Issues in Clinical Linguistics
This volume of selected papers (not a regular volume of proceedings) derives from the conference "Linguistics Theory in Speech and Language Pathology and in Speech therapy", which was held in Padova in August 2000. This conference attracted international interest, because of its timely topic and interdisciplinary focus. Both participants and attenders greatly enhanced the value of the conference by discussing and contributing stimulating data and theoretical problems. The conference was organized with the support of a grant from the MURST in the framework of a national research project (Cofinanziamento Murst 1998-9810197024), coordinated by Alberto Mioni.
## CONTENTS

**Introduction**  
Elisabetta Fava and Alberto Mioni  

I Phonetics and Phonology  

A phonetic and phonological analysis of weak syllable omissions: comparing children with normally developing language and SLI  
Allyson K. CARTER  

A method for assessing the perception of vowels in a second language  
James Emil FLEGE  

Japanese speech errors: theoretical and clinical implications  
Haruko MIYAKODA  

One approach to phonological disorders from Clinical Phonology  
Mª Pilar OTERO CABARCOS  

II Grammar and Lexicon  

Agreement disorders in Broca's aphasia sentence production: a bilingual case study  
Paolo CHINELLATO  

The cognitive basis of syntactic development vs nativist arguments (evidence from Persian)  
Habibeh SAMADI  

The deviant lexicon approaching lexical processing in Williams syndrome  
Julia SIEGMÜLLER & Henrik BARTELS
A METHOD FOR ASSESSING THE PERCEPTION
OF VOWELS IN A SECOND LANGUAGE

James Emil FLEGE
Division of Speech and Hearing Sciences University of Alabama at Birmingham

1. Introduction

Individuals who learn a second language (L2) often speak it with a foreign accent. For example, studies examining the production of English by individuals who have learned English as an L2 have shown a wide range of phonetic divergences from the speech of monolingual native speakers of English. Foreign accents are evident in the sentences produced by individuals who began to learn English in childhood ("early" learners), but are usually much stronger in the speech of individuals who began to learn English in adolescence or adulthood ("late" learners). Age of learning effects have also been observed in the production of individual vowels. For example, late learners’ English vowels are more likely to be misidentified than early learners’ vowels are (Munro 1993; Munro et al. 1996).

The vowel errors observed in L2 speech production raise two important questions. Will L2 learners differ from native speakers of the L2 in perceiving L2 vowels? And, if so, do differences in perception cause differences in production? The results obtained by Flege et al. (1997) provided an affirmative answer to the first question. The results obtained by Rochet (1995) suggested that native versus non-native differences in perception may, in fact, cause differences in production. Rochet (1995) observed differences in the classification of a continuum of synthetic French vowels by native speakers of French, Portuguese and English. Vowels classified as a front rounded vowel (/y/) by the French participants were classified as a front unrounded vowel (/i/) by the Portuguese participants, but as a back rounded vowel (/u/) by the native English participants. The participants’ imitation of vowels paralleled the perceptual results. When asked to imitate tokens of French /y/, the Portuguese participants tended to produce /i/-
quality vowels whereas the native English participants tended to produce /u/-quality vowel. From this, Rochet concluded (1995: 404) that some L2 vowel production errors have a perceptual origin, that is, may be the “consequence of the target phones having been assigned to an L1 category.”

It order to further test the hypothesis offered by Rochet (1995), it will be necessary to develop valid and reliable tests of L2 segmental perception. There is, at present, uncertainty concerning how best to evaluate the segmental perception of L2 learners. The first aim of this article, therefore, was to review methods that have been used in previous research to assess L2 vowel perception. The second aim of the article was to present the results of an experiment using one of the perception techniques that was reviewed. It assessed the perception of English vowels by late learners of English who were native speakers of eight different languages (Korean, Japanese, Czech, Hungarian, Arabic, Portuguese, German and Dutch).

Presentation of the perception test results focused on the question of whether the test yielded stable results. Evaluating test stability was important because of the need to evaluate L2 perceptual learning from a longitudinal perspective. If the scores yielded by an L2 perception test increased on a second administration given shortly after the initial administration, or as the result of a modest amount of training, then the test would not be suitable for use in longitudinal research. This is because an increase in test scores obtained under such conditions would make it impossible to determine if the gains observed in a longitudinal design were due to an effect of test familiarization, or were due to perceptual learning.

2. Perceptual testing techniques

The perception of speech encompasses a variety of processes and multiple levels of structure. It is, therefore, insufficient to determine whether a group of L2 learners do or do not differ significantly from a group of native speakers. As discussed by Flege (1992a, 1992b, 1995), three specific questions must ultimately be addressed with respect to L2 perceptual learning.

The first question is whether each individual L2 learner is able to perceptually distinguish all vowels in the L2 inventory from one another. If the L2 possesses more vowels than the learner’s L1 does, it is likely that the learner will readily distinguish certain pairs of L2 vowels in the absence of any perceptual learning. That is, from the moment of first exposure to the L2, some L2 vowels will be easy to distinguish as the result of being identified as instances of two distinct L1 vowels.

It is unlikely, in the case just described, that the learner will perceive the L2 vowels exactly like individuals who are monolingual native speakers of the target L2. The vowel inventories of any pair of human languages are apt
A method for assessing the perception of vowels in a second language

to differ not only in terms of the number of contrastive vowels, but also in terms of the phonetic implementation of the vowels the two languages might be said to share. It is rarely the case that vowels found in two languages are truly identical. Even if the properties defining the L1 and L2 vowels have the same average values, the range of values possible for the L1 and L2 vowels, or the effect of contextual factors, may differ cross-linguistically.

This leads to the second important question that must be addressed in research examining the perception of L2 vowels (or consonants). When phonetic differences exist between L1 and L2 vowels, do non-natives detect them?

The final question that must be addressed is how L2 learners who have detected differences between L2 vowels and the closest L1 vowel(s) will represent the L2 vowel in long-term memory. One possibility is that information derived from the detection of L1-L2 phonetic differences will modify the long-term representation of the L1 vowel. For example, native Spanish adults who began to learn English as young children, and who perceive English vowels much like native English speakers (Flege 1992b), might represent English vowels as variants of Spanish vowels. For example, early Spanish-English bilinguals might represent English /i/ as a “low” Spanish /i/, and English /e/ as a Spanish /e/ produced with an unusually large amount of formant movement. Alternatively (Flege 1995), Spanish learners of English might establish new categories in long-term memory for certain English vowels.

**Indirect methods**

One way to evaluate the perception of L2 vowels (and consonants) is to use non-behavioral techniques that do not require conscious mediation or overt responses. For example, Winkler et al. (1999: 638) provided electrophysiological evidence that Hungarians who arrived in Finland between the ages of 13-32 years developed “cortical memory representations” for Finnish vowels not found in their L1 (viz., /e/ and /æ/) that permitted “preattentive” categorization of Finnish vowels not found in Hungarian.

The efficacy of using event-related potentials to study L2 speech acquisition was established by Näätänen et al. (1997), who found that automatic cortical change-detection responses reflected by mismatch negativity (MMN) patterns were substantially greater in response to changes from one L1 vowel to another than from an L1 vowel to a neighboring foreign vowel. Winkler et al. (1999) examined Finnish vowels (/e/, /æ/) located in a portion of the phonetic vowel space occupied by just one Hungarian vowel, /e/. Hungarian-Finnish bilinguals discriminated /e/-/æ/ as well as Finns did whereas Hungarian monolinguals responded at a chance level. Native Finnish speakers and Hungarian-Finnish bilinguals showed
larger magnitude MMNs when /e/ changed to /æ/ than the Hungarian monolinguals did. However, the native Finnish speakers, Hungarian-Finnish bilinguals and Hungarian monolinguals showed MMNs of similar magnitude when /e/ changed to /y/ (a vowel found in both Hungarian and Finnish).

Unfortunately, research involving event-related potentials is difficult to control, and is beyond the technical expertise of most L2 researchers. As a result, most previous research has made use of behavioral techniques that required overt responses from the participants being tested.

**Identification tests**

The most direct behavioral method for evaluating the perception of L2 vowels is to ask L2 learners to identify the L2 vowels. Percent correct identification scores might be obtained for all of the vowels in the L2 inventory. When misidentifications occur, the pattern of perceptual confusions may provide insight into how the L2 learners' perception differed from that of L2 native speakers. For example, the frequent misidentification by native Spanish speakers of English /i/ tokens as /i/, and /i/ tokens as /i/, would suggest that the two English vowels were confused with one another because Spanish has a single vowel, /i/, in the portion of vowel space occupied by English /i/ and /i/. A perceptual assimilation test (e.g. Strange 1995; Strange et al. 2001), if administered, might indicate that both English vowels tend to be identified as instances of the Spanish /i/ category, perhaps with a difference in degree of goodness of fit (e.g., Guion et al. 2000).

One important justification for obtaining perceptual identification data for L2 vowels (or consonants) is that such data can be readily related to segmental production data. For example, if a Spanish learner of English were found to identify naturally produced English /i/ tokens in 60% of instances, but 90% of her English /i/ productions were heard as intended by native English-speaking listeners, then one might reasonably conclude that the learner produced English /i/ better than she perceived this vowel. Should such evidence be obtained, it could be taken as counter-evidence to the hypothesis (e.g., Rochet 1995; Flege 1995) that L2 segmental perception will tend to "lead" production (i.e., that the production of an L2 phonetic segment can only be as accurate as the perceptual representation that has been developed for the segment).

Unfortunately, methodological problems have made it difficult for L2 researchers to obtain the kind of perceptual identification scores just described. A serious problem is finding labels that can be used unambiguously by L2 learners who are not thoroughly familiar with English vowel categories, spelling conventions, or phonetic symbols. Teaching L2 learners to use phonetic symbols constitutes a form of training that might alter the
A method for assessing the perception of vowels in a second language

object of investigation (the well-known “observer’s paradox”). Using orthographic labels is problematic because a many-to-one relation exists between letters and vowel quality in English. For example, ‘e’, ‘ea’, ‘ei’ and ‘ee’ can be used to spell the English vowel /i/. Phenomena such as this led Diehl et al. (1981) to doubt that participants can accurately report their perception of English vowels through the use of vowel symbols or letters.

Keywords might be used. However, when a vowel token is ambiguous, non-native participants might be biased to use the more familiar of two possible keywords (Flege et al. 1995; see also Gottfried 1984). Furthermore, even if difficulties concerning response alternatives can be resolved, important questions arise as to how many labels (keywords, symbols or letters) should be provided (see Beddor & Gottfried 1995 for discussion). The possibility for confusions—even by native speakers who are thoroughly familiar with the vowels being examined—increases as the number of response alternatives increases (Flege 1992a). Offering too many labels might bias participants away from the correct response. Offering too few labels, on the other hand, may force participants to under-report what they perceive.

The vowel perception test used by Horii et al. (1971) to assess the perception of English vowels by native English speakers obviated some of the problems just described. The test was developed to examine the “differentiation” of 12 English vowels presented in noise. The participants’ task was to identify each aurally presented word (e.g., ‘bead’ /bid/) from a written set of alternatives, of which one was the correct answer (e.g., ‘bed, bead, bid, bird, bad, bawd’). An important assumption of the test was that every possible response the participants may have wanted to give was included among the response alternatives provided. The test was found to provide stable scores after repeated exposures. Unfortunately, it may not be appropriate for use with nonnative speakers who are not thoroughly familiar with English spelling conventions (see above).

Problems associated with the choice of response labels are somewhat less serious for experiments in which just two response alternatives are provided, that is, two-alternative forced-choice (2AFC) experiments. However, other problems arise with this kind of test. For example, an increase in the use of one response alternative necessarily means the use of the other response alternative must decrease (see, e.g. Major 1987). This might lead to instances in which the perceptual learning of one L2 vowel creates the appearance of a deterioration in the perception of an adjacent L2 vowel. This might actually happen if the learner were developing a single new category that subsumed both L2 vowels. However, there would be nothing in the pattern of 2AFC data that would permit the investigator to determine if the response pattern were, or were not, an artifact of the testing technique.
Problems also arise for 2AFC experiments examining the identification of the members of a speech sound continuum. L2 learners might decisively partition the members of a synthetic continuum on a partly or completely auditory basis without phonetic knowledge of the two nominal continuum endpoints (Yamada & Tohkura 1992). One can infer such behavior when, for example, the phoneme boundary obtained from nonnative participants, but not for native speakers, is located at the exact midpoint of the continuum.

Other interpretative problems arise as well. For example, the native Spanish participants examined by Flege et al. (1997) identified the members of synthetic ‘beat-to-bit’ (/i-/u/) and ‘bet-to-bat’ (/e/-/æ/) continua. Spectral quality (F1 frequency) and vowel duration were varied orthogonally in the stimulus sets. The native Spanish participants were able to partition members of both continua in a 2AFC test. However, they may have done so differently than the native English controls did. Many native Spanish participants relied on vowel duration when identifying the stimuli from the first continuum as ‘beat’ or ‘bit’ rather than making use of both frequency and temporal cues as the native English speakers did. It is uncertain if these native Spanish participants tended to ignore the spectral cue because they did not discern the spectral quality difference between English /i/ and /u/, or because they gave greater weight to duration than the native English speakers did.

The native Spanish participants examined by Flege et al. (1997) differed less from the native English speakers for the ‘bet-to-bat’ continuum than they did for the ‘beat-to-bit’ continuum. However, this did not necessarily mean that the native Spanish participants perceived English /e/ and /æ/ more accurately than they perceived English /i/ and /u/. They may have judged members of the ‘bet-to-bat’ continuum as instances of the Spanish /e/ and /a/ categories, not as instances of the nominal continuum endpoints (viz. English /e/ and /æ/). If so, the decisive partitioning of the /e/-/æ/ continuum did not mean that the native Spanish participants possessed long-term memory representations for English /e/ and /æ/.

The problems associated with orthography and with the use of a 2AFC might be circumvented by training L2 learners to transcribe the vowels of their L2 using phonetic symbols such as those proposed by the International Phonetic Association (IPA). Individuals trained in this way could simply be asked to transcribe L2 lexical items presented via a loudspeaker rather than being asked to write the words using conventional orthography (or label them using keywords). However, the use of IPA symbols would engender other interpretative problems. For example, a native Spanish speaker’s failure to differentiate English /i/ and /u/ might have a perceptual basis (e.g., the lack of distinct perceptual representations for the two English vowels). However, it might also be due to inadequate phonetic transcription training.
Conversely, accurate responses might be attributed to an auditory rather than to a phonetic level of analysis. The ability of phoneticians to transcribe vowels in an unknown foreign language derives from an acquired ability to auditorily relate foreign phones to known referents in long-term memory (often vowels in their L1, less often cardinal vowels), or to identify component features of the phones (e.g., transcription of a phone as [ɸ] through the detection of bilabiality, frication and voicelessness). However, neither process typifies normal phonetic perception. Finally, and perhaps most importantly, the L2 vowel perception of a phonetically trained L2 learner might not generalize to other L2 learners who have not received extensive phonetic training.

Discrimination tests

The problems associated with identification tests are largely circumvented by discrimination tests, but at the cost of ecological validity. Discrimination refers to a listener’s ability to perceptually distinguish two phones. In everyday speech perception, listeners recognize words based on their overall spectral pattern, supplemented (when necessary) but a more fine-grained analysis of syllable- and segment-sized units. Listeners rarely if ever recognize a word by discriminating a phonetic segment in it from a phonetic segment in another word they have just heard. An important interpretative problem for discrimination tests is that the results they yield almost always reflect an auditory component instead of, or in addition to, information stored in long-term memory representations.

One method that has been used to assess the perception of vowels in an L2 is the same/different (AX) discrimination test. For example, Politzer and Weiss (1969) presented tape-recorded tokens of minimally paired French words (e.g., ‘basse’, ‘bosse’), or two exact repetitions of one member of the pair, to native English-speaking children ranging in age from 6-14 years. The percent correct discrimination scores were found to increase steadily with age. L2 vowels are known to be produced with an increasing degree of divergence from L2 phonetic norms as the age of L2 learning increases (Munro et al. 1996). Thus, if the Politzer and Weiss finding were accepted at face value, it would indicate that a relation does not exist between L2 vowel production and perception.

It is more likely, however, that the AX procedure used by Politzer and Weiss (1969) did not provide a *phonetically relevant* indication of L2 vowel perception. If so, concluding that a relation does not exist between L2 vowel production and perception is unwarranted. Vowels are inherently more discriminable than consonants are (see e.g. Stevens et al. 1969; Strange 1995). It is unlikely that the children examined by Politzer and Weiss were making
phonetic judgments concerning the *categorical* identity of the French vowels they were asked to discriminate (see Shigeno 1992). The older children may have been better able than the younger children were to attend to the task, and to make use of auditory information present in the stimuli.

The AX procedure has two serious problems as a test of L2 segmental perception. When two phones are presented for same/different judgments at a short inter-stimulus interval, both phones can be held in auditory short-term memory, and judgments can be made on the degree of mismatch (if any) of the two auditory representations. This problem is further exacerbated by the fact that any auditorily detected difference, regardless of its relevance for distinguishing pairs of speech sounds, can provide the basis for correct “different” judgments. For example, a native Spanish might correctly judge an English /i/ token and an English /I/ token to be different in an AX test as the result of detecting a difference in loudness rather than, say, detecting a phonetically relevant difference in F1 frequency or vowel duration.

The problems just described can be reduced, to some extent, through the use of a categorial test. A “categorial” test employs multiple natural tokens of each speech sound category of interest (e.g., Gottfried et al. 1985). Such a technique encourages participants to respond in a general rather than a token-specific manner. The use of a categorial procedure changes the nature of the research question being asked. The question is no longer whether the participant can detect a physical (i.e., auditory) difference between a pair of stimuli but, rather, whether she judges them to be instances of different *categories*. In other words, the participants in a categorial test are called upon to discriminate stimuli based on “name” identity rather than physical identity. For example, five tokens of English /i/ (produced by a single talker or by five different talkers) and a five tokens of English /I/ might be presented in categorial AX discrimination test. The correct response to /i/₁–/i/₅ and /I/₁–/I/₅ trials (where subscripts denote different talkers or tokens) would be “same”, even though acoustic differences might be detected auditorily. A problem still exists, however. A correct “different” response to an /i/₁–/I/₅ trial or to an /I/₁–/I/₅ trial might be based either on a detection of phonetically relevant dimensions, or on the detection of dimensions that are not phonetically relevant.

In some discrimination tests, more than two stimuli are presented on each trial. In a “triadic” test, for example, three stimuli are presented. One important benefit of a triadic test is that it forces the participant to hold two stimuli in auditory short-term memory while perceptually appraising a third stimulus. This decreases the likelihood of auditory-based judgments, and encourages judgments based on the comparison of phonetic codes, especially if a relatively long inter-stimulus interval is used.
Polka (1995) used a categorial AXB test to assess vowel perception. Tokens presented in the A and B positions of each trial were drawn from a different category. The participants' task was to determine if the X token was "the same" as either the A token or the B token. The three vowels presented on each trial were always physically different. Polka examined native English participants' discrimination of two pairs of German vowels. Both pairs contained one German back rounded vowel (/u/ or /u/) that was similar to a vowel in English, and one German front rounded vowel without an English counterpart (/y/ or /y/). English monolinguals discriminated the tense vowels /y/-/u/ at a near-perfect rate but they discriminated the lax vowels /u/-/y/ poorly.

Polka (1995) interpreted her findings within the framework of Best's Perceptual Assimilation Model (e.g. Best 1995; Best et al. 2001). According to this model, the native English participants discriminated /y/-/u/ well, even though instances of both German vowels were likely to have been classified as instances of English /u/, because the German /u/ tokens were judged to be better instances of English /u/ than the German /y/ tokens were. That is, the German /u/ tokens may have been coded as 'good' English /u/ tokens while the German /y/ tokens were coded as 'poor' English /u/ tokens. The German /u/ and /y/ tokens, which were discriminated poorly, were not distinguished by a difference in category goodness.

Gottfried (1984) used a categorial ABX test to evaluate the discrimination of eight pairs of French vowels by three groups: native speakers of French living in the United States, English monolinguals, and native English speakers of French. The participants' task was to decide if the X stimulus in each trial was the same as the A stimulus or the B stimulus. (Vowels presented in the A and B positions were always drawn from different categories.) The inter-stimulus interval between the three stimuli in each trial (always spoken by different talkers) was 1.0 sec. As might be expected, the English monolinguals made more errors than the native English speakers of French (M = 34% vs. 27%), who made more errors then the native French speakers did (M = 22%).

One surprising finding reported by Gottfried (1984) was that the native English speakers of French made so many errors (M = 38%) discriminating the French vowels /i/ and /e/. One might have expected them to identify these vowels as instances of two different English vowels (/i/ and /e/), and thus to have obtained high discrimination scores (see above). Perhaps the native English speakers of French did not identify the French /e/ tokens as instances of English /e/ because the French /e/ tokens showed less formant movement than English /e/ typically does. It was notable that the native French speakers also made many errors (M = 37%) discriminating French...
/i/-/e/. In fact, there was a correlation between the native French speakers’ error rates, on the one hand, and those obtained for the native English speakers of French, $r(6) = 0.81$, $p < 0.05$, and the English monolinguals, $r(6) = 0.69$, $p = 0.06$. A possibility suggested by the findings of Flege et al. (1999) for late Italian-English bilinguals who had lived in Canada for many years is that the French speakers’ perception of French vowels had changed as the result of using English for many years. Alternatively, the load imposed on memory by the ABX format may have been sufficient to cause errors in the perception of even L1 vowels (see Beddor & Gottfried 1995).

In AXB and ABX tests the serial position of the vowel to be compared to instances of two distinct categories—the vowel in the ‘X’ position—is fixed. This reduces task uncertainty. Oddity tasks are similar to AXB and ABX tests in that the participant must identify the one stimulus that differs from all of the remaining stimuli. However, the position of the “odd item out” is not fixed. This increases task uncertainty, and thus the overall difficulty of the task.

The stimuli used in an oddity task by Weiher (1975) were tape-recorded sequences of four non-words (e.g. /æmps/ /æmps/ /æmps/ /æmps/) spoken by a native English speaker. The error rates obtained for German 10- and 11-year olds were surprisingly low, even for pairs of English vowels (e.g. /æ/, /e/) that one might expect to be identified in terms of a single German vowel (see, e.g., Bohn & Flege 1990). Perhaps the participants’ young age contributed to their success.

Strain (1963) modified the standard oddity procedure. Sequences of English sentences (e.g., ‘Give me a bit ... Give me a beet ... Give me a bit’) were presented live-voice to native Japanese college students in Japan. The participants circled ‘1’, ‘2’ or ‘3’ on the answer sheet to indicate the serial position of the odd item out, if they heard one. They were instructed to circle ‘0’, however, if they did not hear an odd item out (i.e., one vowel that differed from the other two vowels). Apparently, no trials consisted of three instances of a single vowel. The ‘0’ response seems to have been included to permit the participants to report not hearing a difference between the stimuli, if this happened to be the case. Perhaps due in part to the inclusion of a ‘0’ response alternative, the native Japanese participants’ correct detection rates were low for nearly all of the English vowel contrasts examined.

Flege et al. (1999; see also Guion et al. 2000) used an oddity discrimination format that included no-change as well as change trials. The correct response to a change trial such as /ɪ/–/ɪ/–/ɪ/ was “2” (the serial position of the odd item out). The correct response to no-change trials such as /ɪ/–/ɪ/–/ɪ/ or /ɪ/–/ɪ/–/ɪ/ was “none” (i.e., no odd item out). The decision to include both change and no-change (or “catch”) trials consisting of three
A method for assessing the perception of vowels in a second language

physically different tokens of a single category was motivated by the view (e.g., Liberman 1957; Kuhl 1980; Guenther et al. 1999) that two things happen when a phonetic category is established during language acquisition. There will be an increase in sensitivity to differences between instances of the newly formed category and other categories; and there will be a reduction in sensitivity to token-to-token variation within the newly formed category. The no-change trials were included to encourage the participants to respond only to phonetically relevant differences, not to any auditorily detectable difference between the stimuli.

A disadvantage associated with the use of the ABX (or AXB) format is that participants can guess the correct answer in 50% of instances. However, an advantage of these testing formats is that they minimize response bias. Guessing will yield the same error rate regardless of whether the participant has a propensity to guess that the variable (X) stimulus occurred in the A or the B position. Flege et al. (1999) reduced response bias in their oddity-format categorial discrimination test by including an equal number of change and no-change trials. The dependent variable, A' (a-prime), was based on the proportion of both hits and false alarms. “Hits” were defined as the proportion of correct detections of the odd item out in change trials; “false alarms” were defined as the proportion of incorrect selections of an odd item out in no-change trials that did not actually include an odd item out.

3. A categorial discrimination test

Introduction

The categorial discrimination test (CDT) used by Flege et al. (1999) was used here to assess the perception of English vowels by individuals who had learned English as an L2. The experiment had two aims. The first aim was to provide preliminary evidence concerning which English vowel contrasts would prove difficult for native speakers of Korean, Japanese, Czech, Hungarian, Arabic, Portuguese, German and Dutch to discriminate. The second aim was to assess test stability. Two weeks after taking the CDT for the first time, most of the participants returned to receive training on vowel contrasts that had been relatively difficult for them to discriminate during the first administration. Immediately following the training, they were given the CDT for a second time. The question of interest was whether the scores obtained for the trained contrasts, the untrained contrasts, or both would be higher on the second than the first administration of the CDT.
Methods

Participants. The 48 non-native participants had studied English at school before coming to the United States, but nearly all of them received their first extensive exposure to spoken English upon arriving in the United States after the age of 13 years. These participants (27 males, 21 females) were all living in Birmingham, Alabama at the time of the study. In addition, a group of 10 native English speakers (3 males, 7 females) with a mean age of 31 years were also tested in Birmingham. All 58 participants passed a pure-tone hearing screening before taking the CDT.

As shown in Table 1, the non-native participants were 28 years of age, on the average, had lived in the United States for an average of 7.8 years, and reported using English 67% of the time. As expected from previous research, the participant variables just mentioned were correlated. The later the participants had arrived in the United States, the less often they tended to use English according to self-report, \( r(46) = -0.31, p < 0.05 \), and the older they tended to be at the time of testing, \( r(46) = 0.45, p < 0.01 \). Also, the longer the participants had lived in the United States, the more often they tended to use English, \( r(46) = 0.49, p < 0.01 \), and the older they tended to be at the time of testing, \( r(46) = 0.81, p < 0.01 \).

The eight L1-defined subgroups were heterogeneous. For example, the native German participants arrived in the United States at an earlier age than the native Japanese participants had (\( M = 22 \) vs. 34 years of age). To take another example, the Czech participants had lived longer in the United States (\( M = 16.5 \) years) than either the native Hungarian or Korean participants had (\( M = 1.1 \) and 2.2 years, respectively). The heterogeneous nature of the eight non-native groups meant that it was not possible to separate the influence of differences in the L1 and L2 vowel systems from language experience. For example, a difference in the discrimination of English /u/-/u/ by the German and Japanese subjects might be attributed either to a difference in the likelihood of the two English vowels being identified with two distinct L1 vowels (much more likely for the German than Japanese participants), to differences in length of residence in the United States (much longer for the German than Japanese participants), or to some combination of both.
A method for assessing the perception of vowels in a second language

Table 1. Characteristics (mean, SD, range) of the native speakers of eight languages (N = 6 per group) who participated.

<table>
<thead>
<tr>
<th></th>
<th>AOA</th>
<th>LOR</th>
<th>% Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech</td>
<td>30(5)</td>
<td>16.5(5.5)</td>
<td>78(14)</td>
</tr>
<tr>
<td></td>
<td>21-35</td>
<td>7.0-21.0</td>
<td>65-100</td>
</tr>
<tr>
<td>Japanese</td>
<td>34(3)</td>
<td>0.8(0.3)</td>
<td>47(9)</td>
</tr>
<tr>
<td></td>
<td>30-39</td>
<td>0.3-1.0</td>
<td>33-57</td>
</tr>
<tr>
<td>Hungarian</td>
<td>28(4)</td>
<td>1.1(1.6)</td>
<td>64(31)</td>
</tr>
<tr>
<td></td>
<td>25-35</td>
<td>0.1-4.0</td>
<td>10-98</td>
</tr>
<tr>
<td>Dutch</td>
<td>27(4)</td>
<td>17.4(18.4)</td>
<td>87(19)</td>
</tr>
<tr>
<td></td>
<td>23-33</td>
<td>0.6-38.0</td>
<td>60-100</td>
</tr>
<tr>
<td>Portuguese</td>
<td>27(4)</td>
<td>5.2 (4.8)</td>
<td>52(19)</td>
</tr>
<tr>
<td></td>
<td>22-34</td>
<td>1.0-14.0</td>
<td>18-68</td>
</tr>
<tr>
<td>German</td>
<td>22(6)</td>
<td>12.0(14.6)</td>
<td>85(21)</td>
</tr>
<tr>
<td></td>
<td>14-32</td>
<td>1.0-34.0</td>
<td>47-100</td>
</tr>
<tr>
<td>Korean</td>
<td>27(3)</td>
<td>2.2(1.2)</td>
<td>41(25)</td>
</tr>
<tr>
<td></td>
<td>23-31</td>
<td>1.0-4.0</td>
<td>0-77</td>
</tr>
<tr>
<td>Arabic</td>
<td>27(13)</td>
<td>7.2(3.1)</td>
<td>80(13)</td>
</tr>
<tr>
<td></td>
<td>17-52</td>
<td>3.0-11.0</td>
<td>62-92</td>
</tr>
</tbody>
</table>

M = 28 7.8 67

Note: AOA, age of arrival in the United States, in years; LOR, length of residence in the United States, in years; % Use, self-reported percentage use of English.

There was also a substantial amount of intra-group variability among participants in the L1-defined subgroups. For example, the native Portuguese participants had lived in the United States from 1-14 years, and reported using English 18-68% of the time. It is likely, therefore, that some of the native Portuguese participants had received considerably more English input than others had. If experience in English leads to a better discrimination of L2 vowels, one would expect a substantial amount of intra-group variability in the Portuguese participants' discrimination scores.

Stimuli. Multiple natural tokens of the English /i/, /u/, /e/ V, /æ/, /er/, /a/ /o/ /æ/, /u/, /er/, /a/ and /o/ served as stimuli. Five adult male native speakers of American English, all from the Midwest, recorded these 12 vowels in a /b Vt! context at the end of a carrier phrase ("I say a."). All but one of the /b Vt! stimuli (viz. /but/) was a real word. The middle of three available tokens of each item was digitized (10.0 kHz, 12-bit resolution) and normalized for peak intensity. Prevoicing, if present, was removed from the word-initial /b/. The word-final /t/ release bursts, produced by all five talkers, were left intact.
James Emil Flege

The duration values of the 12 vowels were measured. The between-vowel duration differences were those expected for citation-style speech (e.g. Peterson & Lehiste 1960). For example, the /i/ tokens were slightly longer on average than the /u/ tokens were (M = 150 vs. 131 msec), the /e/ tokens were longer than the /e/ tokens were (M = 191 vs. 160 msec), and the low vowel /a/ was longer than the high vowel /u/ was (M = 211 vs. 149 msec). The frequencies of the first two vowel formants (F1, F2) in the test stimuli were measured using linear predictive coding (LPC) analysis at three locations: the acoustic midpoint, 25% into the vowel, and 75% into the vowel. As expected, the diphthongs /au/ and /au/ showed considerable formant movement, and the mid vowels /e/ and /o/ showed more movement than the remaining simple vowels.

Procedures. The stimuli were presented in triads to test the 14 vowel contrasts shown in Table 2. Three contrasts (/o/-/au/, /æ/-/æ/, /æ/-/æ/) each included one diphthong. All of the nonnative speakers of English were expected to discriminate these contrasts well because the vowels in these contrasts could be coded differentially as one-vowel and two-vowel sequences. The other contrasts involved pairs of vowels that are close in vowel space. Previous research with adult native speakers of English have shown that these vowels are confused occasionally even by native speakers of English (e.g., Peterson & Barney 1952; Verbrugge et al. 1976) and show relatively long reaction times in a speeded AX discrimination test (McCabeley 1984). Nine of these contrasts involved a one-feature difference (/u/-/u/, /i/-/i/, /e/-/e/, /i/-/i/, /i/-/i/, /o/-/o/, /æ/-/æ/, /æ/-/æ/, /æ/-/æ/), one involved a two-feature difference (/u/-/u/), and one involved a three-feature difference (/æ/-/æ/).

Table 2. The 14 English vowel contrasts that were examined. Contrasts 1-7 comprised Part 1 and contrasts 8-14 comprised Part 2.

<table>
<thead>
<tr>
<th>#</th>
<th>Vowels (keywords)</th>
<th>#</th>
<th>Vowels (keywords)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>/i/-/e/ (beat-bet)</td>
<td>8</td>
<td>/u/-/u/ (book-boot)</td>
</tr>
<tr>
<td>2</td>
<td>/i/-/u/ (beat-bit)</td>
<td>9</td>
<td>/u/-/u/ (book-but)</td>
</tr>
<tr>
<td>3</td>
<td>/e/-/e/ (bet-bait)</td>
<td>10</td>
<td>/o/-/u/ (boat-boot)</td>
</tr>
<tr>
<td>4</td>
<td>/æ/-/æ/ (bet-bit)</td>
<td>11</td>
<td>/o/-/æ/ (boat-bout)</td>
</tr>
<tr>
<td>5</td>
<td>/æ/-/æ/ (bet-bit)</td>
<td>12</td>
<td>/æ/-/æ/ (bottle-book)</td>
</tr>
<tr>
<td>6</td>
<td>/æ/-/æ/ (bat-bottle)</td>
<td>13</td>
<td>/æ/-/æ/ (bottle-but)</td>
</tr>
<tr>
<td>7</td>
<td>/æ/-/æ/ (bat-bite)</td>
<td>14</td>
<td>/æ/-/æ/ (bottle-bout)</td>
</tr>
</tbody>
</table>
A method for assessing the perception of vowels in a second language

Each vowel contrast was tested by eight change and eight no-change trials. For example, the /i/-/u/ contrast was tested by eight trials such as /i/₁-i/₃/i/₄ and /i/₃/i/₁/i/₄ (half with an /i/ token as the odd item out and half with an /u/ token as the odd item out) and eight trials such as /i/₁-i/₃-i/₄ and /i/₃/i/₁/i/₄. The three tokens presented on each trial were always spoken by a different individual. In change trials, the odd item out occurred in the three possible serial positions with near-equal frequency.

The test trials were recorded (Technics Model M235X) for later off-line presentation. The inter-stimulus interval between the three stimuli in each trial was 1.3 sec, and the inter-trial interval was fixed at 2.8 sec. The two parts of the test (Parts A and B, see Table 2), which each contained seven contrasts, were presented in counterbalanced order. The stimuli were presented at a comfortable level over headphones to the participants, who were tested individually in a sound booth. The participants were told to circle '1', '2' or '3' on an answer sheet to indicate the position of an odd item out, if they heard one. They were told to circle a fourth response, 'none', if they heard three instances of a single vowel.

The participants were given practice on a non-test contrast (/i/-/u/) before the experiment began. After a familiarization block in which the correct responses for eight trials were pre-marked on the answer sheet, the participants were given practice without feedback. They were required to respond correctly to eight trials in a row before beginning the experiment. Most of the participants, including the non-natives, met this requirement on the first practice block. This is because the English /i/ and /u/ stimuli were likely to have been coded as instances of high front and high back vowels in the non-native participants’ LI, thereby permitting a high level of discrimination.

An A' score was calculated for each contrast using the formula provided by Grier (1971). These scores were based on the proportion of "hits" (i.e., the number of times an odd item was correctly selected in different trials; maximum 8), and the proportion of "false alarms" (i.e., the number of incorrect selections of an odd item out in catch trials; maximum 8). The formula used was:

\[
A' = 0.5 + \frac{(\text{HIT} - \text{FA}) \times (1 + \text{HIT} - \text{FA})}{(4 \times \text{HIT}) \times (1 - \text{FA})}
\]

An A' score of 1.00 indicated perfect sensitivity (i.e., correct responses to all 8 change and all 8 no-change trials) whereas a score of 0.5 indicated a lack of sensitivity.
Results

Between-group differences. The mean A' scores obtained for the eight groups of L2 learners and the native English controls are shown in Table 3 (Part A) and Table 4 (Part B).

The native English speakers obtained an average score of 0.99 for the 14 contrasts, indicating near-perfect sensitivity for each contrast. The average scores obtained for the eight non-native groups ranged from 0.82, for the native Portuguese speakers, to 0.95 for the native Arabic speakers. There was considerable variation among the scores obtained for the 14 contrasts for most non-native groups. For example, the scores obtained by the native Japanese participants ranged from 0.49, for /u/-/u/, to 0.98 for both the /i/-/e/ and /e/-/e/ contrasts. To take another example, the native Korean participants' scores ranged from 0.55, for /u/-/u/, to 0.99 for both the /o/-/u/ and /e/-/e/ contrasts.

As mentioned earlier, the heterogeneity of the non-native groups made it impossible to draw conclusions as to why certain L1 groups obtained higher scores than other groups did, or to draw inferences regarding the influence of language experience.

To illustrate this point, Pearson product-moment correlations were computed between the subject variables shown in Table 1 and the average A' score obtained for each L2 learner. The correlation between the average discrimination scores and age of arrival in the United States was non-significant, \( r(46) = -0.14, p > 0.10 \), as was the correlation with length of residence in the United States, \( r(46) = 0.26, p > 0.05 \). A modest correlation was obtained, however, between the discrimination scores and self-reported percentage use of English, \( r(46) = 0.41, p < 0.01 \).

The correlation just reported might be taken as evidence that a frequent use of the L2 contributes to a relatively accurate perception of L2 vowels. Although this may be true, it is possible that the correlation arose from a confound between L2 use and L1 background. Many of the nonnative participants who tended to use English frequently were speakers of L1s that have large vowel systems that are likely to engender the mapping of the two English vowels in a contrast onto two distinct L1 vowels. For example, the Dutch participants had lived far longer in the United States than the Japanese participants had (\( M = 17.4 \) vs. 0.8 years of residence). The Dutch participants were also more likely than the Japanese participants were to map two English vowels in a contrast onto two distinct L1 vowels. This is because Dutch has more vowels than Japanese does (13 vs. 5 non-diphthongal vowels).
A method for assessing the perception of vowels in a second language

Table 3. Mean discrimination scores obtained for the seven vowel contrasts in Part A. Contrasts that were especially difficult for the subjects in each L1 group are boldfaced (see text).

<table>
<thead>
<tr>
<th></th>
<th>/i/-/E/</th>
<th>/i/-/l/</th>
<th>/e/-/E/</th>
<th>/E/-/u/</th>
<th>/E/-/o/</th>
<th>/o/-/A/</th>
<th>/o/-/a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng.</td>
<td>0.99(0.01)</td>
<td>0.99(0.02)</td>
<td>1.0(0.01)</td>
<td>0.98(0.03)</td>
<td>0.98(0.02)</td>
<td>1.0(0.00)</td>
<td>1.0(0.00)</td>
</tr>
<tr>
<td>Cze.</td>
<td>0.99(0.01)</td>
<td>0.97(0.03)</td>
<td>0.99(0.02)</td>
<td>0.81(0.16)</td>
<td>0.78(0.06)</td>
<td>0.82(0.27)</td>
<td>0.96(0.05)</td>
</tr>
<tr>
<td>Jap.</td>
<td>0.98(0.03)</td>
<td>0.89(0.13)</td>
<td>0.98(0.03)</td>
<td>0.78(0.06)</td>
<td>0.62(0.24)</td>
<td>0.65(0.26)</td>
<td>0.94(0.04)</td>
</tr>
<tr>
<td>Hun.</td>
<td>0.98(0.03)</td>
<td>0.93(0.07)</td>
<td>1.0(0.00)</td>
<td>0.81(0.09)</td>
<td>0.71(0.19)</td>
<td>0.92(0.06)</td>
<td>0.96(0.03)</td>
</tr>
<tr>
<td>Dut.</td>
<td>0.99(0.01)</td>
<td>0.98(0.02)</td>
<td>0.99(0.01)</td>
<td>0.93(0.05)</td>
<td>0.89(0.04)</td>
<td>0.94(0.06)</td>
<td>0.97(0.03)</td>
</tr>
<tr>
<td>Port.</td>
<td>0.91(0.10)</td>
<td>0.84(0.10)</td>
<td>0.96(0.05)</td>
<td>0.83(0.16)</td>
<td>0.52(0.22)</td>
<td>0.85(0.18)</td>
<td>0.94(0.05)</td>
</tr>
<tr>
<td>Ger.</td>
<td>0.99(0.02)</td>
<td>0.98(0.02)</td>
<td>0.98(0.03)</td>
<td>0.93(0.06)</td>
<td>0.76(0.19)</td>
<td>0.93(0.05)</td>
<td>0.94(0.05)</td>
</tr>
<tr>
<td>Kor.</td>
<td>0.96(0.03)</td>
<td>0.72(0.17)</td>
<td>0.99(0.02)</td>
<td>0.75(0.12)</td>
<td>0.60(0.22)</td>
<td>0.95(0.08)</td>
<td>0.98(0.03)</td>
</tr>
<tr>
<td>Ara.</td>
<td>0.99(0.01)</td>
<td>0.99(0.01)</td>
<td>0.99(0.02)</td>
<td>0.86(0.05)</td>
<td>0.91(0.07)</td>
<td>0.99(0.01)</td>
<td>0.99(0.01)</td>
</tr>
<tr>
<td>M</td>
<td>0.98(0.04)</td>
<td>0.93(0.11)</td>
<td>0.99(0.02)</td>
<td>0.86(0.12)</td>
<td>0.77(0.21)</td>
<td>0.90(0.16)</td>
<td>0.97(0.04)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

The following procedure was adopted to identify vowel contrasts that posed special difficulty for the participants in each non-native group. A series of one-way ANOVAs indicated that the effect of Vowel Contrast (14 levels) was significant for each non-native group (p < 0.01). Tukey’s post-hoc tests with an alpha level of 0.05 were then carried out to identify significant pair-wise differences between vowel contrasts. An English vowel contrast was identified as being “difficult” for a particular non-native group if it received a significantly lower score than at least two other vowel contrasts did. The difficult contrasts have been bold-faced in Tables 3 and 4.

The pattern of difficult contrasts varied across the L1-defined subgroups. For example, the /u/-/u/ contrast was difficult for the native Japanese and Korean participants but not for the native Czech or Portuguese participants. Conversely, the /a/-/a/ contrast was difficult for the native Czech and Portuguese participants but not for the native Japanese or Korean participants. Additional research will be needed to determine why particular English contrasts were difficult for speakers from various L1 backgrounds. It seems likely, however, that a contrast was usually difficult because both English vowels in it tended to be identified as instances of a single L1 vowel category (Best 1995). For example, the English /u/-/u/, /i/-/i/ and /e/-/ae/ contrasts were likely to have been difficult for the native Korean participants because the two English vowels in each contrast occupy a portion of vowel space where only one contrastive vowel exists in Korean (Yang 1996).
Table 4. Mean discrimination scores obtained for the seven vowel contrasts in Part B. Contrasts that were especially difficult for the subjects in each L1 group are boldfaced (see text).

<table>
<thead>
<tr>
<th></th>
<th>/ʌ/-/u/</th>
<th>/ɛ/-/ε/</th>
<th>/oʊ/-/oʊ/</th>
<th>/ɑ/-/ɒ/</th>
<th>/ɛ/-/æ/</th>
<th>/æ/-/æ/</th>
<th>/ɑ/-/ɑ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng.</td>
<td>.99 (.01)</td>
<td>.98 (.03)</td>
<td>1.0 (.00)</td>
<td>.99 (.01)</td>
<td>.97 (.04)</td>
<td>.99 (.02)</td>
<td></td>
</tr>
<tr>
<td>Cze.</td>
<td>.85 (.04)</td>
<td>.96 (.06)</td>
<td>.99 (.03)</td>
<td>.98 (.03)</td>
<td>.96 (.06)</td>
<td>.76 (.14)</td>
<td>.96 (.06)</td>
</tr>
<tr>
<td>Jap.</td>
<td>.49 (.23)</td>
<td>.87 (.07)</td>
<td>.97 (.04)</td>
<td>.95 (.04)</td>
<td>.89 (.05)</td>
<td>.73 (.15)</td>
<td>.93 (.05)</td>
</tr>
<tr>
<td>Hun.</td>
<td>.67 (.23)</td>
<td>.94 (.05)</td>
<td>.94 (.08)</td>
<td>.99 (.02)</td>
<td>.95 (.04)</td>
<td>.81 (.16)</td>
<td>.97 (.02)</td>
</tr>
<tr>
<td>Dut.</td>
<td>.78 (.19)</td>
<td>.90 (.13)</td>
<td>.98 (.03)</td>
<td>.97 (.04)</td>
<td>.96 (.04)</td>
<td>.90 (.06)</td>
<td>.94 (.07)</td>
</tr>
<tr>
<td>Por.</td>
<td>.73 (.24)</td>
<td>.79 (.22)</td>
<td>.97 (.02)</td>
<td>.92 (.07)</td>
<td>.74 (.35)</td>
<td>.59 (.30)</td>
<td>.85 (.18)</td>
</tr>
<tr>
<td>Ger.</td>
<td>.94 (.05)</td>
<td>.98 (.02)</td>
<td>.98 (.03)</td>
<td>.98 (.05)</td>
<td>.98 (.03)</td>
<td>.84 (.18)</td>
<td>.99 (.02)</td>
</tr>
<tr>
<td>Kor.</td>
<td>.55 (.24)</td>
<td>.88 (.10)</td>
<td>.99 (.02)</td>
<td>.98 (.05)</td>
<td>.93 (.08)</td>
<td>.79 (.07)</td>
<td>.98 (.02)</td>
</tr>
<tr>
<td>Ara.</td>
<td>.89 (.06)</td>
<td>.95 (.04)</td>
<td>.98 (.02)</td>
<td>.99 (.01)</td>
<td>.97 (.03)</td>
<td>.84 (.04)</td>
<td>.98 (.03)</td>
</tr>
<tr>
<td>M</td>
<td>.78 (.23)</td>
<td>.92 (.11)</td>
<td>.98 (.04)</td>
<td>.97 (.04)</td>
<td>.93 (.13)</td>
<td>.82 (.17)</td>
<td>.96 (.08)</td>
</tr>
</tbody>
</table>

One might infer that the native Korean participants did not generate two distinct codes when processing trials testing the /ʌ/-/u/, /ɛ/-/ε/ and /ɛ/-/æ/ contrasts. They may instead have identified the two vowels making up each contrast as being instances of a single Korean vowel. Given the structure of the discrimination test used here (see above), this would have prevented accurate discrimination. Research examining the perceptual assimilation of English vowels by native speakers of Korean will be needed, however, before such a conclusion can be reached with certainty.

Collins and Mees (1984) observed that native speakers of Dutch tend to judge instances of both the English /ɛ/ and /æ/ categories to be “allophones” of a single Dutch vowel, /ɛ/, and thus to “sound the same”. They also tend to judge instances of English /u/ and /u/ to be allophones of Dutch /u/. These observations lead to the expectation that the Dutch participants would generate a single phonetic code when hearing instances of English /ɛ/ and /æ/, and when hearing instances of /u/ and /u/. In fact, the English /u/-/u/ but not the /ɛ/-/æ/ contrast was classified as “difficult” for the native Dutch participants. However, additional analysis revealed that /ɛ/-/æ/ contrast was indeed relatively difficult for the Dutch participants. The Dutch participants obtained scores for 10 contrasts that significantly exceeded a chance level (p < 0.05 according to one-sample t-tests). However, neither their score for /ɛ/-/æ/ or /u/-/u/ significantly exceeded the chance level (t = 1.70 and 0.44, respectively; p > 0.10).
A method for assessing the perception of vowels in a second language

The need for perceptual assimilation data is illustrated by the results obtained for the native Japanese participants. These participants obtained a very low score for the /u/-/o/ contrast (M = 0.49) but a very high score for the /i/-/t/ contrast (M = 0.98). This might seem puzzling given that Japanese has a single high front vowel in the portion of the vowel space occupied by English /i/ and /u/ (viz. /i/), and a single high back vowel in the portion of the vowel space occupied by English /u/ and /o/ (viz. /u/). The difference in the scores obtained for these two English vowel contrasts can probably be attributed to a difference in how the two vowels in each contrast were perceptually assimilated by Japanese vowels. Japanese adults with little English-language experience tend to identify English /i/ and /u/ tokens in terms of a single Japanese vowel (/i/). However, they tend to identify English /i/ and /u/ tokens as being instances of two different Japanese vowel categories, /i/ and /e/ (Flege et al. 1998; see also Miranda & Strange 1989).

Stability of test results. It has been assumed that scores on the COT depended importantly on whether a non-native participant generated two distinct codes when processing the stimuli in the trials testing each contrast. In many cases, the difficulty of a particular English vowel contrast might be predicted based on a consideration of the acoustic-phonetic properties of the English vowels and the closest L1 vowel(s) (see, e.g., Yang 1996), and also empirical evidence concerning how the English vowels are related perceptually to vowels in the L1 (see, e.g., Strange et al. 2001). However, individual differences may arise as the result of perceptual learning. For example, two of the six native German participants obtained scores for /e/-/æ/ that were similar to the native English speakers' scores. However, one German participant obtained a score that was slightly below the chance level of 0.50. The results for the first two Germans agree with the finding (Bohn & Flege 1990) that a substantial amount of English-language experience may lead to an improved perception of English /e/ and /æ/ by native German speakers. The possibility exists that the participants showing high scores for /e/-/æ/ did so as the result of establishing a category for English /æ/ that was distinct from the category they had established as children for German /e/.

The foregoing conceptualization of the processes underlying responses to CDT items leads one to expect that the CDT will yield much the same scores from one administration to the next. That is, if discrimination improves only as the result of (a) establishing a new L2 vowel category, or (b) beginning to relate two distinct L2 vowels to two distinct L1 vowels, where earlier there had been a 2-to-1 mapping pattern, then perceptual changes should occur slowly over time. The purpose of the analysis presented in this section was to determine if these expectations hold true. As
mentioned earlier, training was administered on difficult English vowel contrasts two weeks after the first administration of the CDT. The test was administered a second time immediately following the training. The scores obtained on the first and second administrations of the CDT were compared. The CDT might be considered "stable" only if the scores changed very little or not at all across the two administrations, even for the contrasts that were trained.

Forty-four of the 48 nonnative participants were invited to return two weeks after the first CDT administration. (Four others who obtained high scores on all 14 contrasts were not invited to return.) These participants' scores for the 14 contrasts were rank ordered. Thirty-nine participants who obtained low scores on multiple contrasts were assigned to receive training on the contrasts receiving the lowest score and the third lowest score. Five other participants who obtained low scores for only two contrasts were assigned to receive training on the contrast receiving the lowest score. Twenty-five participants received training on the /ɛ/-/æ/ contrast, 21 on /ɑ/-/ʌ/, 19 on /ʊ/-/u/, 10 on /e/-/ɪ/, 4 on /æ/-/æ/, 3 on /ʊ/-/ʌ/ and 1 on /ɑ/-/æ/.

When the participants returned, they were shown a list of minimally paired English words containing the pair (or pairs) of vowels to be trained. After practice producing vowels in the minimal pairs contrastively, the participants were given identification training with feedback. The stimuli used in training were the same as those used in the CDT. For each contrast that was trained, five tokens of two English vowels were randomly presented 10 times each. The participants identified each token using two sets of keywords (the minimal pairs used earlier in production training). Feedback was provided after each identification response. The participants' correct response rate increased significantly, from 75-89%, over the ten blocks of training, \(F(9,261) = 7.04, p < 0.01\). Most of the improvement occurred during the first three random presentations of the stimuli.

The scores obtained for all 14 contrasts on the first and second administrations of the CDT were examined in (2) Time x (14) Vowel Contrast repeated-measures ANOVA. The Time factor was non-significant, \(F(1,43) = 0.52, p > 0.10\), because the scores obtained on the two administrations (first: 0.892; second: 0.897) were virtually identical. However, a significant two-way interaction was obtained, \(F(13, 559) = 2.95, p < 0.001\), suggesting that the scores for some of the 14 contrasts changed. Simple effects tests indicated that significantly higher scores were obtained on the second than first administration of the CDT for two contrasts: /ɪ/-/ɛ/ (first: 0.97, second: 0.99,) and /ɛ/-/æ/ (first: 0.74, second: 0.81). The change for /ɛ/-/æ/ might be attributed to training, for 25 participants were trained on this contrast. However, none of the participants were trained on /ɪ/-/ɛ/. 

\[\]
A method for assessing the perception of vowels in a second language

A second analysis assessed the effects of vowel training, if any, on discrimination. This analyses compared scores obtained on the two CDT administrations for contrasts that were trained and for equally difficult contrasts that were not trained. An average score was computed when two contrasts were trained. These average scores were compared to averages of two equally difficult contrasts that were not trained. (For participants trained on single contrast, the scores obtained for the trained contrast and an equally difficult untrained contrast on the first and second administrations of the CDT were compared.) Higher scores were obtained on the second than the first administration both for the trained vowel contrasts (second: 0.73, first: 0.67) and for the contrasts for which training was not provided (second: 0.80, first: 0.77). A (2) Time x (2) Type ANOVA examining these scores yielded a significant main effect of Time, \( F(1,37) = 5.52, p < 0.05 \), and a non-significant two-way interaction, \( F(1,37) = 0.95, p > 0.10 \).

In summary, familiarization with the CDT did not, in itself, seem to have increased test scores. Much the same scores were obtained on the two administrations of the CDT. A significant improvement was noted for just one contrast (/i/-/e/) which had already obtained a high score on the first administration. Feedback training on the CDT stimuli led to only a modest improvement in test scores, which were no greater than the improvement observed for equally difficult contrasts for which no training was provided. Taken together, these results indicate that the CDT yields stable test scores.

Summary
This study explored the discrimination of English vowels by speakers of eight languages who had learned English as an L2. Two things were necessary for the non-native participants to obtain high scores on the CDT for an English vowel contrast. The participants had to note differences between tokens of vowels drawn from different English vowel categories. They also had to ignore audible but phonetically irrelevant acoustic differences between tokens drawn from a single English vowel category. The results suggested that only those participants who generated two distinct codes for a pair of English vowels were able to obtain high discrimination scores. In some instances, this was likely due to the assimilation of two English vowels by two different L1 vowels. In other instances, the L2 learners may have established a new long-term memory representation for a vowel found in English but not the L1 (e.g. Flege 1995).

As expected, the nonnative participants’ perceptual sensitivity to the English vowel contrasts varied according to L1 background. It is likely that some of these differences were simply a reflection of how many vowels existed in the L1 inventory, and how often two English vowels would be
“mapped onto” two different L1 vowels. However, firm conclusions on this point cannot be reached based on the evidence presented here. Cross-language perceptual assimilation data were not obtained in this study. Moreover, there was a great deal of heterogeneity both across and within the L1 groups. Some between-group and intra-group variation may have reflected perceptual learning rather than different patterns of cross-language phonetic interference.

The most important finding of the study was that the CDT yielded stable scores across two test administrations. This finding is consistent with the view that high scores for a contrast depended on the generation of two distinct phonetic codes, and that such codes change slowly over time inasmuch as they are based on long-term memory representations that develop slowly. The findings indicate that the CDT would be appropriate for use in a longitudinal study of speech learning. A large increase over time in the scores obtained for an English vowel contrast might be taken as support for the claim (Flege 1995) that L2 learners are capable of establishing new phonetic categories for L2 vowels.

Acknowledgements

This research was supported by grant DC00257 from the National Institute for Deafness and Other Communicative Disorders. The author thanks Laurie Skelton and Upali Lockridge for their contributions.

References


A method for assessing the perception of vowels in a second language


A method for assessing the perception of vowels in a second language


