deal of recent discourse (e.g., Field, 2008d; Siegel, 2014; Vandergrift, 2004) has suggested that ESL listening instruction must place a greater focus on the *process* of listening, rather than just the *product* of listening in the form of correct answers to comprehension questions. This attention to process can emphasize top-down skills, such as explicit instruction in metacognitive listening strategies (Vandergrift & Goh, 2012), or bottom-up skills, such as diagnosis of specific aural decoding problems followed by practice in those areas (Field, 2008d). A balance of these two approaches seems most likely to meet students' needs, but the literature indicates an imbalance in current teaching practices, with more attention needed to bottom-up skills (Field, 2008d; Siegel & Siegel, 2015; Vandergrift, 2004).

The ability to quickly and automatically decode the speech stream into known words is a key skill for successful listening. Tsui & Fullilove (1998) found that strong bottom-up skills distinguish stronger from weaker performers on a listening test. To help students improve these skills, Field (2008d) proposed a diagnostic approach in which the teacher ascertains which bottom-up processes are causing challenges and designs short instructional activities to practice precisely these processes. In order to apply a diagnostic approach to listening instruction, however, it is necessary to find out what learners hear when they listen.

The Present Study

We are instructors in a moderately large Intensive English Program (IEP) at a moderately large public university. As at many other universities, our students can begin their university studies when they reach an intermediate to high-intermediate language level. The ability of students at this level to decode connected speech has been found to be remarkably low, with around 60% of words decoded on average, as compared to around 95% for native speakers (Estes, 2014; Field, 2008a, 2008c, 2011).

We were interested in learning more about the decoding ability of our own intermediate-level learners. Past studies have found that learners decode content words more accurately than function words, in spite of the greater frequency of function words. We were interested in this result, and we also wondered how articulation rate would

affect decoding, since students often state a belief that they cannot understand when the text is fast. We also hypothesized that students' specific errors in paused transcription would offer clues to diagnose which subskills of listening were challenging for them, and therefore this method could be a useful tool in the classroom.

Thus our research questions are:

1. How completely do our students decode listening texts at various levels?
2. Will students decode more content words than function words?
3. Will students decode more words with a slower articulation rate?
4. Can students' transcriptions provide insight into their listening processes?

Method

Since aural decoding and comprehension occur inside the mind, they cannot be directly observed. Researchers have approached this problem using think-aloud protocols and retrospective interviews (e.g., Goh, 2000; Zielenski, 2008), paused transcription (e.g., Estes, 2014; Field, 2008c), and priming studies (e.g., Cutler, 2012), among others. Paused transcription has the advantage that it focuses specifically on aural decoding, but without divorcing the target phrases from a natural context in connected speech and discourse or preventing learners from also applying top-down processes as they would in natural listening. In paused transcription, subjects are asked to listen to an extended text into which pauses have been inserted at irregular intervals. During each pause, subjects write down the last phrase (4-5 words) that they heard. The written phrases can then be compared to the original text and coded for accuracy.

> The rationale for this method is that it taps into a listening process that replicates a real-world one. Subjects listen to the recording with a view to following its meaning, and it is only when a pause occurs that they switch attention to word level. Memory effects are limited by the fact that subjects are asked to transcribe around four or five words – well within the range of Miller's (1956) seven plus or minus two. Furthermore … listeners retain verbatim word forms until major clause boundaries and only then "wrap them up" by replacing them with representations in propositional form. (Field, 2008b, pp. 16-17)

Participants

Study participants were students in intact listening and speaking classes at a university-based Intensive English Program. Participants (*N*=77) included 48 upper-level students and 29 midlevel students who spoke Chinese (65.4%), Japanese (10.2%), or Arabic (24.4%) as their first language. They had already studied in the US for an average of about 11 months, and a *t*-test showed that the length of residence was not significantly different between students in the two levels.

Materials

Three listening texts were used for the paused transcription study. The first two texts were from listening textbooks and graded for easy comprehension at the two proficiency levels. A third text was taken from an authentic university lecture available online. In addition, a very short text was prepared for use as a sample/warm-up activity to clarify the paused transcription procedure.

All three audio texts were similar in length (see Table 1). Each was structured as an academic talk or lecture, with a relatively informal tone and some features of oral language (the textbook recordings were scripted and performed by actors, but some of these features were written into the script). All speakers had standard North American accents.

Table 1
Origin, Topic, and Length of Listening Texts

	Warm-up	*Text 1*	*Text 2*	*Text 3*
Origin	*Pathways 2*	*Pathways 2*	*Learn to Listen, Listen to Learn*	Open Yale Courses
Topic	Comparing people	Changes in our world	Women and work	Our relationship to food
Length	0:44	2:58	3:32	3:21
Words	104 (142 wpm*)	387 (130 wpm)	498 (141 wpm)	561 (167 wpm)

Note. *words per minute.

For each audio text, Cobb's (n.d.) VocabProfiler was used to select four-word phrases for transcription. Twelve phrases were selected from each audio text, for a total of 144 words (see Appendix A). Of

these, 141 were found among the 1,000 most commonly used words in English based on the General Service List (West, 1953), and three were among the second thousand words of the General Service List ("dance," "repeat," and "probably"). These words were estimated to be familiar to students at both levels. Thus study participants could be expected to be familiar with most or all of the words selected for transcription.

Procedure

The study was conducted as a listening exercise during class time. The first author conducted all sessions of the study. After reading instructions and giving consent in their L1, participants completed a brief questionnaire about their language background and then the warm-up paused transcription activity. They were then instructed to explain the activity to each other in their L1. Once all participants understood the instructions, the three texts were played, always in the same order (Text 1, Text 2, Text 3). Participants wrote their transcriptions on a paper packet. At the end of each audio text, participants rated their comprehension of the text from 1 to 5 and then turned the page for the next audio text.

Three class instructors chose to participate in the study, transcribing in the pauses as their students did. All three had 100% correct transcriptions.

Data Analysis

Each transcribed target word was coded as correct or incorrect. Only the target words (last four words spoken before the beep/pause) were coded and any extra words were ignored. Missed words were coded the same as incorrect words. When words were present but transcribed out of order, they were still coded as correct. Words with morphological errors (generally in endings for tense and number) were coded as correct. Misspelled words were also coded as correct, if they could clearly be identified as the intended word. The first author coded all words and the second author coded a subset of 10%. Interrater agreement was found to be 98.1%. Examples of coding can be found in Table 2.

During the process of coding for quantitative analysis, interesting transcriptions were highlighted for qualitative analysis. In addition, an overall difficulty score was calculated for each phrase (an average of the percent correct for the four words), and the most difficult phrases were flagged for further qualitative error analysis. For selected phrases, transcription errors were tallied and categorized. The researchers listened again to the target phrases, made notes about the speaker's

Table 2
Sample Coding for Target Word Transcriptions

Target word	Transcription	Coded
raised	Raise	correct
raised	Rave	incorrect
woman	Women	correct
dress	Drees	correct
dress	Drac	incorrect
have	had, has	correct
their	The	incorrect

delivery, and speculated about the origin of specific errors. In this process, several broad types of errors emerged as common and significant in the data. All transcriptions of the difficult phrases were then reanalyzed with reference to these error types.

Results and Discussion

Research Question 1

How completely do our students decode listening texts at various levels?

With 144 target words and 77 participants, there were 11,088 target tokens. Of these, 7,414 target tokens were coded as correctly transcribed, a correct transcription rate of 67%. Upper-level students (intermediate proficiency) transcribed 73% correctly, while midlevel (preintermediate proficiency) students were successful with 54% of the target tokens. The percent of correctly transcribed tokens by text and student level can be seen in Figure 1.
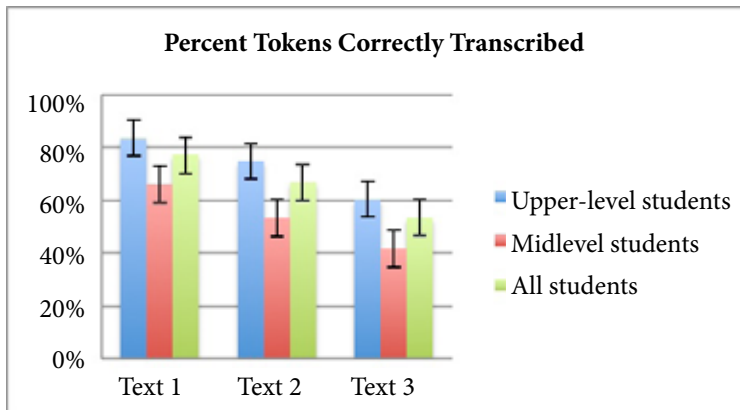

Figure 1. Percent of tokens correctly transcribed.

An ANOVA confirmed that differences in overall transcription accuracy were significant by student group, $F(1, 282) = 48.80$, $p < .001$, and by text, $F(2, 282) = 24.76$, $p < .001$. Full statistics can be found in Appendix B, Tables 1 and 2.

Both groups of students experienced significant gaps in their aural decoding, with less than three quarters of the words decoded in every group except the upper-level students listening to the easiest text. The upper-level students were a few weeks away from exiting the IEP and beginning university classes, yet they could decode only about 60% of the words in the first four minutes of the first lecture of an undergraduate class (Text 3). A lexical coverage of 90-95% has been found to be sufficient for adequate listening comprehension (Van Zeeland & Schmidt, 2012). We can therefore see that when international university students enter with minimally acceptable English language proficiency, decoding perhaps 60-70% of the words in a typical lecture, they will be at a significant disadvantage in lecture comprehension.

Research Question 2
Will students decode more content words than function words?

Overall, study participants were able to correctly transcribe 76% of content words and 54% of function words. A *t*-test confirms that transcriptions of content words (*n*=80, *M*=0.75, *SD*=0.19) were significantly more accurate than those of function words (*n*=64, *M*=0.54, *SD*=0.24), $t(142) = 6.06$, $p < .001$. The results are presented in Figure 2.

This finding aligns with results of previous studies that have found that language learners can decode more content words than function
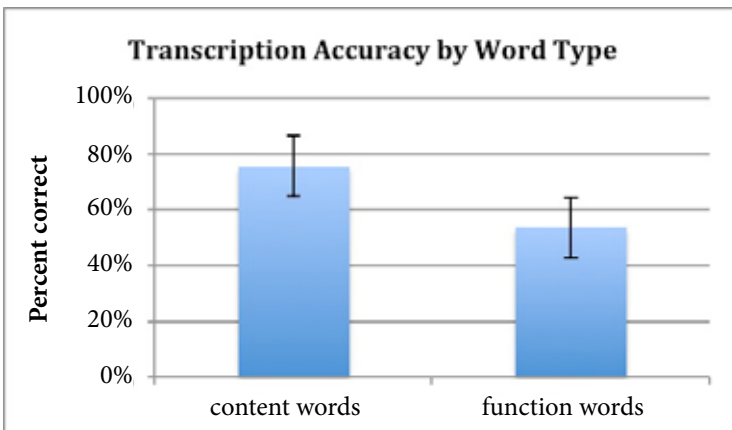

Figure 2. Average transcription accuracy by word type.

words. ESL students at these levels are likely familiar with all function words and encounter them frequently, but these words are often reduced in speech and also are usually less essential to understanding the overall message of an utterance. In fact, even L1 listeners have been found to rely on context to fully decode function words (Herron & Bates, 1997, as cited in Field, 2008c).

With limited available attention, a focus on decoding content words is probably an effective choice for L2 listeners. At times, however, function words can have a significant effect on meaning. Consider, for example, the effect of misunderstanding a preposition or pronoun in the sentence "I bought it for you." Also, if students can hear and understand function words, then listening becomes an avenue for them to improve their productive language skill through exposure to correct grammar in context. Field (2008c, 2008d) suggests activities to help language students pay attention to function words in listening. For example, teachers can train learners to infer function words after perceiving content words by pausing an audio text (or dictation) before a function word and asking students to predict what word will come next, or teachers can have their students explicitly practice perceiving unstressed function words and suffixes through a variety of targeted dictation exercises.

Research Question 3

Will students decode more known words with a slower articulation rate?

Language students often state a belief that difficulties in listening comprehension arise from faster audio delivery (e.g., Goh, 2000), but studies on speed and listening comprehension have found mixed results. It appears that pauses are helpful to L2 listeners, and increased speed can negatively affect comprehension, but slower rates do not always improve comprehension and students often misattribute other causes of difficulty to speed (Bloomfield et al., 2011).

In the current study, a simple measure of articulation rate (phrase time divided by pronounced syllables) was calculated for each four-word target phrase (n=36, M=4.704, SD=0.899). A basic measure of phrase difficulty was calculated by averaging the percent of participants who correctly transcribed each of the four words (n=36, M=0.658, SD=0.161). No significant correlation was found between these two measures, r = -0.253, n = 36, p = .137, indicating a lack of strong relationship between within-phrase articulation rate and success in decoding the words of the phrase. Figure 3 shows the relationship between transcription accuracy and articulation rate for the 36 phrases.
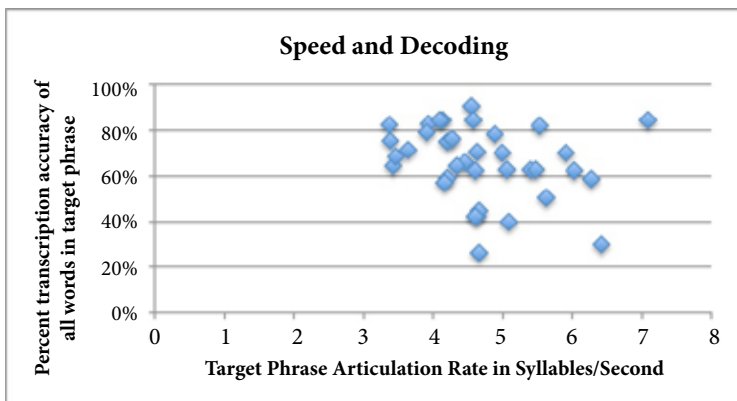
Figure 3. Phrasal articulation rate and average transcription accuracy.

This result is not surprising against the background of research mentioned above, but still it might come as a revelation to some teachers and many students. Simply informing students of these findings could have an impact on students' emotions about listening comprehension. Since listener anxiety has been found to have a powerful effect on comprehension scores (Bloomfield et al., 2011), affective issues are one key to helping students listen more successfully. Finally, when teachers select recorded authentic texts for classroom use, they may often base decisions on "speed" of delivery. These results add to data suggesting that teachers should consider the speaker's use of pauses rather than overall words per minute or articulation rate.

Research Question 4

Can students' transcriptions provide insight into their listening processes?

Qualitative examination of transcription errors led to a variety of insights about participant misunderstandings and gave hints about the listening processes they struggled with. We focused our error analysis on the phrases that proved most difficult for participants, based on average words transcribed correctly. Both researchers examined these phrases, considering the frequency and possible origin of each error.

Several categories of errors emerged that we will discuss individually, giving example participant transcriptions for each. We will also suggest some simple classroom activities that could be used to draw students' attention to these issues and practice skills (both bottom-up and top-down) that may underlie or support them. The categories are *word segmentation, phonemes, unknown words and phrases,* and *top-down fabrications.*

Word Segmentation

One challenge of L2 listening is to locate the beginnings and ends of words, since there are usually no silent spaces between them. Listeners employ several strategies to meet this challenge, including vocabulary knowledge (recognizing one word will also locate the beginning of the next word), knowledge of language-specific rules about which phonemes and combinations of phonemes can appear in word-initial and word-final positions (phonotactics), and strategies involving stress and rhythm. The most effective strategy for listeners of English is to initially assume that each stressed (unreduced) syllable begins a new content word and adjust as needed based on other strategies (Cutler, 2012). For the most part, the word-segmentation errors in our study resulted in transcriptions that also followed this primary strategy. In other words, participants did not incorrectly place stressed syllables in the middle of transcribed words. Three example phrases are analyzed below.

Text 2 phrase 6—"Some of the factors a woman might want **to take into account**—"

Incorrect transcription	N	*Error analysis*
… taking to account	17	/tek/ is a stressed syllable, which begins a content word. In this common error, /tek/ is still correctly placed at the beginning of a word. /ɪntu/ is a function word of two unstressed syllables, and students have mistakenly assigned the first unstressed syllable of /ɪntu/ as an unstressed suffix of the preceding content word. This is reasonable from the standpoint of word-segmentation strategy, but syntax and subtle clues in delivery could have helped disambiguate the phrase.
… a count … count … a corn, a comet … a(n)- [no following word transcribed]	10 4 2 9	/caʊnt/ is a stressed syllable, so it is reasonable to guess that it will begin a content word and therefore to assume that the preceding /ə/ is a separate function word. Here knowledge of English collocations could help disambiguate the phrase.

Text 1 phrase 5—"Native American music **used to be played**—"

For this phrase it is noteworthy that study participants did not command the grammar in "used to be played"—70% of all participants were able to transcribe some form of both content words ("use" and "play"), but only 22% were able to transcribe the whole phrase with correct function words and morphemes. Many omitted one or more of the content words (e.g., "used to play" *n*=13).

Incorrect transcription	N	Error analysis
Usually play	2	This phrase included four syllables, with a stress on the first and fourth syllables. Like the previous example, the rule of assuming that stressed syllables begin content words resulted in more than one possible interpretation, and these four participants selected an incorrect interpretation that had the same rhythm and vowels, but meant that they transcribed two consonants incorrectly. In addition to the consonants, syntax could have disambiguated this phrase.
Usually like to play	1	
Usually to played	1	

Text 1 phrase 1—"Changes take place over time, so we **don't always notice them**—"

Incorrect transcription	N	Error analysis
We don't always know this sound	1	The frequent word-segmentation error represented here is a perception of the second (unstressed) syllable of "notice" as a separate (unstressed) function word. As above, this interpretation follows the basic word-segmentation assumption. Various phonemic changes are associated with this shift in word boundaries, and the results vary in their syntactic and semantic plausibility.
We don't always know the sound	1	
We don't always know this song	1	
We don't always know the change	1	
We don't always know understand	1	
We do not always understand	1	
So we don't understand	1	
Don't always don't the sound	1	

We don't always know this We don't always know that We don't always know them Always no them We don't know	1 3 9 1 1	These are similar to the above, except that one syllable is missing—either the unstressed syllable of "notice" or the last function word. It is thus unclear whether they represent word-segmentation errors or a missed word.
We don't know all with them	2	Here, "always" has been split into two words (and there is a reversal of words/sounds as well).
We always listen	1	Here we see a different segmentation, with the unstressed second syllable of "notice" misperceived as a stressed initial syllable of a difference content word ("listen"), along with some phoneme errors.

In most of the clear examples of incorrect word segmentation, participants were found to have maintained the pattern of stressed (unreduced) syllables' beginning content words. Participants applied a nativelike strategy to segment words, successfully segmenting a great majority of the words they heard. The examples presented here are the clearest incidences of word-segmentation error precisely because they maintain some of the rhythm and phonemes of the original. Less-transparent segmentation errors may underlie other incorrect transcriptions as well.

When listeners misperceive word boundaries, it can cause lasting confusion. For language learners, aural misperception of word boundaries is a more common and longer-lasting phenomenon than for more expert listeners. The learner's smaller number of known words and uncertainty in phonemic matches can lead to more frequent errors, and a lack of confidence in general comprehension can impede learners' recognition and correction of previous mistakes in decoding (Field, 2008b).

Instructional Suggestions for Word Segmentation

- **Dictation**: Brief dictation exercises can be an excellent tar-geted-listening task, as long as the target sentences are spoken with a natural speech rate and style. While maintaining

this natural delivery, length, lexical choices, and grammatical complexity can be adjusted to student proficiency levels. Students will practice word segmentation as they listen and transcribe sentences and phrases.

- **Elicited imitation**: This technique is similar to dictation, except that comprehension is displayed via speaking rather than writing. Students listen to phrases spoken naturally and repeat back what they hear. Extremely short phrases may be repeated back phonetically, but with more than a few syllables repetition requires comprehension (see Yan, Maeda, Lv, & Ginther, 2016, for a meta-analysis of elicited imitation as a measure of L2 proficiency).

- **Paused transcription detectives**: With teacher guidance, students can find segmentation errors in their own paused transcription practice and examine the pronunciation differences between the spoken phrase and their transcription, pronouncing and practicing the phrases. They should also examine co-text for semantic or syntactic clues to correct word segmentation.

Phonemes

Research has indicated that word codas are less salient than onsets, and that students have more trouble correctly identifying vowels than consonants (Cross, 2009; Field, 2004; Rost, 2016). The participants in our study did have a tendency to transcribe wrong words beginning with the right sounds, and to transcribe syllables with correct consonants and incorrect vowels. However, we also found opposite examples, in which participants transcribed wrong words ending with the right sounds, and examples in which the vowel was correct but the consonants were inaccurate. Two example phrases are analyzed below.

In the example Text 2 phrase 10, we can see that the /st/ onset of "study" was quite salient, and the final /i/ of the word was also maintained in several of these erroneous transcriptions. The middle of the word was not maintained in any erroneous transcriptions.

For the function word "was," the first phoneme was maintained in erroneous transcriptions. Participants never mistook this word for a content word, instead substituting other function words beginning with /w/. Both function words in this phrase were often omitted.

Five percent of all participants wrote "down" for "done." In this case, initial and final consonants were both maintained, but the vowel was not decoded correctly. The erroneous transcription "stone" for done may have had some relationship with the /st/ of "study," but since the full transcription in this case was "stay with stone," we know that

"stone" was an attempt at "done." The final consonant is correctly decoded, and the middle vowel is similar to the target but still incorrect.

Text 2 phrase 10—"I'd like to tell you about a **study that was done—**"

Target word	Study		That		Was		Done	
	Error	**N**	*Error*	**N**	*Error*	**N**	*Error*	**N**
	Stay	4	It	1	With	4	down	4
	Stiy	1	And	2	Will	1	Stone	1
	Staied	1	We	1				
	Stains	1	What	1				
Incorrect transcriptions	Still	1	The	1				
	Stand	1	Language*	1	Language*	1		
	Story	2	Almost*	1	Almost*	1		
	State	1						
	Outside	1						
	Research	1						
Omissions	16		56		40		30	
Correct transcriptions	47		13		30		42	

Note. *These two-syllable words seemed to replace both function words.

In the example Text 3 phrase 11, the second word of this phrase, "wouldn't," was the only word with a 0% correct transcription rate in this study. Forty-two erroneous transcriptions are presented in the chart. The other 35 participants did not transcribe this word. The great majority of erroneous transcriptions (39/42) maintain the correct initial phoneme. Participants who wrote "would" were correct about the entire first syllable (although the meaning of the sentence will still be misunderstood), while others were able to transcribe some of the word-final consonants, for example, "want."

For "seem," the most common error was a failure to perceive the final /m/ sound, resulting in transcriptions of "see," which indicates correct perception of the word-initial consonant and the vowel (various morphological endings added to "see" may have been related to the application of top-down skills). However, other participants maintained the word-final consonant but not the vowel ("same"), while others maintained only the /i/ vowel sound ("think," "technique").

More than half of the erroneous transcriptions for the final word of this phrase, "like," maintained the correct vowel sound. None maintained the correct consonants in word-initial or word-final position.

Text 3 phrase 11—"Burning more calories creating a paper than you guys have too. **That wouldn't seem like—**"

Target word	That		Wouldn't		Seem		Like	
	Error	N	*Error*	N	*Error*	N	*Error*	N
	(Now) I	14	One	13	See*	28	My	6
	Then	3	Was	8	Same	3	Why	3
	The	3	Will	5	Think	3	A lot	2
Incorrect transcriptions	The	2	Would	5	Say	2	Have	2
	Him	1	Want	4			Might	2
	It	1	We	3			Wise	1
	There	1	Can	2			How	1
			May	1			As	1
			When	1				
Omissions	18		35		14		26	
Correct transcriptions	34		0		27		33	

Note. *Some form of "see" (see, seen, sees, seeing).

When students perceive a phoneme incorrectly or ambiguously, it can lead to identification of the wrong word, as we see in these examples. Even when it does not lead to incorrect word identification, it can slow down and complicate aural decoding by introducing additional competition from "phantom words" (Broersma & Cutler, 2008) into the process of word recognition. Therefore, teachers should help their students practice identifying phonemes, focusing as much as possible on the specific areas where students struggle.

Instructional Suggestions for Phonemes

- **Vowel/consonant homework**: Individual students can work with phonemes that are difficult for them to distinguish, being sure to practice with the sounds in a variety of phonetic contexts. For example, teachers can assign work with http://www.englishaccentcoach.com/.
- **Partial dictation**: Phrases or sentences are printed with a blank, and students fill in the missing part. The blanks can be word codas (e.g., "That woul_____ seem like"), pre-/suffixes (e.g., "In from larg____ distances"), or word middles (e.g., "That wouldn't s____m like"). It is preferable to concentrate on one position for the blanks in each short exercise.

- **Gating and prediction**: The teacher can stop the audio text after the first sound or syllable of a word and have students predict what the rest might be (e.g., the teacher says, "Food was raised lo-" and students talk to a partner about what word might follow, and then they discuss with class). This activity helps students practice applying top-down skills to make up for gaps or ambiguities in phoneme perception.

Unknown Words and Phrases

In designing the paused transcription materials, we tried to target only words that were known to participants to see if they would decode them in context. However, some unrecognized words and combinations of words may have been treated as unknown words by participants. We could infer that this had occurred when participants wrote letter combinations that did not correspond to any English word. Here are some examples of single words that appeared to be unrecognized.

Target word(s)	Transcriptions
Locally (Text 3 phrase 1—"food was raised locally")	Recoaly, Ridlly, Grobally, Recloliy, Quackly, Workly, Ulgerly, Bigulgle, Locanary, Revly
Distances (Text 3 phrase 6—"in from larger distances")	Siystances, Digness, Indecnit, Adegescence, Destious, Margien

Field (2004) discusses three strategies that learners might select when they encounter an unknown word in listening. They might take a phonological approach (attempt to transcribe the sounds they heard), a lexical approach (attempt to match approximately to a known word), or a zero approach (no transcription). Each of these approaches has advantages and disadvantages for learner comprehension. If learners take a strictly phonological approach, they recognize that a word has been missed and begin to learn the sounds of the new word, but they do not take the opportunity to apply schema and make an educated guess that will support their overall understanding of the text. If they choose a lexical approach, learners engage actively in trying to make meaning of the text, but they may forget the provisional nature of the lexical match and fail to revise their hypothesis when needed. Field (2004) found that his subjects selected a lexical approach more frequently than expected, and that lexical matches often were not semantically appropriate. Finally, a zero approach to new words can be seen as an instance in which the learner either did not recognize

that another word was spoken or could not remember anything about that word. These instances may occur when the listener "couldn't keep up" with the input, often resulting in a perception that the input was fast, regardless of its actual speed (see Bloomfield et al., 2015; Goh, 1999). Certainly, increased vocabulary knowledge can help improve students' listening comprehension, especially if the vocabulary is well known in its spoken form (Staehr, 2009; Van Zeeland, 2013; Van Zeeland & Schmitt, 2012). In fact, aural word recognition in context has been shown to correlate strongly with general listening comprehension scores (Matthews & Cheng, 2015).

One of the most difficult phrases for our participants to transcribe completely was "over an open fire." It was transcribed with 40% accuracy, compared to 66-90% accuracy for all other phrases in Text 1. Most participants wrote some words correctly, but very few transcribed both "over" and "open." The phrase is a common collocation, a formulaic expression that may be unfamiliar to many English language learners.

Text 1 phrase 7—"Instead of cooking **over an open fire**—"

Incorrect transcriptions	N	Analysis
Open fire	20	42 students transcribed "open" but not "over."
Cooking (in/with/ on) (an/ the/0) open fire	10	
Cooking (and/or) open (an/the/0) fire	7	
Open cooking fire	1	
Open (the/on/a) fire	4	
Cooking over fire	8	10 students transcribed "over" but not "open."
Stopping over the fire	1	
Over and over fire	1	
Cooking over an open fire	1	Only one student transcribed all four words correctly. Two additional students transcribed both "over" and "open," but missed the word "an."
Cooking over and open fire	1	
Cooking over open in fire	1	

The remaining 22 students omitted both "over" and "open" from their transcriptions.

- **Look up unknown words from listening:** Teachers can dictate sentences that include an unknown word. Students approximate the spelling to look up the word and compare meanings to the co-text (Sheppard, 2013). Field (2008b) suggests using proper nouns and even nonwords that conform to target language phonology in dictation and matching exercises.
- **Learn aural forms:** Teachers can easily incorporate aural forms into vocabulary study by having students listen to and repeat the words, identify syllables and stress, and hear the target words in the context of phrases and sentences.
- **Notice new expressions**: To encourage students to develop the habit of noticing and investigating word combinations, the teacher can pause after speaking or hearing a common idiom or collocation and asking students to discuss it. Dictation of common phrases or formulaic expression can also be a good method to raise student awareness.

Top-Down Fabrications

In some instances, participant transcriptions had little similarity to some or all of the four target words, either semantically or phonetically. Often these phrases were related to previous content from the audio text. In other cases, learners used the "lexical strategy" for unrecognized words as described above, selecting a familiar word with some similar characteristics. In these cases, the resulting phrase often made sense but did not fit semantically with the co-text. Finally, there were instances in which participants wrote words or phrases that did not match the target phonetically but had a similar meaning. These last instances can be seen as examples of successful application of top-down skills to repair small gaps in bottom-up processing. Two example phrases are analyzed below.

Text 3 phrase 6—"Food is shipped **in from larger distances**—"

Incorrect transcriptions	*Error analysis*
Food get different relationship	The topic of the text is "our relationship with food" and this phrase is also part of recent co-text.
Food relationship	

Close the relationship	The phrase "a distant rather than a close relationship" is part of less-recent co-text (about 1 minute ago).
Ship to logically places	Some sounds from "distances" are maintained or nearly maintained in "places," and "logically" has the same initial phoneme as the target word. The preposition is completely changed. The phrase does not make sense.
In from long distances	Long is a reasonable word for this context. The meaning is not changed, even though the participant did not write "larger." This can be seen as a successful semantic interpretation.

Text 3 phrase 1—"They were physically close to it and psychologically close to it. **Food was raised locally—**"

Incorrect transcriptions	*Error analysis*
Food will increase normally	Some of the sounds are maintained and some nearly maintained (e.g., /i/ for /e/ is a common mishearing), but different fairly sensible word choices are substituted. The phrase makes sense by itself but does not fit the co-text.
Food was grown rekoly The food was grown locally	Transcription of the third word substitutes a semantically sensible alternative for "raised"—in that sense it can be seen as successful. In one of the two instances, the last word was not recognized (although a number of phonemes are maintained).
The food was reason locally Food is look locally	Less-successful substitutions for the third word are seen here. In the first instance we see some matching phonemes, and in the second perhaps some effect of the following phonemes.
The good was great lovely The food was lovely	A different word with several similar sounds is substituted for the fourth word of the phrase. In the first example, a phonetically similar word is also substituted for "raised." In the second example, "raised" is omitted, leading to a phrase that makes sense by itself and could stretch to make sense with the co-text so far, but this interpretation will still add challenge to interpretation of following co-text.

Food was very (*n*=3)	This is a plausible beginning for a sentence in this context, and "very" does incorporate some phonemes from both of the words it replaces. The missed concept of "locally" will, however, add to the challenges of listening in the next sentences.

Applying top-down skills to guess in the face of inadequate decoding is a valuable strategy, but learners need to remember that guesses may need to be revised in light of further input. Misperception of words in a key sentence can lead some learners to maintain incorrect beliefs about the topic of a text even when further co-text makes it clear that something is wrong. Field (2008b) suggests that this may occur when learners do not trust their comprehension of later co-text enough to discard their investment in what they heard before, especially since they cannot go back and listen again. Thus teachers should encourage students to use top-down skills to make guesses but also remind students to revise those guesses as needed.

Instructional Suggestions for Top-Down Fabrications

- **Monitor comprehension**: Students must learn to check their understanding of the text-so-far for consistency with what they think they are understanding in the moment. Teachers can tell stories of their own misunderstandings or give think-aloud demonstrations to raise awareness of this point. Teachers can make a habit of asking, "How sure are you?," along with other comprehension questions, to develop in students the habit of assessing their own level of certainty.
- **Making and checking predictions**: A teacher can play the first part of an audio text, then ask students to make predictions about the topic and main ideas together with a partner or group, and then play some more of the text and ask students to discuss whether and in what ways their predictions were right or wrong. They can also discuss possible reasons for misunderstandings.
- **Metacognitive strategy instruction**: Teachers can follow Vandergrift and Goh's metacognitive pedagogical sequence (2012), in which learners are taught to (a) plan for listening, (b) monitor comprehension, (c) solve problems with comprehension, and (d) evaluate the outcome.

Using Paused Transcription in the Classroom

The process of examining student errors in paused transcriptions was enlightening to us as teachers, highlighting common errors and also giving insights into the misperceptions of individuals. It would likely be similarly enlightening for other classroom teachers to examine the patterns of error in paused transcriptions from their students. Using a short text, teachers could deliberately locate pauses to check students' perceptions of certain language features as a diagnostic tool. It may be even more useful (and more practical) for teachers to have students examine their own results from a paused transcription exercise. After the listening activity, teachers could post the full text and ask students to correct their own answers, with instructions to ignore spelling errors if the correct word was intended. They could then ask students to count specific kinds of errors, or simply instruct students to write and share a reflection on a few errors they found interesting, speculating about why they made those mistakes.

We believe that classroom activities involving analysis of paused transcription exercises can help teachers and students better understand the challenges of L2 listening and provide guidance for classroom instruction to improve listening skills. We also believe that such exercises can help develop an attitude of curiosity about errors that can facilitate student engagement and reduce listener anxiety, resulting in a more effective listening classroom.

Conclusions

This study suggests that even known words (or words presumed to be known—see the discussion of limitations below) often are not successfully decoded by intermediate-level language learners. These learners are more likely to decode known words when they are part of a less challenging text. When words drawn from the same list are part of a more challenging aural text, they are less successfully decoded. Content words are decoded more successfully than function words, a finding that confirms results of previous studies. Finally, faster phrases are not necessarily harder to decode, in spite of students' perceptions about speed and listening challenges (Bloomfield et al., 2011; Goh, 1999; Renandya & Farrell, 2011).

The paused transcription methodology used in this study can provide useful information about what individual students perceive when they listen. We recommend that teachers and students employ brief paused transcription exercises in the classroom to analyze listening perception for strengths and weaknesses, raise awareness, and possibly guide instruction. Teachers can choose a short, level-appropriate audio recording and insert 15-second pauses at the end of

several phrases. There is no need to space the pauses equally—varied intervals are preferred. If inserting pauses in the recording is a challenge, the teacher can simply plan locations to pause playback at the ends of phrases. Students listen to the recording, and in each pause write the last phrase (4-5 words) that was heard. Finally, the resulting written phrases are compared to a complete transcript of the audio recording. Teachers can conduct a simple analysis of student results to decide what kinds of activities would be helpful—for example, by checking for a few common categories of errors. Students can analyze their own results to build awareness of their strengths and weaknesses and to report their analysis to the teacher and receive advice.

This study had several limitations. First, we presumed that all research participants were familiar with the 1,000 most common words of English. While this probably is mostly true, word knowledge does vary, even among the most common words. For future paused transcription studies that target known words, this knowledge should be explicitly tested in a session after the paused transcription session. The vocabulary test should target auditory knowledge, not just familiarity with words in their written form. Second, we do not know how well participants understood the overall message of the three audio texts used in this study. It would be valuable for future studies on this topic to include an assessment of overall test comprehension, perhaps with a control group who did not do paused transcription, so we can get a better idea of how the paused transcription methodology might interact with listening processes. Finally, it would have been interesting to include a measure of participant confidence for each phrase transcribed. In this study, we cannot distinguish between errors that are guesses and errors that are strongly believed by the participant. Suggestions for interventions could be different in these two cases.

In our discussion, we have proposed a variety of activities to help students improve specific listening skills. Some of these activities are drawn from the literature, while others are our ideas. More research is needed on effectiveness of these specific interventions to improve listening subskills. In the meantime, we suggest only that teachers try them out and watch carefully for improvements in student listening.

Authors

Beth Sheppard teaches and develops curriculum for ESL listening and speaking at the University of Oregon and is involved in teacher training.

Brian Butler teaches academic reading and writing for international students at the University of Oregon and uses experimental research methods to explore and explain the functions of the English article system.

References

Bloomfield, A., Wayland, S., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2011). *What makes listening difficult? Factors affecting second language listening comprehension.* College Park: University of Maryland Center for Advanced Study of Language (CASL).

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*(1), 22-34.

Brunfaut, T., & Revesz, A. (2015). The role of task and listener characteristics in second language listening. *TESOL Quarterly, 49*(1), 141-168.

Cobb, T. (n.d.) *Web vocabprofile* [an adaptation of Heatley, Nation, & Coxhead. (2002). *Range*]. Retrieved from http://www.lextutor.ca/vp/

Cross, J. (2009). Diagnosing the process, text, and intrusion problems responsible for L2 listeners' decoding errors. *Asian EFL* Journal, 11(2), 31-53.

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

Estes, R. (2014). *Lexical segmentation in L2 Spanish listening* (Doctoral dissertation). University of California, Davis.

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System, 32*(3), 363-377.

Field, J. (2008a). The L2 listener: Type or individual. In *Working papers in English and applied linguistics in honour of Gillian Brown* (pp. 11-32). Cambridge, England: RCEAL.

Field, J. (2008b). Revising segmentation hypotheses in first and second language listening. *System, 36*(1), 35-51.

Field, J. (2008c). Bricks or mortar: Which part of the input does a second language listener rely on? *TESOL Quarterly, 42*(3), 411-432.

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

Field, J. (2011). Into the mind of the academic listener. *Journal of English for Academic Purposes, 10*(2), 102-112.

Goh, C. (1999). How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics, 4*(1), 17-40.

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*(1), 55-75.

Herron, D., & Bates, E. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. *Journal of Memory and Language, 32*(2), 217-239.

Liu, N-f. (2002). *Processing problems in L2 listening comprehension of university students in Hong Kong* (Unpublished doctoral disserta-

tion). Hong Kong Polytechnic University.

Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System, 52,* 1-13.

Renandya, W., & Farrell, T. (2011). Teacher, the tape is too fast! Extensive listening in ELT. *ELT Journal, 65*(1), 52-59.

Revesz, A., & Brunfaut, T. (2012). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition (SSLA), 35*(1), 31-65.

Rost, M. (2016). *Teaching and researching listening* (3rd ed.). Abingdon, England: Routledge.

Siegel, J. (2014). Exploring L2 listening instruction: Examinations of practice. *ELT Journal, 68*(1), 22-30.

Siegel, J., & Siegel, A. (2015). Getting to the bottom of L2 listening instruction: Making a case for bottom-up activities. *SSLLT, 5*(4). Retrieved from http://pressto.amu.edu.pl/index.php/ssllt/article/view/4322/4386

Sheppard, B. (2013). Defining unknown words from listening. *ORTESOL Journal, 30,* 33-34.

Staehr, L. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition (SSLA), 31*(4), 577-607.

Tsui, A. B. M., & Fullilove, J. (1998). Bottom-up or top-down as a discriminator of L2 listening performance. *Applied Linguistics, 19*(4), 432-451.

Van Zeeland, H. (2013). L2 vocabulary knowledge in and out of context: Is it the same for reading and listening? *Australian Review of Applied Linguistics, 36*(1), 52-70.

Van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24,* 3-25.

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* New York, NY: Routledge.

West, M. (1953). *A general service list of English words.* London, England: Longman, Green.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing, 33*(4), 497-528.

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System, 36,* 69-84.

Appendix A
Target Phrases

	Target phrase	# content words	# function words	Syllables/ second
Text 1	don't always notice them	3	1	3.628
	and make new friends	3	1	4.540
	most of the dances	2	2	3.918
	never done for money	3	1	7.075
	used to be played	1	3	4.449
	might see a woman	2	2	4.122
	over an open fire	2	2	5.076
	still a special time	3	1	3.381
	women wore long dresses	4	0	3.363
	part of our lives	2	2	3.902
	think is beautiful today	3	1	4.079
	like in the future	1	3	4.575
Text 2	to have an opinion	2	2	6.263
	direction of their lives	2	2	4.323
	women must now decide	3	1	4.199
	to stay at home	2	2	4.188
	it is no longer	2	2	4.878
	to take into account	2	2	4.598
	We knew that men	2	2	4.624
	outside of the home	2	2	4.255
	to be about equal	2	2	5.391
	study that was done	2	2	4.621
	women in both groups	3	1	4.990
	let me repeat that	2	2	5.519
Text 3	food was raised locally	3	1	3.417
	person or one step	3	1	3.455
	True in earlier days	3	1	4.648
	than a close relationship	2	2	4.593
	you can see that	1	3	5.900
	in from larger distances	2	2	4.645
	where it came from	1	3	4.154
	that story is something	2	2	5.045
	you probably know this	2	2	5.618
	later in the class	2	2	5.464
	that wouldn't seem like	2	2	6.410
	go across the room	2	2	6.024
Total		80	64	

Table 1
Descriptive Statistics for Transcription Accuracy
by Student Level and Text Level

Variable	Midlevel students (n =144)		Upper-level students (n = 144)		Total (n = 288)	
	M	SD	M	SD	M	SD
Text level 1	0.66	0.22	0.84	0.19	0.75	0.22
Text level 2	0.53	0.23	0.75	0.20	0.64	0.24
Text level 3	0.42	0.28	0.61	0.27	0.51	0.29
Total	0.54	0.26	0.73	0.24	0.63	0.27

Table 2
Student Level by Text-Level Analysis of Variance Summary Table

Source	df	SS	MS	F
Student level	1	2.68	2.68	48.80*
Text level	2	2.72	1.36	24.76*
Student level * Text level	2	0.02	0.01	0.18
Error	30	1668.00	55.60	
Total	35	5793.00		

Note. *$p < .05$.