Second Language Learners' Speech Perception

Bridget A. Goodman
Nazarbayev University Graduate School of Education
Manami Suzuki
Hosei University

Reference Data:

Goodman, B. A., & Suzuki, M. (2017). Second language learners' speech perception. In P. Clements, A. Krause, & H. Brown (Eds.), *Transformation in language education*. Tokyo: JALT.

In this study we examined factors influencing Japanese speakers' discrimination of *can* and *can't* in a spoken English sentence. Drawing on theories of phonemes and L2 speech perception, we prepared 36 sentences containing *can* or *can't* that were read by 7 native speakers of American English. Japanese university students (*N* = 148) listened to a recording of 252 sentences in random order and marked which word they perceived. Bivariate regression analyses showed that subsequent sound, allophones of /t/, and sentence position were significant predictors of identification. Multiple regression analyses of significant phonemic factors and additional nonphonemic factors (the TOEIC IP score and the speaker) indicate that articulation of the /t/ in *can't*, coarticulatory effects, and individual speaker effects are the most significant predictors of comprehension.

本研究では148人の日本人の大学生を対象に、センテンスレベルのcanとcan'tの聞き分けの難しさの要因について音素の理論に基づき分析を実施した。7人のアメリカ人のネィティブスピーカーによって音読されたcanあるいはcan'tを含む252のセンテンスを利用し、音素と非音素の要因（TOEIC　IPテストの結果とスピーカーによる違いの要因など）に関する重回帰分析を実施した。それらの統計的な分析により、can'tの　/t/　の調音、調音結合の影響、スピーカーの相違の3要因が最も有意な予測変数であることを示唆する結果を得た。

The words *can* and *can't* are commonly used in English daily conversation but can be challenging for L2 learners to correctly perceive. ESL pronunciation textbook authors observe that, in American English, *can* is unstressed with a schwa vowel (/kən/) but *can't* is stressed and has a clear /æ/ (Grant, 2001; Miller, 2006). Miller acknowledged, however, that "even native speakers get confused at times" (p. 104) because speaking pace, vowel length, and the relative formality of context can lead to variability in the clarity of the sound of *can* or *can't*. Grant (2001) added that English speakers do not rely on the presence or absence of /t/ for distinguishing *can* from *can't*; however, there was no elaboration regarding any phonetic properties English speakers use instead. It is possible that listeners can rely on redundant syntax- (sentence) or discourse-level contextual cues to differentiate *can* from *can't*, for example, "Yes, we can" or "I'm sorry, he can't come." If these redundant cues were present consistently and reliably, however, there would not be confusion for English language learners let alone native speakers.

The aim of the current study was to identify the relationship between accurate perception of *can* or *can't* by Japanese L2 learners of English and variations in manners or contexts of articulation of these words by native speakers of American English. In addition, we wish to show which variables of articulation are the strongest predictors of difficulty perceiving *can* or *can't*. The research questions for this study are

RQ1. Which phonetic patterns or phonological contexts of *can* or *can't* (as uttered by native speakers of American English) are significantly linked with accurate auditory discrimination of *can* and *can't* by learners of English?

RQ2. Which additional features of speakers and listeners are associated with a Japanese English learner's ability to discriminate *can* from *can't* as spoken in sentences by native speakers of English?

Literature Review

The current study was primarily designed to control for and test the recognition of *can* or *can't* according to Rost's (2002) categorization of factors of speech recognition: (a) different rates of speaking; (b) different sounds preceding or following a particular word of interest (coarticulatory effects); (c) different pronunciation, due to regional native speaker (NS) or nonnative speaker (NNS) accents; (d) different speakers' vocal tract configurations leading to spectrum differences; and (e) incomplete utterances, in which sounds or whole words are omitted (p. 76).

Previous research has also connected L2 speech recognition with markedness and L1 transfer (Eckman, 1991; Eckman & Iverson, 1994; Ellis, 2008; Romaine, 2003). The first language of the current study's participants, Japanese, is a mora-timed language. A mora is a unit in phonology that determines syllable weight. Stress and timing in languages are designated by syllable weight. In Japanese, a mora is a V (vowel) or CV (consonant + vowel) syllable that takes one unit of timing. Native Japanese speakers use vowel sounds as a marker of syllables. Each mora takes almost the same duration of time to pronounce. For example, the Japanese word *dekiru* (can) has three morae, /de/ki/ru/, each containing one vowel. English, on the other hand, is a stress-timed language. Individual syllables in words receive primary or secondary stress. In sentences, words receive stress based on the part of speech, location in the sentence or sentence clause, and desired emotional emphasis. Based on parts of speech, *can* is primarily unstressed as an auxiliary verb and *can't* is primarily stressed as a form of negation. The differences between the two languages might influence Japanese English learners' perception of *can* or *can't* in a sentence.

In addition, English has an exceptionally wide range of consonant-vowel syllable formations, including two- and three-consonant clusters. The word *street*, for example, is one syllable with a CCCVC pattern. In contrast, Japanese permits only CV, V, and CVC; it has no consonant clusters. Therefore, based on L1 transfer theory and empirical observation (Swan & Smith, 1987), we predicted it would be more difficult for Japanese learners to recognize CVCC (e.g., *can't*) than CVCV (e.g., *can* preceding a vowel sound).

Additional research on speech perception is mixed about whether it is the learner's phonetic background or the speakers' articulation that influences perception. Flege and MacKay (2004) found that learners of English may use different linguistic cues than native speakers to differentiate phonemes, causing misidentification. Strange et al. (1998) investigated Japanese speakers' perception of English vowels in words and short sentences across multiple English speakers and concluded that "phonetic variation in the realization of vowels has a significant influence on cross-language perceptual similarities" (p. 327).

Nonphonemic Factors in Perception

Identification of *can* or *can't* may also be correlated with demographic factors such as the listener's age, sex, years of studying English, and English proficiency level (Ellis, 2008). Based on previous studies of the order of acquisition of English morphemes (see Ellis, 2008, for a review), it is predicted in the current study that high proficiency L2 learners will correctly identify *can* or *can't* at a higher frequency than will low proficiency learners. Moreover, to our knowledge no L2 research has examined the relationship between L2 learners' speech recognition of individual words (e.g., *can* and *can't*) in a sentence and their proficiency levels. A few researchers have investigated the interactional effect between repetition and proficiency levels in L2 listening tests (Chang & Read, 2006; Sakai, 2009), and others have explored the relationship between perception and the starting age or length of exposure learners have had to the target language (Flege, Bohn, & Jang, 1997; Flege & Liu, 2001).

Method

The first author wrote 36 sentences that contained one token (instance) of *can* or *can't*, without redundant cues such as *yes* or *no*, and that allowed for the articulation of *can* and *can't* multiple times in initial, medial, and final sentence positions. In the initial and medial positions, each instance of *can* or *can't* was followed by a word starting with one of five sounds: a vowel /i/, /ɛ/, or /o/, or a fricative, /s/ or /t/. The sentences can be found in Appendix A.

Seven native English-speaking students were recruited on a U.S. university campus in the mid-Atlantic region to read the 36 sentences. Three of the speakers were female and four were male. Each speaker originated from a different dialect region of the United States as classified by Labov, Ash, and Boberg (2006): Inland North, Mid-Atlantic, North Central, New York City, South, Western New England, and Western Pennsylvania.

The 252 recorded sentences (36 sentences x 7 readers) were screened for reading mistakes and false starts. Seven sentences were removed and the remaining 245 sentences were placed in random order and copied to a CD.

The next step was implementing a listening test for Japanese learners of English using these speech samples. The participants were 148 Japanese students (114 female, 34 male) in compulsory English courses at three universities in Japan. Their specialties were literature ($n = 41$, X Women's University), modern liberal arts ($n = 46$, Y Women's University), foreign language ($n = 18$, Z University), law ($n = 20$, Z University), and economics ($n = 23$, *Z University*). All students were over 18; the mean age was 18.6 ($SD = 0.7$). Their length of learning English in years was similar ($M = 7.5$, $SD = 2.1$).

The listening tests were conducted in each participant's class in the fourth or fifth week of November 2009. Test sheets were distributed to students (see Appendix B). Students were asked to circle or mark the following choices: *can, can't, or I don't know* after hearing each sentence one time. The same CD was played only once in the test. The total length of the test was 20 minutes.

Data Analysis Procedures

After all listening tests were completed, two measures were applied to evaluate the reliability and validity of the test instrument. First, we used Cronbach's alpha to measure internal consistency of each of the 245 items in the test. A high coefficient of .811 was obtained. In addition, we calculated the Cronbach's alpha score for the 36 sentences. The score for the sentences was lower but still reliable (.732). In neither analysis were there items that would impact the overall alpha score if they were removed.

Second, to determine whether there is a ceiling effect of the test instrument, 10 Americans (7 women, 3 men) studying at the graduate level in English or linguistics listened to the recordings. The mean number of sentences correctly identified was 212.2 (86.61%, *SD* 5.5). The breakdown was similar for *can* (86%, SD 3.7) and *can't* (87%, SD 4.4). These data do not indicate a ceiling effect for native speakers in distinguishing *can* from *can't*. Moreover, the percentages of correct identification are lower and more varied (as evidenced by standard deviation) in the results from the 148 Japanese students.

With the validity and reliability of the instrument confirmed, each subject's (*N* = 148) answers were entered into an Excel database as 1 (correct) or 0 (incorrect or not sure) for each sentence. Variables were created for the number and percent of sentences identified correctly by word (*can* or *can't*) and phonetic category (initial position, medial position, final position, glottal stop deletion of /t/, articulation of /t/, and vowel or consonant as a subsequent sound). To determine the allophone of /t/ in each utterance, two raters trained in phonetics independently identified allophones of /t/ using Eddington's (2007) classification through visual examination of waveform and spectrogram output in Praat, a phonetics software. The interrater reliability for the coding of /t/ as an aspirated /t/, glottal /t/, or deletion, as defined by the Kappa value of the ratings, was .793. As any Kappa value above .70 is generally considered reliable, this is sufficient for further analysis.

These variables along with the subjects' demographic data (without identifiers that could be traced back to the person) were entered into SPSS. Reports of frequencies, means, and standard deviations were run in SPSS for each of these categories. Bivariate regressions were run to determine if the differences in the subjects' correct identification were significant based on pairs of phonemic conditions—subsequent sound (vowel or consonant), allophone of /t/ (articulated or deleted), and sentence position (initial/medial or final). For the differences in correct identification of *can* or *can't* according to speakers, paired *t* tests were run to determine the significance. Finally, the statistically significant factors were subjected to a multiple regression analysis to determine which factors were stronger predictors of correct identification of *can* or *can't*.

Results and Discussion

Descriptive statistics of listener data showed that subsequent sound, sentence position, and manner of articulation of /t/ in *can't* all potentially impact participants' correct identification of *can* or *can't*. On average, participants less frequently identified *can* when it was followed by a consonant than by a vowel. For *can't*, however, participants less frequently identified the word correctly when it was followed by a vowel than by a consonant. Participants more frequently identified *can* correctly when it was in the final position than when it was in the initial or medial positions. Table 1 shows the average identification rates by subsequent sound and sentence position.

Table 1. Correct Identification by Sentence Context (*N* = 148)

Sentence context	*M*	*SD*	*n*
Subsequent sound			
can + vowel	72	13.1	26
can + consonant	61	11.4	56
can't + vowel	49	13.1	27
can't + consonant	62	13.8	53
Sentence position			
can initial or medial	57	12.1	82
can final	63	11.2	42

Note. n = the total number of tokens for that variable.

As for allophones of /t/, participants correctly identified *can't* nearly twice as frequently when the /t/ was aspirated than when it was a glottalized or deleted. Further analysis of /t/ allophone categorization revealed that speakers more frequently glottalized or deleted the /t/ before a vowel sound compared with other possibilities for a following sound. Table 2 shows the average identification rates by allophone of /t/.

Table 2. Correct Identification by Allophone of /t/ (*N* = 148)

	Rater 1			Rater 2		
/t/ allophone	*M*	*SD*	*n*	*M*	*SD*	*n*
Aspirated	81	6.9	42	78	7.8	52
Glottal stop or deletion	48	10.6	60	44	10.6	62
Aspirated prevowel	78	19.7	8	70	16.9	16
Glottal stop or deletion prevowel	59	14.2	36	57	14.6	31

Note. n = the total number of tokens for that variable.

Bivariate regressions confirmed that these patterns in identification were significant for subsequent sound and allophones of /t/ (see Appendix C). The highest correlation coefficients were obtained for *can* based on subsequent sound ($R^2 = .316$) ($p < .001$). For *can't*, subsequent sound ($R^2 = .302$, $p < .001$) and articulation of /t/ prevowel ($R^2 = .388$, $p < .001$) had the highest correlations. Bivariate regression of *can't* by manner of articulation without regard to the prevowel condition (aspirated x glottal stop/deletion) was also significantly and positively correlated, but with a lower predictive value. When bivariate regression was run on *can* (initial or medial) x *can* (final), the result was significant but the predictive value was lowest of all the variables.

There are two conclusions that can be drawn from these results. First, it appears that when the /t/ in *can't* is glottalized or deleted, *can't* is being confused with *can*. This confirms Rost's (2002) theory that speech recognition is more difficult when a sound is incomplete or deleted. Second, when *can* is followed by a consonant, the consonant is being perceived as the final sound of *can* (i.e., *can't*) rather than the initial sound of the next word. This would account for why *can* was more frequently identified at the end of a sentence than at the beginning despite the fact that the vowel articulation of *can* and *can't* in this position were potentially ambiguous.

Nonphonemic Factors in Identification

As noted previously, there was little variability in participants' age or years of English study. Therefore, it is not surprising that when we ran bivariate regressions between these factors and identification rates, the results were not significant. Overall proficiency measures, however, were shown to be correlated with the test results. We ran regressions only with the data from students who reported the TOEIC IP test scores ($n = 61$, $M = 439.0$, $SD = 99.1$). All but eight of the students in this subset were from Z University, where they were required to take the TOEIC IP test. The participants' TOEIC IP scores were significantly related to the percentage of participants' correct identification of *can* ($p <. 0.01$), the percentage of participants' correct identification of *can't* ($p <. 0.5$) and the percentage of participants' correct identification of all items ($p <. 0.01$). See Appendix D.

As for speaker variation, we analyzed the frequency of items correctly identified according to the speaker and found a relatively wide range of identification rates (from 52% for Speaker 6 to 71% for Speaker 7). Paired *t* tests conducted for all 21 possible speaker pair combinations revealed that differences among all speaker pairs were significant except for the Speaker 1-Speaker 7 pair and the Speaker 3-Speaker 6 pair. Effect sizes for the significant differences ranged from .404 (Speaker 1-Speaker 6) to .695 (Speaker 2-Speaker 7). No patterns emerged to further differentiate speakers by sex, location, usage of allophones of /t/, or vowel length.

Multiple Regression

Because multiple variables were significant in predicting participants' ability to distinguish *can* from *can't*, multiple regression analysis was conducted for each word. For *can*, four variables were statistically significant: subsequent consonant ($p = .000$), subsequent vowel ($p = .032$), Speaker 2 ($p = .002$), and Speaker 3 ($p = .006$). For *can't*, the strongest predictor of correct identification was overall /t/ aspiration or deletion ($p = <.001$ for each). The subsequent sound was not a significant predictor when it was a consonant, but was significant when it was a vowel ($p < .05$) with a low predictive value (.053). Speaker 7 was also significant but with a very low predictive value ($p = -.073$). See Appendices E and F for detailed regression data.

Effect of L1 on Perception of L2 Sounds

As reported earlier, the identification rate of *can* + vowel (CVCV) was higher than that of *can* + consonant (CVCC), but the identification rate of *can't* + vowel (CVCCV) was lower than that of *can't* + consonant (CVCCC). Therefore, the effect of L1 transfer with regard to the syllable types was observed only in the case of *can* + consonant; listeners' difficulty identifying *can't* + vowel is considered to be attributable to native speaker variation that conflated *can't* with *can*. The difficulty identifying *can't* + vowel is consistent with previous research on salience including acoustic or phonetic salience and SLA (see DeKeyser, 2005).

Speaker Variables

The current study showed that speakers were an important factor for L2 learners' perception of *can* and *can't*; there was variation in identification rates that was significantly different and robust for all but two pairs of speakers. Unfortunately, the current study's analyses of which speaker factors influenced students' sound recognition (e.g., sex, American English variation, age, vowel length, recording quality, or relative pitch) were inconclusive. Further research in this area is needed. Moreover, speakers' variation within the framework of world Englishes (Bolton, 2004) should be considered in future research because the current study only focused on American English.

L2 Proficiency Level and L2 Perception

Although the TOEIC IP score was not a significant predictor of perception for *can* or *can't* in multiple regression analysis, the finding of an overall correlation proved that higher proficiency level L2 learners could recognize the relative features of phonemes better than lower level proficiency level learners. Although participants had studied English for a similar number of years (M = 7.5 years, SD = 2.1), higher proficiency students may have been more exposed to various examples of the target language phonological and phonotactic structures, as de Jong, Silbert, and Park (2009) suggested based on their study of L2 consonant identification. Moreover, the test we developed shows internal reliability with the TOEIC IP test, which is considered to have a high reliability with the TOEIC. This suggests that being able to distinguish *can* and *can't* is a microskill (Brown & Abeywickrama, 2010) of listening comprehension.

Limitations

In this study we only examined L2 learners' perception of vowels and consonants following auxiliary verbs *can* and *can't*, using a quasiexperimental approach. Because our focus was at the word level, we did not examine L2 learners' ability to identify a wide range of L2 consonants and, therefore, cannot suggest a model of L2 learners' skills in different syllable constructions and prosodic positions (see de Jong, Silbert, & Park 2009 for such analysis). All participants were Japanese adult EFL learners. Other contexts (e.g., ESL, learners' age, natural or classroom learners) and other languages (both L1s and L2s) should be considered to draw a robust conclusion. We did not consider whether participants had become accustomed to British or American English or their dialects. We did not examine the effect of phonology instruction either. Future research should focus on the relative effectiveness of rule-based interventions (e.g., providing students with new explicit rules for distinguishing *can* from *can't*) versus increased input for improving accuracy in distinguishing *can* from *can't*.

Conclusion

Despite these limitations, this study can provide several pedagogical suggestions for SLA, particularly for learning to listen. Although teaching students to pay attention to vowel length may still be an effective starting point for discriminating *can* and *can't* in both reception and production, the data from this study suggest that being able to identify the ending of one word and the beginning of the subsequent word may be a more important factor for understanding American English as well as being aware of when sounds are different from the expected allophone. This is especially true for students whose native language or dialect has a different set of patterns for the clustering of vowels and consonants (e.g., CV, VC, V, CCV, CV, and CVC). A student who demonstrates an ability to distinguish these sounds can be predicted to have greater overall comprehension of English. Thus, the data of the current study demonstrate that students with higher proficiency can distinguish subphonemic differences (e.g., allophones) at sentence level more accurately.

The results further indicate that second language teachers should give different instructional input according to their students' levels. For example, learners at a low proficiency level may be asked to read (aloud) or listen to sentences that draw students' attention to the words *can* and *can't* and the difference in pronunciation—a typical approach in explicit focus on form instruction (Doughty & Williams, 1998; Ellis, Basturkmen, & Loewen, 2002; Lightbown & Spada, 1990). Furthermore, because individual speaker variations also proved to be significant, second or foreign language teachers should be conscious of their own speech in their classroom when articulating potentially ambiguous words (e.g., *can* or *can't*) to beginners (e.g., during task instructions) to avoid learners' confusion. As students progress in their studies of the target language, raising awareness of potential variations in speech, as well as the means of negotiating for understanding in communicative language teaching, is necessary.

At the same time, students may benefit from having extensive input from a range of sources to help with comprehension of these and other sounds. Previous studies have shown that perception of target consonants and vowels can be improved through "training" in which L2 learners listen to a large number of minimal-pair words containing the target sounds (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Nishi & Kewley-Port, 2005) or a perception training protocol of a large set of vowels (Nishi & Kewley-Port, 2007, 2008). More research is needed on the relative effectiveness of these strategies in regards to *can* and *can't* versus other types of input in laboratory and classroom settings. We hope that our study can shed light on L2 learners' listening perception and its effective instruction.

Acknowledgments

The authors are grateful to the following people: the late Dr. Teresa Pica for her support at the initial stages of this research project, Dr. Erling Boe and Clare Waterman for their guidance on statistical analyses, Dr. Joanna Labov for her review of an earlier draft of this manuscript, and Dr. Yuko Goto Butler.

Bio Data

Manami Suzuki is a professor in the faculty of Business Administration at Hosei University, Tokyo, Japan. Her research focuses on second language writing, motivation, and intercultural communication in business.

Bridget Goodman is an assistant professor of multilingual education at Nazarbayev University Graduate School of Education in Astana, Kazakhstan. Her research interests include English as a medium of instruction (EMI) policy and practice in multilingual contexts, cross-linguistic transfer, and second language pronunciation teaching.

References

Bolton, K. (2004). World Englishes. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics*. Oxford, England: Blackwell.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101,* 2299-2310.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). New York, NY: Pearson.

Chang, A. C.-S., & Read, J. (2006). The effects of listening support on the listening performance of EFL learners. *TESOL Quarterly, 40*, 375-397. https://doi.org/10.2307/40264527

de Jong, K. J., Silbert, N. H., & Park, H. (2009). Generalization across segments in second language consonant identification. *Language Learning, 59,* 1-31. https://doi.org/10.1111/j.1467-9922.2009.00499.x

DeKeyser, R.M. (2005). What makes learning second-language grammar difficult?: A review of the issue. *Language Learning*, 55, 1-25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

Doughty, C., & Williams, J. (Eds.). (1998). *Focus on form in classroom second language acquisition*. Cambridge, England: Cambridge University Press.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13,* 23-41. https://doi.org/10.1017/S0272263100009700

Eckman, F. R., & Iverson, G. K. (1994). Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In M. Yavas (Ed.), *First and second language phonology* (pp. 251-265). San Diego, CA: Singular Publishing Group.

Eddington, D. (2007). Flaps and other variants of /t/ in American English: Allophonic distribution without constraints, rules, or abstractions. *Cognitive Linguistics, 18,* 23-46. https://doi.org/10.1515/COG.2007.002

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, England: Oxford University Press.

Ellis, R., Basturkmen, H., & Loewen, S. (2002). Doing focus on form. *System, 30,* 419-432. https://doi.org/10.1016/S0346-251X(02)00047-7

Flege, J. E., Bohn, O-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25,* 437-470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23,* 527-552.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1-34. https://doi.org/10.1017/S0272263104026117

Grant, L. (2001). *Well said* (2nd ed.). Boston, MA: Heinle and Heinle.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English* [Online book]. Berlin, Germany: Mouton de Gruyter. Retrieved from http://www.atlas.mouton-content.com/

Lightbown, P., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effect on second language learning. *Studies in Second Language Acquisition, 12,* 429-448. https://doi.org/10.1017/S0272263100009517

Miller, S. F. (2006). *Targeting pronunciation* (2nd ed.). Boston, MA: Houghton Mifflin.

Nishi, K., & Kewley-Port, D. (2005). Second language vowel training using vowel subsets: Order of training and choice of contrasts. *Journal of the Acoustical Society of America, 118*, 1900. https://doi.org/10.1121/1.4779959

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research, 50,* 1496-1509. https://doi.org/10.1121/1.4786064

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research, 51,* 1480-1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

Romaine, S. (2003). Variation. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 409-435). Oxford, England: Blackwell.

Rost, M. (2002). *Teaching and researching listening.* Essex, England: Pearson Education.

Sakai, H. (2009). Effect of repetition of exposure and proficiency level in L2 listening tests. *TESOL Quarterly, 42,* 360-371. https://doi.org/10.1002/j.1545-7249.2009.tb00179.x

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics, 26,* 311-344. https://doi.org/10.1006/jpho.1998.0078

Swan, M., and Smith, B. (Eds.). (1987). *Learner English: A teacher's guide to interference and other problems*. Cambridge, England: Cambridge University Press.

Appendix A
Sentence List

1. Can Eve play the guitar?
2. I think she can.
3. Can't Eve sing?
4. I heard she can't.
5. Can Eddie sing?
6. I know he can.
7. Can't Eddie dance?
8. Joe said he can't.
9. He can order it tomorrow.
10. They can eat at noon.
11. We can't order it today.
12. You can't eat yet.
13. Can Sue play the guitar?
14. I heard she can.
15. Can't Sue sing?
16. I know she can't.
17. Can Sam sing?
18. I guess he can.
19. Can't Sam dance?
20. Ben said he can't.
21. He can set the table.
22. They can see the city.
23. We can't see the board.
24. He can't say why.
25. Can Tom play the guitar?
26. I hope he can.
27. Can't Tom sing?
28. I'm sure he can't.
29. Can Tina sing?
30. Bob said she can.
31. Can't Tina dance?
32. I know she can't.
33. He can tell us today.
34. They can tell us tomorrow.
35. We can't tell you why.
36. He can't tell us now.

Appendix B

Answer Sheet Excerpt

ID Number

Listen to each sentence and circle the word you hear. If you are not sure, circle "Don't know."

英文を聞いて、それぞれの英文の中で、canと聞こえたらcanに、can't と聞こえたらcan't にマルをつけて下さい。どちらかわからなかった時には、don't know（わからない）にマルをつけて下さい。

1. can can't don't know
2. can can't don't know
3. can can't don't know
4. can can't don't know
5. can can't don't know

Appendix C

Bivariate Regression by Sentence Context and Phonemic Feature (N = 148)

Phonetic condition	R^2
Subsequent sound	
can + vowel x *can* + consonant	.316***
can't + vowel x *can't* + consonant	.302***
Allophones of /t/ =	
/t/ aspirated prevowel x /t/ glottal stop/deletion prevowel—Rater 2	.388***
/t/ aspirated prevowel x /t/ glottal stop/deletion prevowel—Rater 1	.256***
/t/ aspirated x /t/ glottal stop/deletion—Rater 2	.170***
/t/ aspirated x /t/ glottal stop/deletion—Rater 1	.131***
Sentence position	
can (initial) x *can* (medial or final)	.021
can (initial or medial) x *can* (final)	.048**

Notes. $^{**}p < .01$. $^{***}p < .001$.

Appendix D

Regression of Items Correct by the TOEIC IP Score

Measure	1	2	3	4
Percent of *can* items correct	---			
Percent of *can't* items correct	.021	---		
Percent of all items correct	.800**	.617**	---	
TOEIC IP Score	.562**	.316*	.627**	---

Note. $^{*}p < .05$. $^{**}p < .01$.

Appendix E

Regression Analysis Summary for Variables Predicting Percent of Can Tokens Correctly Identified, With the TOEIC IP Score (n = 61)

Variable	*B*	*SE B*	β	*t*	*p*
can + consonant percent correct	.754	.161	.377	4.687	.000
can + vowel percent correct	.596	.271	.162	2.203	.032
Speaker 1	-.239	.254	-.070	-.940	.352
Speaker 2	.824	.253	.284	3.253	.002
Speaker 3	1.033	.359	.227	2.879	.006
Speaker 4	.088	.291	.024	.301	.765
Speaker 5	.304	.296	.077	1.024	.311
Speaker 6	.147	.303	.036	.487	.629
Speaker 7	.345	.338	.084	1.021	.312
TOEIC IP score	.003	.009	.023	.295	.769

Appendix F

Regression Analysis Summary for Variables Predicting Percent of Can't Tokens Correctly Identified, With the TOEIC Score (n = 61)

Variable	*B*	*SE B*	β	*t*	*P*
can't + consonant percent correct	.132	.105	.226	1.257	.215
can't + vowel percent correct	.053	.025	.097	2.076	.043
can't aspirated percent correct	.364	.047	.277	7.695	.000
can't deleted or glottal stop percent correct	.461	.050	.618	9.277	.000
Speaker 1	.049	.028	.063	1.772	.083
Speaker 2	.011	.027	.017	.421	.675
Speaker 3	.012	.035	.012	.334	.740
Speaker 4	-.001	.033	-.001	-.023	.981
Speaker 5	.002	.033	.002	.065	.949
Speaker 6	.019	.028	.022	.680	.500
Speaker 7	-.073	.035	-.078	-2.101	.041
TOEIC IP Score	.000	.000	-.007	-.166	.869
can't aspirated prevowel percent correct	-.001	.035	-.003	-.042	.967
can't deleted or glottal stop prevowel	-.040	.071	-.074	-.558	.580