Perception of silent-center syllables by native and non-native English speakers^{a)}

Catherine L. Rogers^{b)} and Alexandra S. Lopez

Department of Communication Sciences and Disorders, University of South Florida, 4202 E. Fowler Avenue, PCD1017, Tampa, Florida 33620

(Received 8 May 2007; revised 23 April 2008; accepted 7 May 2008)

The amount of acoustic information that native and non-native listeners need for syllable identification was investigated by comparing the performance of monolingual English speakers and native Spanish speakers with either an earlier or a later age of immersion in an English-speaking environment. Duration-preserved silent-center syllables retaining 10, 20, 30, or 40 ms of the consonant-vowel and vowel-consonant transitions were created for the target vowels /i, ɪ, eɪ, ɛ, æ/ and /a/, spoken by two males in /bVb/ context. Duration-neutral syllables were created by editing the silent portion to equate the duration of all vowels. Listeners identified the syllables in a six-alternative forced-choice task. The earlier learners identified the whole-word and 40 ms duration-preserved syllables as accurately as the monolingual listeners, but identified the silent-center syllables significantly less accurately overall. Only the monolingual listener group identified syllables significantly more accurately in the duration-preserved than in the duration-neutral condition, suggesting that the non-native listeners were unable to recover from the syllable disruption sufficiently to access the duration cues in the silent-center syllables. This effect was most pronounced for the later learners, who also showed the most vowel confusions and the greatest decrease in performance from the whole word to the 40 ms transition condition.

© 2008 Acoustical Society of America. [DOI: 10.1121/1.2939127]

PACS number(s): 43.71.Hw, 43.71.Es [MSS]

Pages: 1278–1293

I. INTRODUCTION

To attain “nativelike” proficiency in a second language implies mastering a robustness in speech processing that allows native listeners to adapt to a wide range of listening conditions—from optimal to adverse—that are encountered every day. These daily challenges may result from environmental factors such as noise or reverberation (Miller *et al.*, 1951; Moncur and Dirks, 1967), talker-related factors such as dialect differences or speech impairments (Clopper and Pisoni, 2004; Kent *et al.*, 1990; Hudgins and Numbers, 1942; Monsen, 1983), or linguistic factors such as speaking style (Payton *et al.*, 1994; Lindblom, 1996). Relatively little research on second-language (L2) speech perception has compared native and non-native listeners’ responses to these sources of variation in the speech signal (cf., however, Bradlow and Bent, 2002 and Bradlow and Alexander, 2007, with regard to the effects of speaking style on speech perception by non-native listeners).

One area that has been more extensively investigated is the effects of noise on speech perception by native and non-native listeners. There is accumulating evidence that even early learners of a second language, who may speak their L2 with little or no foreign accent and who perform similarly to monolinguals in quiet, may experience greater difficulty than monolingual listeners in processing speech in noise (Mayo *et al.*, 1997; Meador *et al.*, 2000; Rogers *et al.*, 2006).

Both Mayo *et al.* (1997) and Meador *et al.* (2000) studied non-native English speakers with a first language (L1) of Spanish and found that even early learners of English recognized words in noise presented in sentence context less well than monolinguals. Mayo *et al.* (1997) also found that both early bilingual and monolingual listeners, but not later learners of English, recognized words in noise more accurately in sentences with high semantic predictability than in sentences with low semantic predictability.

Rogers *et al.* (2006) compared the recognition of isolated words presented in noise and in reverberation by monolingual English speakers and non-native English speakers with Spanish L1 and an age of onset of immersion in an English-speaking environment of age six or earlier. They also collected accentedness and self-ratings of language dominance for speaking, listening, reading, and writing for the non-native English speakers. Rogers *et al.* (2006) found that even these early learners of English, who were rated as having little or no foreign accent in English and rated themselves as English dominant or balanced for the skills of listening, speaking, reading, and writing, still recognized significantly fewer words in noise and reverberation than did the monolingual English-speaking listeners. It should also be noted that in all three of the above-mentioned studies, the early learners performed similarly to monolinguals in quiet or at very favorable signal-to-noise ratios (SNRs), while proficient non-natives with a later age of immersion have been shown to perform more poorly than monolinguals in quiet or at favorable SNRs of +15 dB or more (cf. Cutler *et al.*, 2004 and Mayo *et al.*, 1997).

^{a)} Portions of these data were presented at the 147th meeting of the Acoustical Society of America [J. Acoust. Soc. Am., **115**, 2605 (2004)].

^{b)} Electronic mail: crogers@cas.usf.edu

One potential explanation for even relatively early learners' increased difficulty in processing speech in noisy environments is a reduced ability to process speech sounds based on partial information (as may occur in the masking of speech sounds in noisy environments). The hypothesis that learners of a second language may have greater difficulty than monolinguals in identifying speech sounds based on partial acoustic information is compatible with evidence suggesting that some phoneme categories of even experienced non-natives may be intermediate between those of L1 and L2, producing potential mismatches between native and non-native speakers' phoneme categories, perhaps in the form of differences in phoneme boundary locations or differences in cue weighting (cf. [Flege, 1995](#); [Imai et al., 2005](#); and [Flege and MacKay, 2004](#)). To investigate the hypothesis that even proficient non-natives may be less able to identify speech sounds based on partial acoustic information, the present study used a silent-center syllable perception task ([Strange et al., 1983](#); [Parker and Diehl, 1985](#)) to compare the ability of native and non-native listeners to identify consonant-vowel-consonant (CVC) syllables from which varying degrees of the vowel center had been removed.

The silent-center syllable-identification paradigm typically employs stop-vowel-stop sequences in which the vowel center (as measured from the release of the initial voiced stop to the onset of closure for the final stop) is reduced to silence while the consonant-vowel (CV) and vowel-consonant (VC) transitions are retained unchanged ([Strange et al., 1983](#)). [Strange et al. \(1983\)](#) found that native listeners could identify target vowels from which most or all of the vowel "steady states" had been removed nearly as well as the full CVC syllable. [Strange et al. \(1983\)](#) (cf. also [Strange, 1989](#); [Jenkins and Strange, 1999](#); [Jenkins et al., 1999](#); and [Jenkins et al., 1994](#)) interpreted this result to imply that native listeners use dynamic information related to articulatory trajectories into and out of the vowel target for vowel identification, rather than the vowel's target articulatory position (cf., however, [Andruski and Nearey, 1992](#), for an alternative interpretation).

The theoretical interpretation of these data allowed [Strange et al. \(1983\)](#) to account for successful vowel identification in the presence of target undershoot. Since then, several studies and authors have used the silent-center paradigm to determine the amount of the vowel center that can be removed before vowel identification is seriously impaired ([Parker and Diehl, 1985](#)) and to compare vowel identification performance of different populations of native English speakers ([Fox et al., 1992](#); [Sussmann, 2001](#); [Kirk et al., 1992](#); [Murphy et al., 1989](#)). [Fox et al. \(1992\)](#), for example, found that older adults' recognition accuracy for silent-center syllables was significantly lower than that of younger adults, despite similar performance for whole words.

The interpretation of the silent-center results of [Strange et al. \(1983\)](#) does not, however, necessarily imply that all L1s place an equally strong weight on the dynamic information presented in CV and VC transitions. It may be that dynamic information (either vowel intrinsic or that contained in formant transitions) is particularly important for languages such as English with a relatively crowded vowel space, in

which converging sources of acoustic information may be important to differentiate vowels that are close neighbors in phonological space or for vowels that are strongly inherently dynamic (cf. [Kewley-Port and Goodman, 2005](#)). Thus, the ability to identify a vowel based on formant transitions alone may not be equally accessible to all learners of English as a second language and a comparison of these abilities between native and non-native listeners may help us to explain how even proficient non-natives' ability to identify speech sounds may be more strongly degraded in noise than that of monolingual listeners.

Therefore, the present study used a silent-center syllable gating paradigm to compare syllable identification from CV and VC transitions by monolingual English speakers and non-native English speakers with a first language of Spanish and either an early or a later age of immersion in an English-speaking environment. Spanish was chosen as the L1 for comparison because Spanish and English differ markedly in their vowel inventories, Spanish having 5 ([Dalbor, 1969](#)) and American English approximately 12 stressed vowels, excluding diphthongs ([Ladefoged, 1982](#)), and because non-native English speakers with a first language of Spanish constitute a large and rapidly growing minority in the U.S. (approximately 28×10^6 persons at the 2000 Census; [United States Census Bureau, 2000](#)). The silent-center paradigm is attractive because it allows for manipulation of the degree of target acoustic information presented without the use of synthesized speech, which may be perceived less accurately by non-native listeners for reasons other than those we wished to investigate, such as the naturalness or appropriateness of the speech cues themselves (cf. [Logan et al., 1989](#) and [Hillenbrand and Nearey, 1999](#)).

Specifically, the present study sought to compare the performance of monolingual listeners and two groups of non-native listeners differing in age of onset of immersion in (1) identifying target vowels when varying degrees of the CV and VC transitions were presented and (2) using vowel duration as a cue to vowel identification in silent-center syllables. Six target vowels (/i, ɪ, eɪ, ε, æ/ and /a/) that span the vowel space from high to low and occupy a region of the vowel space that is considerably more crowded in English than in Spanish (cf. [Dalbor, 1969](#) and [Ladefoged, 1982](#)) were selected.

Performance on the individual target vowels and vowel confusions (vowels selected other than the target vowel) was also compared across the three listener groups. Specific hypotheses about confusion patterns for individual target vowels were not made because the acoustic differences and perceptual assimilation patterns from Spanish to American English vowels have not been extensively investigated and because a wide variety of dialects of Spanish are spoken in the Tampa Bay area, potentially affecting vowel assimilation patterns in different ways. It was hypothesized, however, that the late learners and perhaps the early learners would show greater decreases in performance than the monolingual listeners when vowel duration was not available as a cue. It was also anticipated that the late learners would show the lowest overall performance and the greatest number of confusions for some target vowels, due to more poorly defined vowel

categories, and that different confusion patterns would be observed when vowel duration was available as a cue than when it was not, perhaps for both the early and late learners of English.

II. METHOD

A. Participants

Three groups of participants were recruited: monolingual native English speakers, relatively early learners of English as a second language (age of onset of immersion of 12 years or earlier), and late learners of English as a second language (age of onset of immersion of 18 years or later). Participants were screened to include only persons between the ages of 18 and 50 with no history of speech or hearing disorders. For the monolingual group, potential participants who reported fluency in a second language or exhibited a regional accent that differed strongly from that of the Tampa metropolitan area were also excluded. All potential non-native English-speaking participants were required to be native speakers of Spanish. No speakers of Peninsular varieties of Spanish were included; however, regional variation within New World varieties of Spanish was not controlled. Potential non-native participants who reported fluency in a language other than Spanish or English were also excluded. All participants were also required to pass a basic hearing screening (20 dB hearing level (HL) at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz) prior to participation in the experiment.

Participants were recruited from flyers posted around the campus of the University of South Florida and from newspaper advertisements. Non-native participants were also recruited from among the second author's personal contacts. Prior to participation, all potential participants were required to fill out a language background questionnaire. Forms for both native and non-native English-speaking participants included items related to participants' age, dialect background, history of speech or hearing disorders, and languages spoken. In addition, the form for the potential non-native participants included items probing parents' native languages, age of onset of learning English (AOL), age of onset of immersion in an English-speaking environment (AOI), number of years living in the U.S., and self-ratings of language dominance (more proficient in English or Spanish) for the skills of listening, speaking, reading, and writing.

Participants were compensated by payment of \$8.00 per hour of participation, or by an equivalent-value gift certificate. All participants were required to have sufficient proficiency in English to read and understand the consent forms and language background questionnaire, which were printed in English. The second author, a native Spanish speaker with an early age of immersion in an English-speaking environment, was able to assist some participants if they had trouble understanding particular words or phrases, although this assistance was seldom needed.

According to the criteria outlined above, 13 monolingual English speakers (MO), 10 early learners of English as a second language (EL), and 8 late learners of English as a second language (LL) completed the tasks. Data for two monolinguals who did not appear to understand or take the

task seriously were removed from the analysis. Data for one additional monolingual speaker were removed from the analysis due to the need to complete counterbalancing of listening conditions across listeners (see below), leaving ten monolinguals whose data were included in statistical analyses.

Although participants of both genders were recruited, volunteers were primarily female, resulting in unbalanced groups (nine females and one male in the MO group, ten females and zero male in the EL group, and five females and three males in the LL group). A literature search revealed several studies showing gender differences in various aspects of speech production and intelligibility, but few that examined the effects of listener gender on speech perception were found. Two studies that indicated no effect of listener gender on the processing of male versus female voices, either behaviorally (Mullenix *et al.*, 1995) or neurophysiologically (Lattner *et al.*, 2005) were found. Thus, when an examination of the individual data for the male listeners did not reveal patterns of performance that were markedly different from the means for the respective groups, data for the male listeners were retained.

Following the listening tasks described below, all participants were recorded in a sound-attenuating booth as they read three English sentences selected from the Harvard sentences, which contain five key words and are semantically appropriate but not highly predictable (IEEE, 1969 and Egan, 1948). All participants were recorded using an Audio-Technica AT4033a microphone. The output of the microphone was routed through a preamplifier and recorded to a Roland VS890 digital recorder at a sampling rate of 44.1 kHz [16 bit analog-to-digital (A/D) converter]. The recorded sentences were digitally transferred to computer and then isolated to file. Each sentence file was amplitude equalized using the rms amplitude of the entire sentence file (which always contained about 10 ms of silence at the beginning and end), in order for the sentences to be presented to the listeners at an approximately equal presentation level.

As part of a related study (Crawford, 2006), 15 adult monolingual English-speaking females heard the three sentences spoken by each of the 28 participants in the present study, presented in random order. The raters were asked to rate each sentence for foreign accentedness on a nine-point scale, with one representing no foreign accent and nine representing a very strong foreign accent. For the present study, the listeners' accentedness ratings were averaged across sentences and listeners to provide a measure of proficiency beyond that provided by the participants' self-ratings of language dominance.

The average ages of the participants in the MO, EL, and LL groups were 26.4, 27.1, and 26.0 years, respectively. The participants ranged in age from 19 to 48 years; the standard deviations (SDs) for the participants' ages were 5.4, 10.1, and 6.5 years for the MO, EL, and LL groups, respectively. The average AOI was 5.6 years (SD=3.3) for the EL group and 25.1 years (SD=6.5) for the LL group. The average number of years in the U.S. was 21.0 (SD=11.1) for the EL group and 1.4 (SD=2.3) for the LL group. Three of the ten EL participants (EL06, EL07, and EL08) were born and

TABLE I. Demographic data for individual participants who were either early (EL) or late (LL) learners of English as a second language. Data are displayed for gender; age, country of origin (of listener or listener's parents if born in the U.S.); age of onset of immersion (AOI); number of years spent living in the U.S.; self-ratings of language dominance (E=English, S=Spanish, and B=balanced) for the skills of listening, speaking, reading, and writing; and average foreign accentedness ratings from a nine-point scale, with one indicating little or no accent and nine indicating a very strong accent.

Listener	Gender	Age	Country	AOI	Years in U.S.	Listen	Speak	Read	Write	Accent
EL01	F	48	Cuba	5	42	E	E	E	E	2.56
EL02	F	25	Colombia	5	7	S	E	S	S	1.63
EL03	F	23	Mexico	10	13	E	E	E	E	1.40
EL04	F	44	Cuba	8	36	B	E	B	E	2.38
EL05	F	22	Cuba	3	21	E	E	E	E	3.37
EL06	F	21	Cuba	3	21	E	E	E	E	2.33
EL07	F	21	Cuba- Colombia	4	21	S	E	E	E	2.22
EL08	F	25	Cuba	0	25	E	E	E	E	1.46
EL09	F	22	Puerto Rico	8	14	S	B	E	E	3.27
EL10	F	20	Puerto Rico	10	10	E	E	E	E	1.38
LL01	F	23	Peru	22	1	E	E	S	S	6.94
LL02	M	24	Colombia	21	8	S	S	S	S	6.71
LL03	F	41	Colombia	40	1	S	S	S	S	7.77
LL04	F	25	Colombia	25	<1	S	S	S	S	5.70
LL05	F	27	EI Salvador	27	<1	S	S	S	S	7.87
LL06	M	23	Nicaragua	23	<1	S	S	S	S	7.03
LL07	M	26	Canary Islands	24	<1	S	S	S	S	6.62
LL08	F	19	Colombia	19	<1	S	E	S	S	5.87

raised in the U.S., but only one (EL08) reported being immersed in an English-speaking environment since birth. The other two participants who were born in the U.S. reported AOIs of 3 and 4 years, but some exposure to English via television and parents' interactions outside the home is likely to have occurred before this age for these two talkers. All three received all of their formal schooling in English. Table I displays the following data for the individual EL and LL participants: (1) age; (2) AOI; (3) country of origin (or country of origin of parents if the listener was born in the U.S.); (4) number of years spent living in the U.S.; [(5)–(8)] self-ratings of language dominance for listening, speaking, reading, and writing; and (9) average ratings of foreign accentedness across the three sentences recorded.

As can be seen from Table I, the EL participants typically rated themselves as English dominant in most domains; eight out of ten EL participants rated themselves as English dominant for reading and nine out of ten rated themselves as English dominant for writing and speaking; only six out of ten rated themselves as English dominant for listening, however. Most of the EL listeners received much or all of their schooling in the U.S., so their self-ratings of English dominance are not surprising, especially for the reading and writing domains. Six of the eight LL participants rated themselves as Spanish dominant in all four domains.

The average accentedness rating was 1.47 ($SD=0.22$) for the MO participants, 2.20 ($SD=0.74$) for the EL participants, and 6.87 ($SD=0.78$) for the LL participants. The range of ratings across participants within each group was 1.22–

2.03 for the MO participants, 1.38–3.37 for the EL participants, and 5.70–7.87 for the LL participants. Thus, all of the MO and EL participants received an average accentedness rating in the lower third of the scale (1.0–3.67), indicating little or no foreign accentedness. Furthermore, the scores of the MO and EL participants overlapped substantially, with four of the ten EL participants obtaining accentedness ratings within the range of scores obtained by the MO participants, suggesting a native or near-native degree of proficiency in spoken English. None of the average ratings for the individual LL participants fell within the range of scores obtained by either the MO or EL participants. All of the scores for the LL participants fell in the upper half of the scale (5.0–9.0), indicating at least a moderate to strong degree of foreign accent for all the talkers.

The accentedness rating data also support the retention of data for two participants whose demographic data do not otherwise fit the pattern of data obtained for the participants within their respective groups: EL02 and LL02 (see Table I). Participant EL02 reports Spanish dominance in three of the four domains queried, but obtained an accentedness rating of 1.63, within the range of scores obtained for the native English speakers, suggesting a native or near-native degree of proficiency in speaking English. This participant also obtained an overall score of 100% correct on perception of the whole words, one of the two highest scores for participants in this group, again indicating nativelike proficiency on this task.

On the other hand, participant LL02 reports a much

longer time of residence in the U.S. than the other LL participants, but obtained a mean accentedness rating (6.71) that was just below the group mean of 6.87 and obtained an average perception score of approximately 69% correct on the whole-word identification condition, which was about 10% below the overall average score for the LL group on this condition. These data would suggest that the perception and production skills of this listener do indeed fit with those of the other LL participants, despite his long time of residence in the U.S. Note also that this participant reports only 3 years of immersion in an English-speaking environment, despite a much longer length of residence in the U.S. Such situations are not unusual for Spanish L1 late learners in Florida, where extensive Spanish-speaking communities exist in cities such as Miami and Tampa, underscoring the need for detailed questionnaires if participants' true age of immersion is to be recorded.

B. Stimuli

1. Target words

Target words in /bVb/ context were used, as in [Strange et al. \(1983\)](#); however, only the following six target vowels were selected: /i, ɪ, eɪ, ε, æ/ and /a/.

2. Speakers and recording procedure

Three monolingual native speakers of American English (two males and one female) with an accent typical of that of persons from the Tampa metropolitan area were recorded saying the target words and nonwords ("beeb, bib, babe, bebb, babb" and "bob") in the following carrier phrase: "I say ——— on the tape." Speakers were instructed to speak at their normal speaking rate. Twenty repetitions of each target phrase were recorded. Recordings from the female talker were used to create example and practice stimuli; recordings from the two male talkers were used to create the experimental stimuli.

Talkers were seated in a sound-attenuating booth and read the target sentences from a sheet of paper placed on a stand. A boom-mounted Audio-Technica AT4033 microphone was placed at a 45° angle and approximately 6 in. from the talker's lips. Output from the microphone was routed through an Applied Research and Technology Professional Tube Mic preamplifier and digitized at a 44.1 kHz sampling rate (16 bit A/D converter) using a Roland VS890 digital recorder. Sound files stored on the digital recorder were transferred directly to computer using the digital input of a high-quality sound card (M-Audio Audiophile 2496) and a signal editing software program ([COOLEDIT 2000, 2000](#)).

3. Word isolation and creation of whole-word stimuli

Three clear tokens of each of the six target words that were judged to maintain some audible differences from token to token were selected for each of the two male talkers. The female talker's recordings were used only for the creation of example and practice stimuli and therefore only one clear token of each target word was selected for this talker. Target words were isolated as follows. First, the release times of the

initial and final /b/ consonants were identified from the wave form, based on the small burst of noise energy associated with the release of the /b/. Next, the 15 ms of energy preceding the initial /b/ release and the 15 ms of energy following the final /b/ release were preserved; all energies preceding (in the case of initial /b/) or following (in the case of the final /b/) these 15 ms buffers were silenced. Linear on-ramping and off-ramping of the first 2 ms of the initial 15 ms buffer and the final 2 ms of the final consonant 15 ms buffer were used to eliminate any clicks associated with the abrupt onset or cessation of energy created by the silencing described above. Thus, up to 15 ms of prevoicing for initial /b/ and a clear release of the final /b/ were preserved, where present. Finally, 15 ms of the silence created at the beginning and at the end of the word were preserved; all other energies were deleted and the token was saved to file.

The resulting 36 sound files for the two male talkers (2 talkers $\times$ 3 tokens $\times$ 6 target vowels) served as the whole-word stimuli and as the basis for the creation of the silent-center stimuli. Six whole-word sound files were created in the same manner for the female talker. To ensure that all these whole-word stimuli would be presented at approximately the same overall intensity, the root-mean-square (rms) amplitude of each target word file was adjusted to equal 15 dB less than the maximum amplitude using an automated procedure ([COOLEDIT 2000, 2000](#)). Stimuli were screened for peak clipping following the amplitude adjustment procedure; no stimuli were found to be peak clipped. Figure 1(A) shows an example "whole word" wave form for the syllable "bebb," spoken by talker 1.

In addition to the times of the onset of release of the initial and final /b/, the time of closure for the final /b/ was measured and used to compute target vowel durations for the two male talkers. Vowel duration was measured as the time from the release of the initial /b/ to the closure for the final /b/ (cf. [Strange et al., 1983](#)). The onset of closure for the final /b/ was measured from the wave form by locating the point in time at which voicing ceased or the point in time at which the wave form of the voicing cycles changed from more complex to more sinusoidal, indicating the onset of low-pass filtering created by lip closure.

Fundamental frequency (F_0) and the frequencies of the first and second formants (F_1 and F_2) at vowel midpoint were also measured for the 36 syllables selected as stimuli. All frequency analyses were made by the first author using PRAAT ([Boersma and Weenik, 2006](#)). Values for F_0 were made using autocorrelation analysis with a pitch range between 100 and 500 Hz in most cases. For five of the stimuli for talker 1, the F_0 was below or within 5 Hz of the 100 Hz minimum, and the low end of the pitch range was therefore set to 80 Hz to ensure appropriate measurement of the F_0 . Formant values were measured using linear predictive coding (LPC) analysis, with an analysis range 0–5500 Hz, a 20 ms analysis window, and between five and seven formants used for LPC tracking. Formant tracks were overlaid on a wide-band spectrogram of the target syllable, and the number of formants used for tracking was adjusted up or down from a default of six, until a good visual match with the formants observed on the spectrogram was obtained. A

FIG. 1. Wave forms showing an example of the target syllable “bed,” spoken by talker 1 and edited to create the following three listening conditions: whole-word (A), 20 ms DP (B), and 20 ms DN (C).

good visual match was obtained in all cases and no analysis by hand of F_0 , F_1 , or F_2 values was judged to be needed.

Table II shows the average vowel duration, F_0 , F_1 , and F_2 values for each talker and target vowel, as well as the average across talkers and the average across vowels. These

durations indicate a somewhat slow rate of speech for these two male talkers. The F_0 , F_1 , and F_2 values are largely compatible with values expected for these vowels spoken in General American English for a male talker, although the F_0 for talker 2 is somewhat higher than average.

4. Creation of duration-preserved silent-center stimuli

Four silent-center versions of each target word were created, with 10, 20, 30, and 40 ms of the initial CV and final VC information preserved. These versions or “gates” were selected based on pilot testing that showed a ceiling effect for gates of 35 ms or greater (15, 25, 35, and 45 ms gates were tested). Silent-center tokens were created by first selecting the desired gate duration (e.g., 20 ms) immediately following the release of the initial /b/. Next, the closure of the final /b/ was located and the desired gate duration (e.g., 20 ms) prior to the release of the final /b/ was selected. All energies between the initial and final selections were then silenced. The edges of the initial and final preserved portions of the syllable were then off-ramped and on-ramped, respectively, using a 2 ms linear ramp, as described above for the initial and final portions of the word. This process was repeated for each of the four desired gates, resulting in 144 additional stimuli (2 talkers $\times$ 3 tokens $\times$ 6 target vowels $\times$ 4 gates). These were termed the “duration-preserved (DP)” silent-center stimuli because no modification of the duration of the silent center was performed. Figure 1(B) shows an example of a 20 ms DP silent-center wave form for the syllable “bebb,” spoken by talker 1.

Six additional DP silent-center stimuli (one for each target word) were created from the female talker’s utterances, using the 30 ms silent-center gate only. These were used as practice stimuli.

5. Creation of duration-neutral silent-center stimuli

To create the duration-neutral (DN) silent-center tokens, the vowel duration of each of the 144 DP silent-center stimuli was adjusted to equal the average vowel duration across all tokens of the target words spoken by the two male talkers (267 ms). This was done by noting the measured duration of the vowel in question (from initial /b/ release to final /b/ closure) and then inserting or deleting the appropri-

TABLE II. Mean duration (in s) and F_0 , F_1 , and F_2 values (in Hz) for each of the six target vowels, with the average across vowels for each talker and the average across talker for each vowel.

Target vowel	Talker 1				Talker 2				Average across talkers			
	Dur	F_0	F_1	F_2	Dur	F_0	F_1	F_2	Dur	F_0	F_1	F_2
/i/	0.291	120	243	2327	0.293	184	349	2507	0.292	152	296	2417
/ɪ/	0.222	125	407	1803	0.183	167	506	1993	0.202	146	457	1898
/eɪ/	0.323	126	386	2100	0.275	170	493	2473	0.299	148	440	2286
/e/	0.237	110	586	1600	0.216	171	640	1907	0.227	141	613	1753
/æ/	0.285	103	718	1595	0.280	165	724	1987	0.282	134	721	1791
/a/	0.295	101	677	1087	0.309	172	766	1323	0.302	137	721	1205
Average across vowels	0.276	114	503	1752	0.259	171	580	2032	0.267	143	541	1892

ate duration of silence in the silent-center portion so that the resulting vowel duration was equal to the average vowel duration across the two male talkers. This procedure resulted in an additional 144 DN silent-center stimuli. No DN silent-center stimuli were created from the female talker's utterances. Figure 1(C) shows an example of a 20 ms DN silent-center wave form for the syllable "bebb," spoken by talker 1. Prior to presentation to listeners, the sample rate of all stimuli was adjusted to 48.8 kHz to accommodate the software program used for stimulus presentation (ECOS/WIN, 1999).

C. Main experiment procedure

1. Listening environment

Listeners were seated in groups of up to four in a quiet, sound-treated room with individual carrels for each listener. Each carrel was separated by a divider and an empty carrel separated each listener from the next. Each listener's carrel was equipped with a flat-screen monitor, keyboard, and mouse. The CPU controlling each independent listening station was located outside the room. The stimuli were presented binaurally over headphones (Sennheiser HD265) at approximately 70 dB sound pressure level (SPL) (based on the whole-word stimuli). Presentation level was controlled using the programmable attenuators (PA5) of the Tucker-Davis Technologies (TDT) *Psychoacoustics System III* (2001) hardware.

2. Calibration

A 1000 Hz tone was used for calibration of stimulus presentation level. The rms amplitude of the tone was adjusted to match that of the whole-word stimuli (15 dB below maximum amplitude), prior to the calibration procedure. During calibration, the tone was played out without attenuation through the headphones. As the tone was played out, the right and left headphones were placed in turn onto the coupler of a sound-level meter (Bruel & Kjaer Model 2235) and the level of input to the sound-level meter was measured in dB SPL and noted. The amount of attenuation needed for a presentation level of 70 dB SPL (average across the two headphones) was then computed and the measurement procedure was repeated with this setting to confirm a presentation level of 70 dB. The attenuation levels of the PA5s for the four listening stations were adjusted accordingly within the software used for presentation of stimuli (ECOS/WIN, 1999).

3. Listener task

Prior to beginning the experiment, the listeners were familiarized with the pronunciation and spelling used for all of the target words and nonwords. For all trials (practice and main experiment), one of the six target words was presented over the headphones and six alternatives were displayed within boxes on the screen (two rows and three columns) in the following order (clockwise): "beeb, bib, babe, bebb, babb" and "bob." To focus the listeners' attention on the target vowels and to ensure accurate reading of the nonwords, a common word with the same vowel as the target

word was displayed on the screen below each target word ("feed, crib, tape, red, crab" and "dog"). Listeners were informed that the more common words were displayed for reference purposes only and that the word presented would always be one of the target (/bVb/) words; they all reported familiarity with the pronunciation of the key words used.

Listeners were verbally familiarized with the task prior to participation. In addition, written instructions were provided on the monitor prior to each set of trials (example, practice, silent center, and whole word). On each trial, listeners were instructed to choose which word they had heard and responded by left clicking with the mouse. Listeners were informed that some items would be more difficult than others and to make their best guess if unsure. The order of presentation of all stimuli and collection of listener responses were controlled automatically using ECOS/WIN (1999). All trials were self-paced; the next item was presented approximately 1 s following the listener's response.

4. Example and practice trials

The six whole-word stimuli created for the female talker were used to create six example trials. On these trials, all six whole-word stimuli were presented in the following order: "beeb, bib, babe, bebb, babb" and "bob." Listeners were informed of the order prior to the beginning the task. For the example trials, the correct response was highlighted in green following the listener response in order to provide visual reinforcement to the listener. Six silent-center *practice* trials were created using six 30 ms gate silent-center stimuli created from the utterances of the female talker. Feedback was not provided on these trials and the words were presented in random order.

5. Main experiment trials

Listeners completed three blocks of trials in the main experiment: 144 DP trials, 144 DN trials, and 36 whole-word trials. To control for any practice effects on the silent-center trials, half of the listeners in each group completed the DP trial block first and half of the listeners completed the DN trial block first. All listeners completed the whole-word trials last, to avoid overfamiliarization with the stimuli prior to presentation of the silent-center trials. Stimuli were presented in random order in all three blocks and no feedback was provided. A required 5 min break was provided prior to the whole-word block; listeners were allowed to take a break following any trial block. The entire experiment, including completion of forms, hearing screening, and all experimental trials, took between 1 and 1.5 h.

The ECOS/WIN program (1999) automatically scored listener responses as correct or incorrect and recorded information on the alternative chosen on each trial. These data were imported to a spreadsheet and the number of correct responses was computed for each target vowel at each of the nine listening conditions (2 listening conditions $\times$ 4 gates + the whole-word condition). Confusion matrices showing the number of items correct and the alternatives chosen for incorrect responses were computed for each listener for both the DP and DN conditions. Prior to data analysis, percent-

FIG. 2. Percent-correct identification performance by listener group and listening condition, averaged across target vowels. The solid lines with filled symbols indicate the performance on duration-preserved (DP) conditions and the dashed lines with open symbols indicate the performance on duration-neutral (DN) conditions. The performance of monolingual listeners (MO) is indicated by circles, the performance of early learners of English as a second language (EL) is indicated by squares, and the performance of late learners (LL) is indicated by triangles. Error bars indicate one standard error of the mean.

correct scores were converted to rationalized arcsine transform unit (RAU)-transformed scores, which correct for correlation of variances with the mean that can occur when proportional data are used, help to correct for ceiling and floor effects, and yield values that are reasonably interpretable with respect to the corresponding percent-correct scores (Studebaker, 1985).¹

III. RESULTS

A. Whole-word performance

Figure 2 displays percent-correct performance for each listener group on each of the nine conditions (2 duration conditions $\times$ 4 gates + the whole-word condition), averaged across target vowels. As shown in the figure, the monolingual (MO) and early learners of English as a second language (EL) listeners performed nearly identically and nearly perfectly on the whole-word condition (96% and 94%, respectively, or 100 and 97 RAU); the performance of the late learners (LL) on the whole-word condition was about 27% (or 30 RAU) below that of the other two listener groups (approximately 69% correct or 68 RAU).²

While the overall performance of the LL listeners on the whole-word condition is substantially poorer than that of the other two groups, it is also about 52% above chance performance (17%) for this six-alternative forced-choice task. Thus, while their phonetic categories for the target vowels may be less well formed than those of the other two groups, these data do suggest that the whole-word stimuli were reasonably well identified by the majority of the LL listeners.

B. Effects of partial vowel information (whole-word versus 40 ms duration-preserved syllables)

Despite the lower overall performance of the LL group, performance appears to differ more between the whole-word

to the 40 ms DP condition for this group (about a 25% or 24 RAU difference in performance between the whole-word and the 40 ms DP condition) than for the MO and EL groups (approximately 9% and 13% or 11 and 15 RAU differences in performance, respectively). To examine just the effects of presenting partial vowel information on performance for the three listener groups, a three-way mixed design analysis of variance (ANOVA) was performed using SPSS (2006), with listener group (three levels) as the between-subjects variable and listening condition (whole word vs. 40-ms DP) and target vowel (six levels) as the within-subjects variables. RAU-transformed (Studebaker, 1985) percent-correct identification performance was the dependent variable. The whole-word condition could not be included in a larger ANOVA with the effects of gate and listening condition because there was no DN whole-word condition. Therefore, this limited analysis was considered the best way of determining whether removing *any* vowel information at all affected one group more than another.

Partial eta-squared (η^2_p) effect-size statistics provided by SPSS (2006) were converted to generalized eta-squared (η^2_G) values, using the formulas provided by Bakeman (2005). According to Bakeman (2005), comparisons of effect sizes between studies with between-subjects variables and studies with any within-subjects variables are not appropriate when η^2_p effect-size values are used because η^2_p can be much larger in within-subjects or mixed designs than in between-subjects designs showing similar effect sizes for a given factor. Bakeman (2005) recommended η^2_G as a measure of effect size that is interpretable as the proportion of variance in the dependent variable accounted by the effect, that is appropriate for designs with within-subjects variables, and that allows for comparisons across studies with between-subjects, within-subjects, and mixed designs.

As recommended by Bakeman (2005), type I sum of squares values were used in the ANOVA and the computation of effect sizes and power, due to the unequal group sizes in the between-subjects variable. Table III shows values of F , degrees of freedom, η^2_p , η^2_G , and power for each of the main effects and interactions, along with the classification of effect sizes as negligible, small, medium, or large, according to the guidelines suggested by Bakeman (2005).

As can be seen from Table III, the three-way ANOVA showed significant main effects of listener group, listening condition, and target vowel; based on η^2_G values, the effect sizes of these main effects were categorized as large, medium, and small, respectively. Two interactions were significant: target vowel by listening condition and listener group by target vowel by listening condition; both of these effects were categorized as small. Statistical power reached the generally accepted criterion level of 0.8 or above for all but two effects: the two-way interaction between listener group and listening condition and the two-way interaction between listener group and target vowel. The effect size for the two-way interaction between listener group and listening condition was classified as negligible, while that for the two-way interaction between listener group and target vowel fell at the low end of the range classified as small. Thus, the only two non-significant interactions that were substantially underpowered

TABLE III. F values, degrees of freedom, p values, power and effect size data for the three-way ANOVA examining the effects of listener group, listening condition (whole word vs DP 40 ms silent-center) and target vowel on RAU-transformed percent-correct syllable identification performance. Both partial eta-squared (η_p^2) and generalized eta-squared (η_G^2) effect size statistics are provided, as well as the recommended classification for η_G^2 (cf. Bakeman, 2005). Bold type indicates effects that reached significance.

Effect	F	df	p	Power	η_p^2	η_G^2	η_G^2 classified as
Main effects							
Listener group	48.50	2,23	<0.0005	1.00	0.81	0.40	Large
Listening condition	64.12	1,23	<0.0005	1.00	0.74	0.09	Small
Target vowel	13.51	5,115	<0.0005	1.00	0.37	0.13	Medium
Two-way interactions							
Listener group X listening condition	1.89	2,23	0.175	0.35	0.14	0.01	Negligible
Listener group X target vowel	0.80	10,115	0.625	0.40	0.07	0.02	Small
Listening condition X target vowel	6.99	5,115	<0.0005	1.00	0.23	0.05	Small
Three-way interaction							
Listener group X listening condition X target vowel	2.11	10,115	0.029	0.88	0.16	0.03	Small

were also classified as negligible to small, suggesting that no important effects failed to reach significance, despite the relatively small number of participants in each of the three listener groups.

A Tukey HSD *post hoc* analysis of the main effect of group showed no significant difference in the performance of the MO and EL listener groups ($p=0.333$) and significantly lower performance for the LL group than for the other two groups ($p<0.0005$ for both comparisons). On average, the EL listeners identified the target syllables about 4% (or 4 RAU) less accurately than the MO listeners and the LL listeners identified the target syllables about 32% (or 32 RAU) less accurately than the EL listeners.

An examination of the main effect of target vowel showed that the six target vowels were perceived in the following order, from most to least accurately perceived, across the groups, and listening conditions: /a/ (94% or 97 RAU), /æ/ (89% or 91 RAU), /e/ (86% or 88 RAU), /i/ (81% or 83 RAU), /ɛ/ (74% or 75 RAU), and /ɪ/ (71% or 71 RAU). Simple main effects of comparisons with Bonferroni adjustment for the 15 comparisons among pairs of target vowels showed that target /a/ was perceived significantly more accurately than all of the other vowels except /æ/; target /æ/ was perceived significantly more accurately than /ɪ/ and /ɛ/ but not differently from /i/ or /e/; target /e/ was perceived significantly more accurately than /ɪ/, but not differently from /i/ or /ɛ/; performance for target /i,ɛ/ and /ɪ/ did not differ significantly. This order did not differ dramatically in the interactions and will not be discussed further in this section.

Figure 3 compares performance across the levels of the three factors in the significant three-way interaction: listener group, target vowel, and listening condition (whole word versus DP 40 ms). Performance for each listener group is shown as a separate panel. The three-way interaction was

explored by pairwise comparisons of listeners' performance on the two listening conditions at each level of group and target vowel and by pairwise comparisons of the performance of the three listener groups at each level of listening condition and target vowel. Bonferroni adjustment for the number of comparisons at each level was used.

As shown in the figure, similar patterns of performance were obtained for the MO and EL listener groups. Performance differed significantly between the whole-word and the DP 40 ms conditions for the target vowels /ɪ/ and /ɛ/ for both the MO and EL groups [by about 20%–28% or 23–31 RAU; see Figs. 3(A) and 3(B)]. For target /i/, performance for the MO but not the EL group differed significantly between the whole-word and the DP 40 ms conditions (about a 13% or 16 RAU difference for the MO listeners and 10% or 11 RAU for the EL listeners), while for target /e/ performance for the EL but not the MO group differed significantly between the whole-word and the DP 40 ms conditions (a 12% or 15 RAU difference for the EL group but only 2% or 3 RAU for the MO group). Performance for target /æ/ and /a/ did not differ significantly between the whole-word and the DP 40 ms conditions for either the MO or the EL listener group (with differences of at most 2% or 4 RAU in each case).

Performance for the LL group was lower than for the other two groups and typically more variable, even on the whole-word condition [see Fig. 3(C)]. Performance for the LL group differed significantly between the whole-word and the DP 40 ms conditions for the target vowels /i,ɪ/ and /a/, for which performance differed by 56%, 36%, and 11% or 51, 34, and 11 RAU, respectively between the two conditions.

Comparisons of group within each level of listening condition and target vowel revealed relatively minor differences. First, the performance of the MO and EL groups did not differ significantly for any target vowel, although the

FIG. 3. Percent-correct identification performance by vowel and information condition, for the whole-word and DP 40 ms conditions and each of the three listener groups: monolinguals (A), early learners (B), and late learners (C). In each panel, darker gray bars indicate the whole-word condition and lighter gray bars indicate the DP 40 ms condition. Error bars indicate one standard error of the mean, and asterisks indicate conditions that differed significantly in performance within each target vowel.

difference approached significance for target /eɪ/ in the DP 40 ms condition ($p=0.064$), in which performance for the MO group was about 13% or 18 RAU higher than the performance for the EL group [see Figs. 3(A) and 3(B)]. Second, the performance of both the MO and EL groups was significantly higher than that of the LL group for all target vowels and conditions, except target /ɛ/ in the DP 40 ms condition and target /i/ in the whole-word condition, for which no group differences were significant [see Figs. 3(A)–3(C)]. The difference in the performance between the MO and LL listeners did approach significance for target /i/

in the whole-word condition ($p=0.056$), in which the performance for the MO group was about 15% or 22 RAU higher than that of the LL group. The performance of the LL listener group ranged from 19% to 57% (or 23 to 57 RAU) lower than that of the other two listener groups for the target vowels showing significant differences. Together, these data show that the EL listeners were able to perform as well as the MO listeners when the whole syllable was provided, or when the complete CV and VC transitions were available. The LL listeners performed more poorly overall and showed overall larger decreases in performance from the whole-word to the DP 40 ms condition than did the other two groups.

C. Effects of group, gate, duration neutralization, and target vowel

To address the main research question regarding the effects of varying transition information on vowel identification, a four-way mixed design ANOVA was used to compare the effects of gate, duration neutralization, and target vowel across listener groups. Listener group (three levels) was the between-subjects variable and gate (four levels), and duration condition (DP versus DN) and target vowel (six levels) were within-subjects variables. As in the first ANOVA, RAU-transformed percent-correct identification performance was the dependent variable; η^2_p values were converted to η^2_G values; and type I sum of squares values were used in the computation of effects, effect sizes, and power. Table IV shows values of F , degrees of freedom, η^2_p , η^2_G , and power for each of the main effects and interactions, along with the classification of effect sizes as negligible, small, medium, or large, according to the guidelines suggested by Bakeman (2005).

As shown in Table IV, the four-way ANOVA showed significant main effects of listener group, target vowel, and gate; based on η^2_G values, the effect sizes of these main effects were categorized as large, medium, and small, respectively. The main effect of duration condition was not significant. Significant two-way interactions were found for the listener group by duration condition, the listener group by gate, and gate by target vowel; all three of these effects were categorized as small or negligible in size (see Table IV). No other interactions were significant, although the listener group by gate by target vowel interaction approached significance ($p=0.061$). Statistical power failed to reach the generally accepted criterion level of 0.8 or above for several effects or interactions, but the effect size for all but one was categorized as negligible. The remaining underpowered interaction (listener group by target vowel) approached significance ($p=0.098$) but its effect size fell at the low end of the range classified as small. Thus, as in the first ANOVA, all of the nonsignificant interactions that were substantially underpowered were also classified as negligible to small, suggesting that no important effects failed to reach significance, despite the relatively small number of participants in each of the three listener groups.

Unlike in the analysis comparing performance on the whole-word and DP 40 ms conditions, a Tukey HSD *post hoc* analysis of the main effect of group showed significant

TABLE IV. F values, degrees of freedom, p values, power and effect size data for the four-way ANOVA examining the effects of listener group, gate, duration condition (DP vs DN syllables), and target vowel on RAU-transformed percent-correct syllable identification performance. Both partial eta-squared (η_p^2) and generalized eta-squared (η_G^2) effect size statistics are provided, as well as the recommended classification for η_G^2 (cf. Bakeman, 2005). Bold type indicates effects that reached significance.

Effect	F	df	p	Power	η_p^2	η_G^2	η_G^2 classified as
Main effects							
Listener group	54.12	2,25	<0.0005	1.00	0.81	0.36	Large
Duration condition	2.22	1,25	0.149	0.30	0.08	0.00	Negligible
Target vowel	26.17	5,125	<0.0005	1.00	0.51	0.22	Medium
Gate	76.71	3,75	<0.0005	1.00	0.75	0.07	Small
Two-way interactions							
Listener group X duration condition	4.20	2,25	0.027	0.68	0.25	0.00	Negligible
Listener group X target vowel	1.66	10,125	0.098	0.77	0.12	0.03	Small
Listener group X gate	2.80	6,75	0.017	0.86	0.18	0.01	Negligible
Duration condition X target vowel	1.97	5,125	0.088	0.65	0.07	0.00	Negligible
Duration condition X gate	2.29	3,75	0.085	0.56	0.08	0.00	Negligible
Gate X target vowel	6.82	15,375	<0.0005	1.00	0.21	0.03	Small
Three-way interactions							
Listener group X duration condition X target vowel	1.22	10,125	0.282	0.61	0.09	0.00	Negligible
Listener group X duration condition X gate	0.23	6,75	0.964	0.11	0.02	0.00	Negligible
Listener group X gate X target vowel	1.46	30,375	0.061	0.98	0.10	0.01	Negligible
Duration condition X gate X target vowel	0.61	15,375	0.868	0.40	0.02	0.00	Negligible
Four-way interaction							
Listener group X duration condition X gate X target vowel	1.38	30,375	0.093	0.97	0.10	0.01	Negligible

differences in performance between all three groups ($p = 0.032$ for MO versus EL and $p < 0.0005$ for LL versus MO and EL). Overall, the MO listeners identified the target syllables about 8% (or 10 RAU) more accurately than the EL listeners, and the EL listeners identified the target syllables about 28% (or 29 RAU) more accurately than the LL listeners. Thus, overall, even the EL listeners were found to identify the syllables significantly more poorly than the monolingual listeners when partial vowel information was provided. The main effect of gate will not be discussed separately because it was changed by its interactions with other variables.

Similar to the first analysis, an examination of the main effect of target vowel showed that the six target vowels were perceived in the following order of accuracy, from highest to lowest, across the groups, and listening conditions: /a/ (87% or 90 RAU), /æ/ (78% or 79 RAU), /eɪ/ (73% or 74 RAU), /i/ (57% or 58 RAU), /ɪ/ (53% or 53 RAU), and /ɛ/ (51% or 52 RAU). Pairwise comparisons using Bonferroni adjustment for the number of comparisons among the target vowels revealed significantly higher performance for target /a/ than all of the other target vowels except /æ/ and significantly higher performance for targets /æ/ and /eɪ/ than for /i,ɪ/ and /ɛ/ (p values ranged from <0.0005 to 0.004 for the significant

comparisons). Overall listener performance did not differ significantly between the target vowels /æ/ and /eɪ/ or among the target vowels /i,ɪ/ and /ɛ/. This order of performance for the target vowels did not differ dramatically across gates in the significant target vowel by gate interaction and therefore will not be discussed further in this section.

1. Group by duration condition effect

To address the question of whether the groups differed in their ability to benefit from duration information in identification of silent-center syllables, pairwise comparisons of performance on the DP and DN conditions at each level of the listener group variable were used to examine the significant listener group by duration neutralization interaction. Contrary to our hypothesis, only the MO listener group showed significantly higher performance (by about 4% or 4 RAU; $p = 0.039$) for the DP than for the DN condition. Although the performance of the EL group was also higher in the DP than in the DN condition by a similar amount (by about 3% or 4 RAU), the difference only approached significance for this group ($p = 0.087$). The performance of the LL group was *lower* in the DP than in the DN condition (by

about 3% or 4 RAU overall), but this difference was not significant ($p=0.113$).

A comparison of groups at each level of the listening condition variable was also made, using Bonferroni adjustment for the number of comparisons among groups at each level of duration neutralization. All groups differed significantly from one another in their performance on the DP condition. The MO listeners identified the target vowels significantly more accurately than the listeners in the other two groups ($p=0.05$ for MO versus EL and $p<0.0005$ for MO versus LL), and the EL listeners identified the target vowels significantly more accurately than the listeners in the LL group ($p<0.0005$). In the DN condition, listeners in both the MO and EL groups identified the target vowels significantly more accurately than the listeners in the LL group ($p<0.0005$ in both cases), but the MO and EL listener groups did not perform significantly differently from one another, although the difference did approach significance ($p=0.066$). In both the DP and DN conditions, the difference in performance between the MO and EL listener groups was 9% (or 10 RAU) or less, while the difference in performance between the LL listener group and the other two groups ranged from 25% to 40% (or 25 to 42 RAU) across the two listening conditions. In summary, the MO listeners were able to benefit to some degree from the vowel duration cues provided in the DP condition, but a similar magnitude benefit for the EL listeners failed to reach significance. The LL listeners showed no evidence that they were able to benefit from the vowel duration cues provided in the DP condition.

2. Group by gate effect

Figure 2 shows percent-correct performance for each listener group as a function of increasing gate duration. As shown in the figure, performance for the MO and EL groups increases by about 15% from the 10 to the 20 ms gate condition and to a lesser degree (about 4% and 9%, respectively) from the 20 to the 40 ms gate condition. Performance for the LL group, on the other hand, improves by only about 8% from the 10 to the 20 ms gate, and improves by about another 5% from the 20 to the 40 ms gate (when averaged across DP and DN). Thus, the EL group's vowel identification performance improves the most across the four gates (about 24%) and the LL group's performance improves the least (about 13%).

To determine the specific gates at which group differences were found, pairwise comparisons of performance were made between the groups at each level of the gate variable in the significant group by gate interaction. Bonferroni adjustment for the number of group comparisons at each level of gate was used. The MO listener group identified the syllables significantly more accurately than the EL group (by about 11% or 11 RAU) for the 10 ms gate condition only ($p=0.007$), although similar magnitude differences did not reach significance for the 30 and 20 ms gate conditions ($p=0.066$ and $p=0.117$, respectively). Both the MO and EL listener groups identified the syllables significantly more accurately than the LL group (by 21%–40% or 21–42 RAU) at all four gate conditions ($p<0.0005$ in all cases).

FIG. 4. Percent-correct identification performance by target vowel and gate, averaged across listener groups and duration conditions (DP and DN). Performance at the 10, 20, 30, and 40 ms gates is indicated by medium gray, white, light gray, and dark gray bars, respectively. Error bars indicate one standard error of the mean, and braces indicate sets of conditions for which performance did not differ significantly.

Pairwise comparisons of performance across the gates within each level of the listener group were used to determine the gate(s) for which performance differed significantly for each group. Again, Bonferroni adjustment for the number of comparisons at each level of group was used. All three listener groups showed similar patterns of significant effects across the gates. For the MO listener group, performance for the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate condition (by 14%–19% or 14–20 RAU; $p<0.0005$ in each case), but performance did not differ significantly among the 20, 30, and 40 ms gate conditions. For the EL listener group, performance on the 40 ms gate condition was significantly higher than the performance in the 10 and 20 ms conditions (by 9% and 24% or 9 and 24 RAU, respectively; $p=0.007$ and $p<0.0005$, respectively). Performance on the 20 and 30 ms gate conditions was also higher than the performance on the 10 ms condition for the EL listener group (by 21% and 16% or 21 and 15 RAU, respectively; $p<0.0005$ in both cases). The difference in performance between the 30 and 20 ms gate conditions (about 5% or 6 RAU) also approached significance ($p=0.061$) for the EL group. For the LL listener group, performance on the 30 and 40 ms gate conditions was significantly higher than the performance on the 10 ms gate condition (by 12% and 11% or 11 and 10 RAU, respectively; $p<0.0005$ and $p=0.005$, respectively), but no other comparisons reached significance.

3. Target vowel by gate effect

Although the three-way interaction between group, gate, and target vowel was not significant, the significant target vowel by gate interaction indicates that the six vowels exhibited different patterns of performance across the gates, but that this pattern did not differ significantly across the groups. Thus, Fig. 4 shows the performance for each target vowel on each of the four gate conditions and averaged across the DP and DN conditions and across listener groups. Pairwise comparisons of performance with Bonferroni adjustment for the

TABLE V. Percent-correct performance for each listener group on each target vowel for the DP and DN conditions, rounded to the nearest whole percentage. For each target vowel, the most frequent (i.e., those with greater than 5%) confusions, or vowels chosen instead of the target vowel are shown, with the confusion percentage in parentheses.

Listener group	Target vowel	DP		DN	
		Percent correct	Confusions (pct)	Percent correct	Confusions (pct)
MO	/i/	78	/eɪ/ (17)	81	/eɪ/ (15)
	/ɪ/	72	/eɪ/ (16), /ɛ/ (12)	59	/eɪ/ (28), /ɛ/ (13)
	/eɪ/	92	/ɛ/ (5)	91	/ɪ/ (5)
	/ɛ/	65	æ/ (26), /ɪ/ (5)	51	/æ/ (38), /ɪ/ (9)
	/æ/	89	/ɛ/ (8)	88	/ɛ/ (10)
	/ɑ/	99		100	
EL	/i/	60	/eɪ/ (20), /ɪ/ (18)	51	/ɪ/ (25), /eɪ/ (21)
	/ɪ/	65	/eɪ/ (20), /ɛ/ (11)	63	/eɪ/ (25), /ɛ/ (10)
	/eɪ/	76	/ɪ/ (12), /ɛ/ (8)	74	/ɪ/ (16), /ɛ/ (10)
	/ɛ/	57	æ/ (24), /ɪ/ (13)	52	/æ/ (30), /ɪ/ (12)
	/æ/	85	/ɛ/ (8)	85	/ɛ/ (8)
	/ɑ/	96		96	
LL	/i/	33	/ɪ/ (51), /eɪ/ (9), /ɛ/ (7)	32	/ɪ/ (51), /eɪ/ (9), /ɛ/ (6),
	/ɪ/	26	/ɛ/ (27), /eɪ/ (26), /i/ (18)	23	/eɪ/ (29), /ɛ/ (24), /i/ (20)
	/eɪ/	46	/ɛ/ (19), /ɪ/ (18), /i/ (15)	51	/ɛ/ (22), /ɪ/ (14), /i/ (12)
	/ɛ/	35	/æ/ (34), /eɪ/ (15), /i/ (6), /ɪ/ (6)	42	/æ/ (21), /eɪ/ (20), /i/ (12)
	/æ/	55	/ɛ/ (21), /ɑ/ (9), /eɪ/ (7)	58	/ɛ/ (22), /ɑ/ (10), /eɪ/ (8)
	/ɑ/	57	/æ/ (40)	66	/æ/ (29)

number of comparisons across the gates within each level of target vowel were used to explore the significant interaction between target vowel and gate. Performance did not differ significantly between any gates for targets /ɑ/ and /ɪ/, which showed the highest and second lowest rates of correct identification, respectively. As shown in the figure, performance is relatively flat across the four gates for these two target vowels.

For target /æ/, which showed the second highest overall rate of correct identification, performance on the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate (by 26%–30% or 27–31 RAU; $p < 0.0005$ in all three cases), but performance did not differ significantly across the three longer gates. Similarly, for target /eɪ/, performance on the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate (by 14%–21% or 12–20 RAU; $p < 0.0005$ to $p = 0.032$), but performance on the 40 ms gate was also significantly higher than the performance on the 20 ms gate (by 7% or 9 RAU; $p = 0.025$); no other gates differed significantly for this vowel. For target /i/, performance on the 30 and 40 ms gates was significantly higher than the performance on both the 10 and 20 ms gates (by 15–25 RAU or 13%–25%; $p < 0.0005$ to $p = 0.013$); no other gates differed significantly for this vowel. For target /ɛ/, for which the lowest overall level of performance was obtained, performance on the 20, 30, and 40 ms gates was significantly higher than the performance on the 10 ms gate (by 19–30 RAU or 21%–31%; $p < 0.0005$ to $p = 0.02$); no other gates differed significantly for this vowel (see Fig. 3).

D. Confusion analyses

The analysis of the confusion matrices was used to compare the groups in terms of the identity and number of vowels they perceived other than the target vowel in both the DP and DN listening conditions. Table V summarizes percent-correct performance and percent confusions for each target vowel, averaged across the four gates. Results are shown separately for each listener group and listening condition (DP versus DN). Within each group and for each target vowel, a percentage is given for each confusion vowel (vowel selected other than the target vowel) that received greater than 5% of listener responses for each target vowel.

As anticipated, the LL group showed a greater number of confusions than the other two groups; one to two more confusion vowels exceeded 5% confusions than for either the MO or EL group for all six target vowels and both listening conditions. Thus, the performance of the LL group was not only lower overall, but these listeners also appeared to be more uncertain of which vowel they had heard in the silent-center conditions. The EL group also showed one more confusion vowel exceeding 5% than for the monolingual group for two target vowels (/i/ and /eɪ/) in both listening conditions.

Although the order of confusions changed relatively little from the DP to the DN listening condition within each listener group, some interesting patterns did emerge. For the MO listener group, the greatest decreases in performance from the DP to the DN condition were seen for the target vowels /ɪ/ and /ɛ/ (13% and 14%, respectively), but the order of confusions did not change in either case. In fact, the pat-

tern changed for the MO listeners from the DP to the DN for only one target vowel (/eɪ/); for this condition, the only confusion vowel equal to or greater than 5% was /ɛ/ in the DP condition and /ɪ/ in the DN condition, which is not surprising considering the substantial shortening of /eɪ/ that occurred for the creation of the DN condition. Thus, although the MO listener group showed the greatest (and only significant) decrease in performance from the DP to the DN condition, the pattern of confusions for these listeners changed very little across the two conditions.

For the EL listener group, the greatest decreases in performance from the DP to the DN condition were seen for the target vowels /i/ and /ɛ/ (5% and 9%, respectively). In the case of target /i/, the most frequent confusion vowel changed from /eɪ/ to /ɪ/ from the DP to the DN condition, as would be expected for the removal of a duration cue to these two neighboring vowels and partially supporting the hypothesis stated in the Introduction that different confusion patterns might be seen for this listener group when duration was not available as a cue. No other confusion vowels change order from the DP to the DN condition for this listener group, however.

For the LL listener group, the order of confusions varied between the DP and DN conditions for only one target vowel (/i/), for which the most frequent confusion was /ɛ/ in the DP condition and /eɪ/ in the DN condition. These data do not support the hypothesis stated in the Introduction that different confusion patterns would be seen for this listener group when duration was not available as a cue. It is notable, however, that the most frequent *response* for target /i/ (not just the most frequent confusion) was /ɪ/ in both the DP and DN listening conditions. Figure 3(C) shows that percent-correct performance for target /i/ decreased by over 50% (from over 80% correct to less than 30% correct) from the whole-word to the DP 40 ms condition. These results suggest that the disruption caused by the silencing of the center made the target vowel /i/ unidentifiable for this listener group in either duration condition.

IV. DISCUSSION

The results of the present study generally support the research hypothesis that even relatively early Spanish L1 learners of English as a second language may have greater difficulty identifying L2 speech sounds based on partial acoustic information, compared to monolingual English speakers. Although the performance of the monolingual and early learner groups was nearly identical for the whole-word condition and did not differ significantly for the DP 40 ms condition, the EL listeners showed significantly lower performance overall on the silent-center conditions. In the *post hoc* analysis of the significant listener group by gate interaction, the performance of the MO listeners was significantly higher than that of the EL listeners for the 10 ms condition and approached significance for the 30 ms condition. Both the MO and the EL listeners performed with significantly higher accuracy than the LL listeners on most conditions, including the whole-word condition.

Furthermore, only the MO listener group appeared to be able to use the vowel duration information provided in the DP silent-center syllables effectively because this was the only group that showed significantly higher performance in the DP than in the DN listening condition. At first glance, this result would seem to differ from previous results that suggest that vowel duration information is often weighted as heavily or more heavily by speakers of English as a second language, even when vowel duration is not a contrastive cue in the L1 (Bohn, 1995). However, it is possible that the EL and the LL listeners *would* have made as much (or more) use of the duration information as the MO listeners in the intact syllable, but that they were unable to overcome the disruption of the syllable sufficiently to *use* the duration information effectively in the DP condition.

This interpretation would suggest that the phonemic representations of the EL listeners may not be substantively different from that of the MO listeners, but rather that their representations may be less *robust* than those of the MO listeners. In fact, the EL listeners did not perform significantly more poorly than the MO listeners in the DN condition; rather, their performance failed to *improve* significantly in the DP condition, although the difference did approach significance. The LL listeners, on the other hand, showed no evidence of improved identification rates from the DN to the DP condition and performed significantly more poorly than the other two groups in both the DP and DN conditions. Performance for the LL group also declined more dramatically than for the other two groups from the whole-word to the DP 40 ms condition; this result is partially accounted for by a much greater decline for this listener group for target vowel /i/ than for the other two groups. These data suggest that the LL listeners may have a reduced ability to recover from the loss of *any* target vowel information, compared to the other two groups. Once the initial disruption in the syllable occurred, however, the differences among the groups did not increase dramatically as the amount of information presented decreased (i.e., from the 40 to the 10 ms gate conditions).

A comparison of the number of confusions above 5% (shown in Table V) shows one to two more confusion vowels for the LL listeners than for the MO and EL listeners for all target vowels. Although the performance of the LL listeners on the whole-word condition is well above chance for each target vowel, the increased number of confusions in the silent-center conditions suggests that these listeners' category representations may be much more fragile than those of the other two groups, in that when the syllable was disrupted they appeared to be less certain of what vowel they heard than the MO and EL listeners.

The level of performance attained by the monolingual English-speaking listeners in the present study is similar to that of adult listeners in other studies of silent-center syllable perception, supporting the validity of the present data. The 30 ms DP and 30 ms DN conditions in the present study are most similar to two experimental conditions in Strange (1989) because (1) the target words in Strange (1989) were produced in a carrier phrase (although at a faster rate than the present study); (2) the initial and final portions in Strange

(1989) contained 26 and 34 ms of the CV and VC transitions, respectively; and (3) [Strange \(1989\)](#) also examined perception of syllables with DN silent centers. [Strange's \(1989\)](#) task did, however, employ more target vowels and a ten-alternative forced-choice task. Nevertheless, the performance by the present group of monolingual listeners and that of Strange's listeners is quite similar, with 96% correct performance on the whole-word data, 90% correct performance on the 30 ms DP stimuli, and a 4% reduction in performance on the 30 ms DN stimuli in the present study, compared to 98% correct whole-word performance, 84% silent-center performance, and a 5% reduction in performance on the DN stimuli in [Strange's \(1989\)](#) study. The present results also parallel those of [Strange et al. \(1983\)](#), in that the monolingual listeners in the present study showed greater decreases in identification performance for the short vowels (/i/ and /ε/) and the midlength vowel (/i/) than for the longer vowels (/ei, æ, and α/), despite the relatively long duration obtained for the "midlength" vowel /i/ found in the present study.

[Parker and Diehl \(1985\)](#) attempted to neutralize duration by the use of similar duration syllable triads as their alternatives and used four CV and VC transition durations (60%, 70%, 80%, and 90% vowel deletions, corresponding roughly to the 10, 20, 30, and 40 ms duration CV and VC conditions used in the present study). Despite the difference in number of alternatives in the forced-choice task [six in the present study and three in [Parker and Diehl \(1985\)](#)], the performance of the monolingual listeners in the present study was consistently within about 4%–8% of that of listeners in [Parker and Diehl's \(1985\)](#) study.

[Fox et al. \(1992\)](#) used a number of different syllable types and an open-set identification task to compare the performance of older and younger monolingual listeners on whole-word and silent-center stimuli. As in the present study, they found no significant difference between the listener groups in the performance on the whole-word condition but significantly lower performance (by about 7%) on the identification for older adults than for younger adults in the silent-center condition. In the present study, no significant difference between the performance of the MO and EL listeners was found in the whole-word condition, but the EL listeners' performance was consistently and significantly less accurate (by about 6%–10%) than that of the MO listeners across most of the silent center conditions. Thus, the effects of early learning of a second language and the effects of aging on the listeners' ability to identify syllables based on partial vowel information would appear to be similar in magnitude, although they may have quite different origins. This similarity is intriguing and direct comparison of silent-center syllable perception by non-native and elderly listeners may yield interesting results.

V. CONCLUSION

The results of the present study indicate that even relatively early learners of English as a second language may have more difficulty in identifying speech sounds based on partial acoustic information. This difference may therefore account for a portion of the increased difficulty that even

early learners of English appear to have in processing speech in noise and/or reverberation. The source of this difference at the phonetic level may lie in a reduced robustness in recovering from syllable disruption, less flexibility in switching perceptually to focus on the available speech cues when some cues are obscured, differences in phoneme boundaries or cue weighting, or other potential explanations. Whatever the reason for the differences observed in the present study between the native and non-native listeners' ability to identify the target vowels based on CV and VC formant transitions alone, it is likely that they only account for a portion of the difficulty that non-natives experience processing speech in difficult environments in the real world. As suggested by the results of both [Bradlow and Alexander \(2007\)](#) and [Cutler et al. \(2004\)](#), differences may well be observed between native and non-native listeners at each level of linguistic processing and from both top-down and bottom-up sources. Thus, further systematic comparisons of speech processing using different tasks and investigating different levels of processing are necessary to understand this problem. Furthermore, in light of the small but significant differences found in the present study for even early learners with little or no observable foreign accent, future research should detail language background variables carefully.

Despite the significant results, the small number of participants in each listener group is a limitation of the present study. Thus, replication and further analysis of the use of partial information for both vowels and consonants with larger numbers of listeners from varying L1s and L2s is necessary to generalize from the present data. Replication of the present study using consonants with lingual articulations (e.g., /d/ or /g/) would be particularly interesting because the effects of CV and VC transitions on syllable identification might be expected to be stronger for these consonants, due to the greater degree of coarticulation between consonant and vowel needed for consonants with lingual articulations.

Despite its limitations, the results of the present study suggest that later learners of a second language and perhaps even some early learners may benefit from perceptual training that may help listeners to develop greater flexibility in using alternate sources of phonetic information when some information is unavailable. Such training methods may be helpful in enabling non-native speakers to perform more like monolingual listeners in challenging listening environments such as noise or reverberation.

ACKNOWLEDGMENTS

This research was supported by NIH-NIDCD Grant No. 5R03 DC005561 to Catherine L. Rogers and by a University of South Florida Graduate Student Travel Award to Alexandra S. Lopez. We thank Stefan A. Frisch, Joseph Constantine, Gail S. Donaldson, and Theresa H. Chisolm for their helpful suggestions.

¹The RAU-transformed scale extends from –20 to 120, rather than 0% to 100%, but the scale is designed so that scores in the middle of the percent-correct range change very little when transformed to RAUs. The maximum possible performance (six out of six correct) translated to approxi-

mately 105 RAU in the present case, rather than the theoretical maximum RAU of 120.

²Data for two LL listeners were dropped in this analysis because raw data files were lost due to computer failure, but data for all eight LL participants were available for the analysis of gated conditions and overall data for the whole-word condition indicate similar performance on that condition for these two participants.

- Andruski, J. E., and Nearey, T. M. (1992). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables," J. Acoust. Soc. Am. **91**, 390–410.
- Bakeman, R. (2005). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods Instrum. Comput. **37**, 379–384.
- Boersma, P., and Weenik, P. (2006). *PRAAT: Doing phonetics by computer* (version 4.4.30), retrieved September 7, 2006 from <http://www.praat.org>.
- Bohn, O.-S. (1995). "Cross-language speech perception in adults: First language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore, MD), pp. 279–304.
- Bradlow, A. R., and Alexander, J. A. (2007). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," J. Acoust. Soc. Am. **121**, 2339–2349.
- Bradlow, A. R., and Bent, T. (2002). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.
- Clopper, C. G., and Pisoni, D. B. (2004). "Some acoustic cues for the perceptual categorization of American English regional dialects," J. Phonetics **32**, 111–140.
- COOLEDT 2000 (version 1.1) (2000). Syntrillium, Inc., Phoenix, AZ.
- Crawford, K. E. (2006). "The relationship between degree of foreign accent-ness and vowel perception of Spanish-English bilinguals," thesis, University of South Florida, Tampa, FL.
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.
- Dalbor, J. B. (1969). *Spanish Pronunciation: Theory and Practice* (Holt, Reinhart and Winston, New York, NY).
- ECOS/WIN (version 1.3) (1999). AVAAZ Innovations, Inc., London, Ontario.
- Flege, J. E. (1995). "Second language speech learning: Theory, findings and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore, MD), pp. 233–277.
- Flege, J. E., and MacKay, I. R. A. (2004). "Perceiving vowels in a second language," Stud. Second Lang. Acquis. **26**, 1–34.
- Fox, R. A., Wall, L. G., and Gokgen, J. (1992). "Age-related differences in processing dynamic information to identify vowel quality," J. Speech Hear. Res. **35**, 892–902.
- Hillenbrand, J., and Nearey, T. M. (1999). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.
- Hudgins, C. V., and Numbers, F. C. (1942). "An investigation of the intelligibility of the speech of the deaf," Genet. Psychol. Monogr. **25**, 289–392.
- IEEE (1969). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust., **AU-17**, 225–246.
- Egan, J. P. (1948). "Articulation testing methods," Laryngoscope **58**, 955–991.
- Imai, S., Walley, A. S., and Flege, J. E. (2005). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," J. Acoust. Soc. Am. **117**, 896–907.
- Jenkins, J. J., and Strange, W. (1999). "Perception of dynamic information for vowels in syllable onsets and offsets," Percept. Psychophys. **61**, 1200–1210.
- Jenkins, J. J., Strange, W., and Miranda, S. (1994). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**, 1030–1043.
- Jenkins, J. J., Strange, W., and Trent, S. A. (1999). "Context-independent dynamic information for the perception of coarticulated vowels," J. Acoust. Soc. Am. **106**, 438–448.
- Kent, R. D., Weismer, G., Sufit, G., Rosenbek, J. C., Martin, R. E., and Brooks, B. R. (1990). "Impairment in speech intelligibility in men with amyotrophic lateral sclerosis," J. Speech Hear. Disord. **55**, 721–728.
- Kewley-Port, D., and Goodman, S. S. (2005). "Thresholds for second-formant transitions in front vowels," J. Acoust. Soc. Am. **118**, 3252–3260.
- Kirk, K. I., Tye-Murray, N., and Hurtig, R. R. (1992). "The use of static and dynamic vowel cues by multichannel cochlear implant users," J. Acoust. Soc. Am. **91**, 3487–3498.
- Ladefoged, P. (1982). *A Course in Phonetics*, 2nd ed. (Harcourt, Brace, Jovanovich, New York).
- Lattner, S., Meyer, M. E., and Friederici, A. D. (2005). "Voice perception: Sex, pitch and the right hemisphere," Hum. Brain Mapp **24**, 11–20.
- Lindblom, B. (1996). "Role of articulation in speech perception: Clues from production," J. Acoust. Soc. Am. **99**, 1683–1692.
- Logan, J. S., Greene, B. G., and Pisoni, D. B. (1989). "Segmental intelligibility of synthetic speech produced by rule," J. Acoust. Soc. Am. **86**, 566–581.
- Mayo, L., Florentine, M., and Buus, S. (1997). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.
- Meador, D., Flege, J. E., and MacKay, I. R. A. (2000). "Factors affecting the recognition of words in a second language," Bilingualism: Lang. Cognit. **3**, 55–67.
- Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.
- Moncur, J., and Dirks, D. (1967). "Binaural and monaural speech intelligibility in reverberation," J. Acoust. Soc. Am. **10**, 186–195.
- Monsen, R. B. (1983). "The oral speech intelligibility of hearing-impaired talkers," J. Speech Hear. Disord. **48**, 286–296.
- Mullenix, J., Johnson, K., Topcu-Durgun, M., and Farnsworth, L. (1995). "The perceptual representation of voice gender," J. Acoust. Soc. Am. **98**, 3080–3095.
- Murphy, W. D., Shea, S. L., and Aslin, R. N. (1989). "Identification of vowels in 'vowelless' syllables by 3-year-olds," Percept. Psychophys. **46**, 375–383.
- Parker, E. M., and Diehl, R. L. (1985). "Identifying vowels in CVC syllables: Effects of inserting silence and noise," Percept. Psychophys. **36**, 369–380.
- Payton, K. L., Uchanski, R. M., and Braida, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.
- Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (2006). "Effects of bilingualism, noise and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. **27**, 465–485.
- SPSS (version 15.0) (2006). SPSS, Inc., Chicago, IL.
- Strange, W. (1989). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.
- Studebaker, G. (1985). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 494–509.
- Sussman, J. E. (2001). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?," J. Acoust. Soc. Am. **109**, 1173–1180.
- TDT SYSTEM III (2001). Tucker-Davis Technologies, Inc., Gainesville, FL.
- United States Census Bureau (2000). 2000 U.S. Census. Retrieved May 8, 2007, from <http://www.factfinder.census.gov>.