CHAPTER 8

Acoustic variability and perceptual learning

The case of non-native accented speech

Allard Jongman and Travis Wade

Jim Flege's research on category formation has introduced or emphasized several key concepts, including equivalence classification and the distinction between new and similar phones. The research described in this chapter addresses these concepts by investigating the role of acoustic variability in the formation of new categories as well as the extent to which this variability may hinder or help native and non-native listeners. A production study comparing Spanish-accented and native English vowels reveals a much greater degree of variability in nonnatives' use of the English vowel space. Results from a subsequent training study where vowel variability was systematically manipulated, suggests that for the most easily maintained distinctions, learning benefited from the high-variability training paradigm. In contrast, for very difficult distinctions, advantages were found for training only with minimal variability (prototypes). Finally, results are presented from a lexical decision task in which English and Dutch listeners responded to native and Dutch-accented English. While Americans prefer native English speech, the Dutch prefer the Dutch-accented stimuli. In addition, Dutch listeners are less efficient in processing words containing sounds that do not occur in Dutch even when listening to a native English speaker

Introduction

The field of phonetics of second language acquisition owes Jim Flege an enormous debt of gratitude. In fact, the existence of this field is largely due to his efforts over the past 25 years. His research continues to have a major impact on the areas of speech science and second language acquisition. Flege's research on category formation has introduced or emphasized several key concepts, including equivalence classification and the distinction between new and similar phones. The research described in this chapter addresses these concepts by investigating the role of acoustic variability in the formation of new categories as well as the extent to which this variability may hinder or help native and non-native listeners.

Acoustic variability and perceptual learning

We will start by summarizing our recent research on the role of acoustic variability in the perception of non-native accented speech (Wade, 2003; Jongman, Wade, & Sereno, 2003; Wade, Jongman, & Sereno, 2006). While greater acoustic variability has been shown to result in increased difficulty in identification (e.g., Mullennix, Pisoni, & Martin, 1989), the acquisition of non-native contrasts by adult second-language learners seems to benefit from exposure to greater variability during training. Namely, using a 'high variability training paradigm', Pisoni and colleagues have shown that non-native segmental contrasts are better acquired and longer retained when learners are exposed to these contrasts in different phonetic contexts and produced by a variety of speakers (e.g., Lively, Logan, & Pisoni, 1993; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). Recent research in our laboratory has successfully extended this high variability training paradigm to the acquisition of suprasegmental contrasts in Mandarin Chinese (e.g., Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2001, 2003; see also Sereno & Wang, this volume).

The underlying notion of the high variability training paradigm is that learning new sounds and adapting to native-produced variants requires exposure to sufficient variability. The question remains whether the same is true when dealing with non-native speech. It is commonly assumed that due to non-constant proficiency across speakers there is a much greater range of acoustic variability in non-native accented than native productions. It is therefore unclear if and to what extent listeners can be trained to recognize non-native speech sounds using the same methods that work for unfamiliar native sounds. To date, only a few studies have explored this issue, with mixed results. While listeners adapt to productions of individual accented speakers rapidly and robustly as a result of varied exposure, training effects generalizable across speakers of a particular accent are more elusive, and perhaps even limited to sentential stimuli (e.g., Wingstedt & Schulman, 1984; Clarke, 2000; Weil, 2001; Bradlow & Bent, 2003).

We therefore set out to train native speakers of English to comprehend Spanish-accented English using the high variability training paradigm. Six native speakers of Latin American Spanish (three females, three males) varying in exposure to English were recorded while producing lists of phonetically balanced English words (Egan, 1948). These lists consisted of 50 monosyllabic common words each.

In a testing phase, we used two of these speakers whose speech was of average recognition difficulty. One list of 50 words spoken by one of these speakers was used at pre-test (before training); a second list of 50 words by the same speaker ('old speaker') as well as a new list of 50 words by the other speaker ('new speaker') were used at post-test (after training). The remaining four speakers were used for training. On each of three consecutive days, trainees were exposed to 200 new words (50 by each of the four speakers). Training thus consisted of hearing 600 words produced by four different speakers.

Thirty participants took part in a listening experiment, including 15 trainees and 15 controls. Both trainees and controls were pre- and post-tested. Only the trainees

participated in the training phase. During training, trainees (college students with little or no exposure to Spanish or Spanish-accented English) listened to the accented stimuli and responded to each word by typing it on a computer keyboard. Maximal feedback was provided: the typed response, accompanied by bell (correct) or buzz (incorrect), appeared on the computer screen, as well as the intended word. After 1500 ms, the auditory stimulus was repeated while the visual feedback remained on the screen. The testing phase for both the trainees and control participants was similar to training, except that participants did not receive any feedback.

The results are shown in Figure 1. Overall, both the trainees and controls performed significantly better at post-test, presumably due to familiarity with the task. Additionally, the new speaker was more difficult for both groups of participants at post-test. However, there was no overall training effect or Speaker by Training interaction. Thus, the high variability training paradigm failed to produce an overall advantage in subjects' performance on new items produced by a previously encountered or new accented speaker. Indeed, the observed difference in improvement is opposite the expected direction for such an advantage: controls performed slightly better at post-test, and improved more over pre-test, than trainees.

However, the robust improvement across subject groups from pre- to post-test requires some explanation, as it seems unlikely that such a large difference in performance would occur entirely by chance across word lists balanced for phonetic difficulty. Moreover, trainees and controls did not demonstrate exactly the same pattern in post-test. The fact that words produced by the new speaker were more difficult for both subject groups to comprehend may be attributed to lower overall intelligibility of the new speaker, perhaps due to the speaker's lower proficiency or related factors. Additionally, though, there appears to be an interaction whereby trainees show a slight advantage for the new speaker while controls prefer the previously encountered one. While this effect was not significant, its relative persistence across subjects hints at the following possible scenario. The overall post-test improvement in accuracy for controls was likely due to a combination of (1) familiarization with the previously encountered speaker achieved during the pre-test even in the absence of a training-with-feedback mechanism, (2) similarly acquired familiarity with the task, and (3) any spurious effects of list difficulty. For trainees, factor (3) should have applied similarly, and may be to blame for part of the improvement for the old speaker across subject groups. However, the effects of factors (1) and (2) might have been attenuated slightly for trainee subjects due to forgetting the specific characteristics of the pre-test speaker's voice over the training process and its associated task. This process, then, might have been sufficient to obscure the effects of a sufficiently weak fourth factor for the trainees, namely the expected perceptual learning of accent-specific characteristics over training, which should have surfaced as an advantage for both new and previously encountered voices. This result, then, would represent a case of speaker-related training effects confusing or taking precedence over language population-specific training for a set of related sounds.

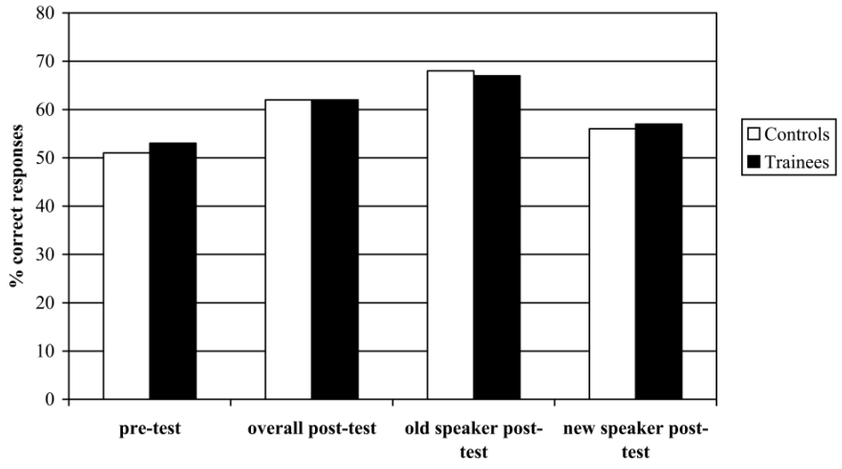

Figure 1. Overall accuracy in perception of accented speech for trainee and control participants at pre-test and post-test, including old speaker and new speaker post-test results

Additional evidence for speaker-familiarity effects comes from examination of trainee performance on the four speakers encountered in training over the three training sessions. Statistical analysis demonstrated that overall performance each day superseded that of the previous day. Thus, while training clearly gave subjects an advantage in attending to *speaker*-specific information, there was a conspicuous absence of learning for similar *accent*-level characteristics. This pattern, considering the previous cross-speaker success of very similar training methods in teaching (native produced) foreign language sounds (e.g., Wang et al., 1999) is consistent with the existence of processing difficulty unique to non-native productions, and can be extrapolated to also explain the ambiguous findings of previous studies on accent learning. We propose that for non-native speech, the most readily learnable patterns may exist at the level of the accented speaker, not the accented population, which itself cannot be adequately described as a homogenous set of learnable deviations from the standard pronunciation.

In the case of the present study, any potentially generalizable features of the Spanish accent itself must have been either (1) not sufficient in quantity to produce any advantage in trainee recognition of new accented words, or (2) so easily acquired that control subjects were able to perform similarly to trainees relying only on familiarity of pre-test items and on their limited past experience with the accent. Trainees and controls alike were probably drawing in part from their past encounters with the Spanish-English accent, limiting trainees' room for improvement to an extent that would not be expected for truly foreign or unfamiliar sounds. However, the presence of clear, sizeable speaker effects, and the fact that identification never approached 100% accu-

racy, demonstrates that there were in fact features to be learned, but that most of these were at the speaker level.

We have argued that the high variability training paradigm failed because the degree of acoustic variability in non-native accented speech is unusually high. However, we have not directly demonstrated that this is the case. In order to pursue this, we compared the acoustic characteristics of the tokens produced by the 6 Spanish speakers to those produced by a comparable group of 6 native speakers of English. We focused on the vowels since they seemed the source of a number of errors during training. Specifically, in order to compare the variability in the native and non-native use of the English vowel space, the 8 vowels [i, ɪ, ɛ, æ, ɑ, ʌ, ʊ, u] were analyzed. Vowel measurements were represented as points in a two-dimensional vowel space based on Miller's formulations of height and backness since this works well across speakers with widely differing F0 ranges (Miller, 1989). Within this framework, a speaker's 'sensory reference' (SR) related to his or her average F0 is calculated as follows: $SR = 168(GMf0/168)^{1/3}$, where GMf0 represents the geometric mean of the speaker's fundamental frequency. Height is then expressed as log (F1/SR), and backness as log (F2/F1).

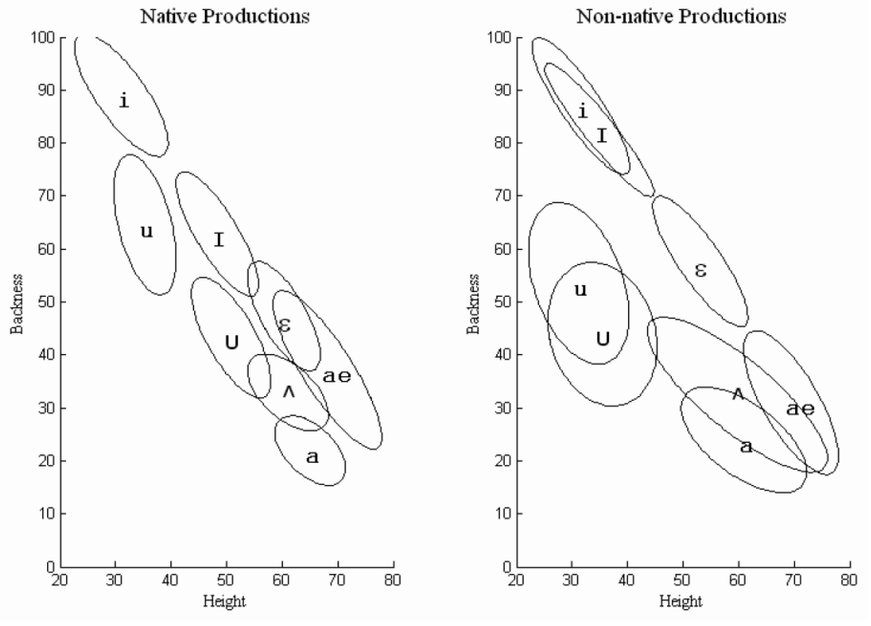

Figure 2. Vowel spaces for native and non-native speakers for eight monophtongal English vowels. Ellipses represent equal-likelihood contours

Figure 2 shows normalized equal likelihood contours for the native and non-native vowel productions to demonstrate the shape and relative location of each of the vowel

categories. It is clear that there are differences in the absolute location of many non-native categories, including those with similar Spanish sounds. This is presumably due to partial equivalence classification (e.g., Flege, 1987). In addition, the incompleteness of category formation related to distinctions that are not present in Spanish is evident. This is particularly clear for the English tense-lax pairs /i-ɪ/ and /u-ʊ/. While statistically distinct, the members of these pairs are both closer together and more variable in the non-native vowel space. In order to quantify these impressions, standard deviations calculated from the observed height and backness values for each vowel were compared. This comparison indicated that, relative to native productions, non-native productions had significantly greater standard deviations for both height and backness (see Wade, 2003 for detailed statistical comparisons). Non-native speakers consistently varied their productions more than native speakers, by a factor of about 1/3.

In sum, compared with a parallel set of words produced by native speakers, non-native productions demonstrated much more variability in the use of the perceptual vowel space. Non-native categories involved large, irregular distributions and in general a greater degree of category overlap than native correspondents. It is therefore clear that the non-native accented productions that proved difficult in the training study were more variable than the native ones would be. An obvious question is whether this unusually high degree of variability caused the high variability training paradigm to be ineffective.

In order to answer this question, we directly manipulated the degree of variability that participants were exposed to during training in a following study (Wade, 2003). Participants were trained to recognize hybrid synthetic/natural hVd words containing vowels with typical height and backness values based on the previous distributions and variability that was either artificially small, typical of native productions, or typical of non-native productions. They were then tested on distributions representative of the most variable (non-native) group. More specifically, mean F1 and F2 values and their correlation coefficients were based on observed values for either native or non-native speakers and were held completely constant within each group, maintaining the overall shape and central location of each vowel. Within each of these 2 categories, height and backness values of individual vowels used to train participants were taken randomly from normal distributions specified by the appropriate mean values, height-backness correlation and one of 3 SDs: an arbitrarily low value set to 0.01, a native SD or a non-native SD. Thus, there were 6 training conditions (3 types of variability combined with native or non-native means).

Seventy-two college students (12 per condition) participated in this experiment. During training, participants indicated which word they heard by clicking on one of eight response alternatives ('heed', 'hid', 'head', 'had', 'hod', 'hud', 'hood', 'who'd') displayed on the computer screen. Feedback was provided. Training continued until 30 of 50 consecutive responses were correct and each word had been correctly identified once. The test phase then immediately followed, consisting of 10 tokens of each possible word derived from a distribution with the appropriate vowel mean condition on which the participant was trained, with non-native variability.

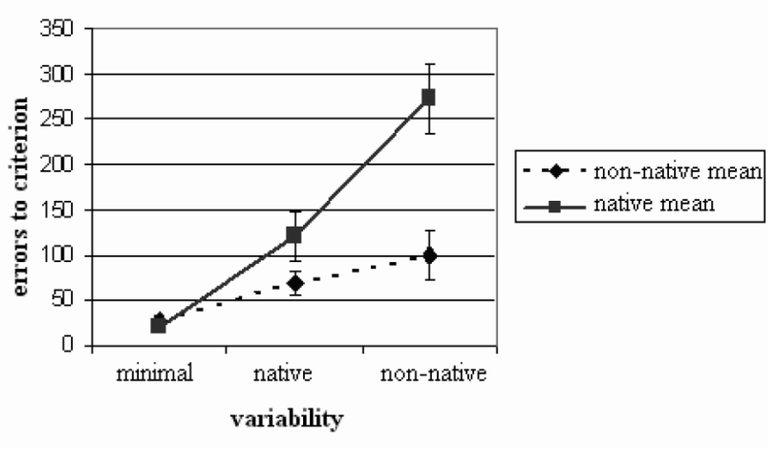

Figure 3. Errors to criterion in vowel ID by trainees exposed to native or non-native vowel means with three levels of acoustic variability. Criterion was 30 correct responses to any 50 consecutive tokens for each of the 8 words at least once

Table 1. Average distance between each vowel and all remaining vowels for the native and non-native mean conditions.

Vowel	Native mean	Non-native mean
i	0.51	0.45
ɪ	0.28	0.41
ɛ	0.23	0.29
æ	0.28	0.37
ɑ	0.36	0.37
ʌ	0.27	0.32
ʊ	0.24	0.30
u	0.33	0.31
Average distance	0.31	0.35

Figure 3 presents the number of errors during training as a measure of difficulty. It is clear that increasing variability leads to difficulty in learning. Interestingly, distributions based on non-native means were easier to learn, and the advantage of non-native means increased as variability increased. This pattern of results clearly counters the assumption that comprehension problems are merely due to deviation from the standard pronunciation and can therefore easily be overcome with experience. If that

were true, we would have expected an advantage for native distributions across variability conditions. Instead, the proximity among other pairs of vowels may have played a greater role. To evaluate this possibility, we calculated the average linear distance in terms of height and backness between each vowel's central position and that of each of the other vowels for the native and non-native conditions, using Euclidean distances. As shown in Table 1, the average distance between each vowel and all other vowels is greater in the non-native mean condition. While the most extremely positioned (tense high) vowels were more isolated in the native mean condition, the non-native condition involved a wider overall spread. As variability increased, this trend resulted in less total category overlap for non-native means. In turn, while native-mean participants enjoyed a slight, non-significant advantage in the minimal variability condition, their distributions grew comparatively more difficult as variability increased. These results suggest that perceptual learnability is determined by the *overall predictability* of a sound (its tendency not to occur in locations where it may be confused with other sounds) rather than its approximation of a typical native category.

Figure 4 provides insight into the pattern of learning during training. Average accuracy is shown for training sessions divided into thirds. Statistically, participants exposed to native and minimal variability improved throughout training. However, participants exposed to non-native variability showed no improvement.

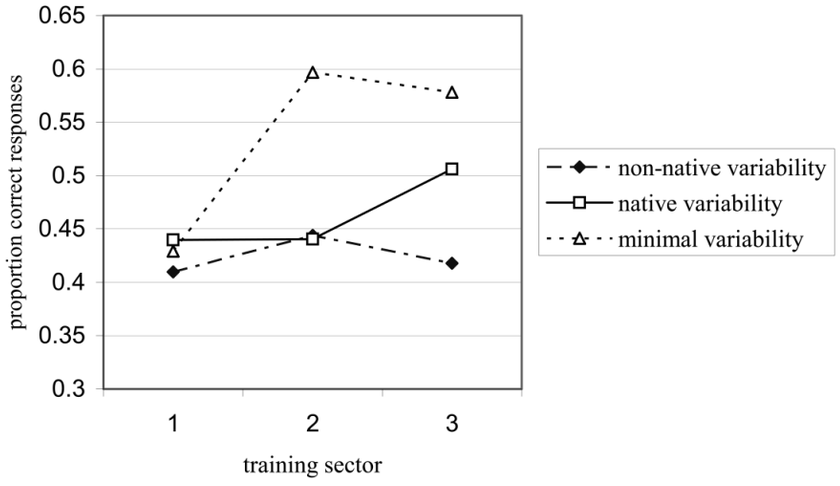

Figure 4. Accuracy in vowel identification over the course of training (first, second, and final third) across three levels of acoustic variability

The (post training) test results were very similar to those observed during training. Performance was generally better on nonnative-mean conditions, but no overall accuracy differences were observed based on degree of variability in training. In other

words, the nonnative-level variability involved in the test presents a difficulty that is not mitigated by greater exposure to variability in training. However, overall percent correct may not be the best measure of subject performance. We therefore looked at listeners' sensitivity to individual vowels and calculated and compared *d'* separately for each vowel. In general, within a language-mean condition, participants typically performed quite similarly across types of variability in training for most vowels. However, as shown in Figure 5, there was one vowel for which a clear effect of variability was observed, namely /i/. Posthoc tests revealed that the effect was in precisely the opposite direction across language-mean groups. In nonnative-mean tests, subjects trained with *minimal* variability outperformed native and non-native variability subjects. In contrast, in native-mean tests, subjects trained with *non-native* variability were most sensitive to /i/, outperforming native and non-native variability subjects.

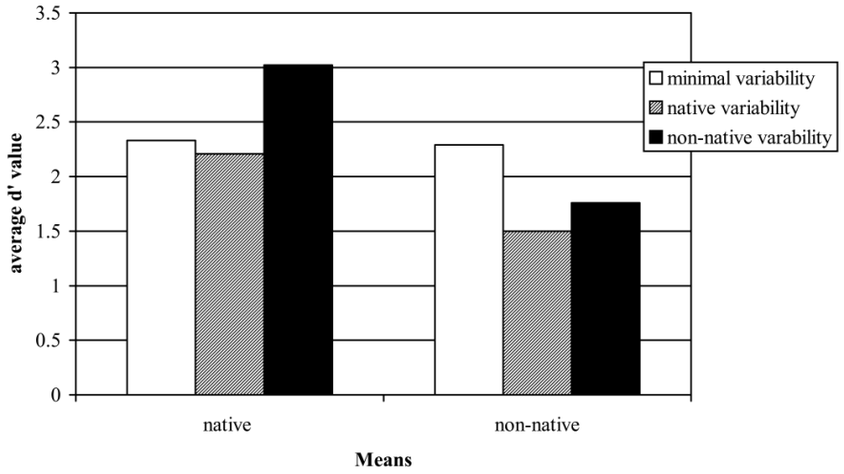

Figure 5. Post-test sensitivity (average *d'* values to the vowel [i] across training groups exposed to native or non-native vowel means

In other words, for the native-mean conditions, learning of /i/ followed the canonical high-variability training pattern: performance was best when training involved *maximal* variability. As shown in Table 1, /i/ is by far the most isolated vowel in native-mean distributions. Moreover, additional calculations revealed that its distance from its nearest neighbor, /u/, (0.25), is the largest distance separating any two neighboring vowels. As a result, performance on /i/ is better across native-mean subjects than for any other vowel. The vowel /i/ stands out so clearly that it is easily learned even when acoustic variability is extreme.

Non-native /i/ presents the opposite picture. For the non-native-mean conditions, performance improved only when subjects had been exposed to *minimal* variability

around the vowel's central location. /i/ is still the most isolated non-native vowel overall, but far less so than native /i/. In fact, its distance from its nearest neighbor, /ɪ/, (0.06), is the smallest. The non-native vowels /i/ and /ɪ/ are almost completely overlapping and the distinction was among the most difficult in the non-native mean condition. Thus, this represents a contrast that is so difficult to perceive that only when exposed to prototypical examples with minimal variability in training can listeners become aware of the relevant parameters of the distinction.

These results suggest that the high degree of acoustic variability found in non-native accented speech does indeed prevent improvement in the perception of accented speech when using the high variability training paradigm. For most individual vowels and overall accuracy, no high-variability effect was obtained and accent-level learning was minimal. For the most easily maintained distinctions, learning was still possible, and the high-variability effect was observable. Finally, for very difficult distinctions, advantages were found for training only on prototypes.

While it does not speak to the precise mechanisms responsible for creating and maintaining phonetic categories, this result does suggest a few things about the overall structure of the information used in recognition. In particular, it seems likely that more than one type of information may be learned during category learning. Taking into account the present results and those of previous studies employing a high variability training paradigm, we posit that listeners may become aware of at least (1) the structure and range of variability typical of a category, and (2) more abstract properties such as its central location. It seems likely that (2) might be learned most straightforwardly from prototypical category instances, while (1) would require exposure to multiple, varied exemplars. Figure 6 gives a qualitative demonstration of how these classes of knowledge might aid in the recognition of different types of categories and productions. In this sketch, we consider the relative categorization advantage provided by each type of information across category confusability, representing some combination of production variability and proximity to neighboring categories, and across token difficulty, presumably related to a production's idiosyncrasy and/or proximity to a category center. As shown in Figure 6, the advantage offered by awareness of variability is greatest for easily separable categories. Recognition becomes more difficult as category overlap increases and as individual tokens become more difficult, but the advantage is always positive, perhaps asymptoting at zero when there is no learnable information and categories overlap completely. Advantages resulting from more abstract, prototype information, it seems, should apply differently. These advantages will increase as tokens more closely resemble a prototype, but should not be affected by overall category difficulty. If these two effects—they may alternatively be thought of as recognition or learning strategies—are oriented as shown, circumstances where prototype training might surface are clear. Most native category productions would probably fall somewhere near the center of this space, where the advantages of variability awareness outweigh those of abstract property awareness. However, for the easiest tokens of the most difficult categories, a crossover effect such as that seen

in our examination of the nonnative-mean [i] is expected, whereby listeners with a better awareness of abstract category information actually show advantages. Probably, most of the non-native distributions we observed could be said to represent something intermediate to these possibilities, with distributions straddling the intersection of the two surfaces so that no net advantages were observable.

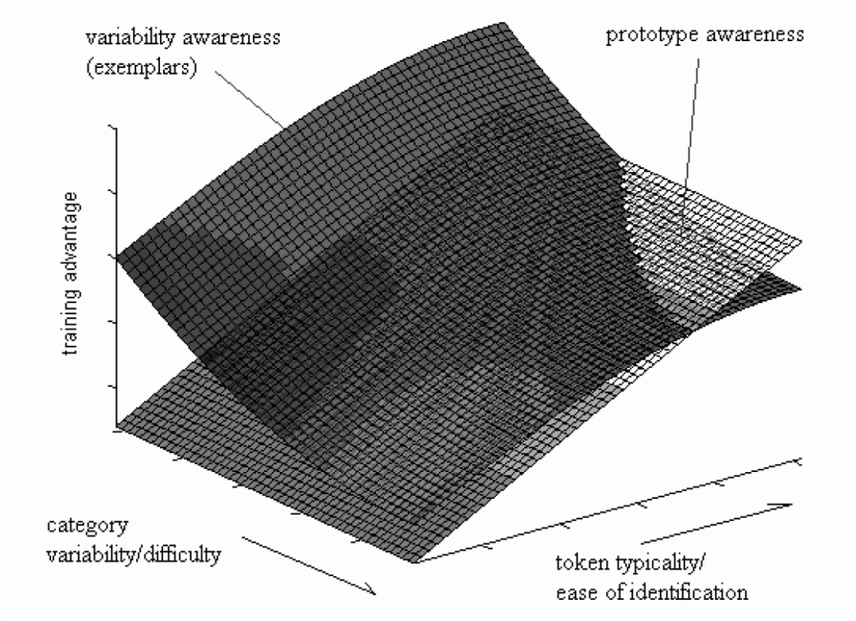

Figure 6. Possible interaction of prototype and variability awareness advantages in category identification

Perception of native and foreign-accented English by native and non-native listeners

The previous section summarized our recent research on the role of acoustic variability in the perception of non-native accented English by American listeners. Little is known, however, about the perception of non-native accented English by non-native speakers of English. This is an important issue since English has increasingly become the 'lingua franca' such that it is spoken by more non-native than native speakers. In other words, in countries where the native language is not English, it is possible that the English to which listeners are most often exposed is spoken by L2 speakers of English. It is not clear what effect this may have on the formation of perceptual representation and ultimately, on comprehension.

The most relevant study to date was conducted by Bent & Bradlow (2003). These researchers recorded native speakers of Chinese, Korean, and English as they spoke a set of simple English sentences. These sentences were then embedded in white noise (signal-to-noise ratio of 5 dB) and presented for comprehension to a set of either English, Chinese, or Korean listeners, as well as a mixed listener group from various native language backgrounds. Listeners listened to the sentences and wrote down what they heard. Accuracy was scored in terms of the number of keywords (approximately three per sentence) that were perfectly transcribed. Results showed that, for the native English listeners, the native English speaker was most intelligible. However, for non-native listeners, speech from a relatively high-proficiency non-native speaker sharing the same native language was as intelligible as a native speaker. This advantage due to a match in native language between a non-native speaker and non-native listener was termed the "matched interlanguage speech intelligibility benefit". Interestingly, however, a similar advantage was obtained when the non-native listener and non-native speaker did *not* share a native language.

The present study (see also Sereno, McCall, Jongman, Dijkstra, & Van Heuven, 2002) provides a more detailed look at the comprehension of non-native accented speech by non-native listeners. Specifically, we explore comprehension by means of an on-line lexical decision task. In addition, we evaluate Flege's hypothesis that during the acquisition of second language speech, phonetic categories interact through mechanisms of category assimilation and category dissimilation (e.g., Flege, 1995a). The dichotomy between assimilation and dissimilation depends, in part, on the perceived phonetic similarity of the first and second language sounds. Similar second-language phones are sounds judged to be realizations of one category in the first language. These similar phones are contrasted to new second language phones which do not have a counterpart in the first language.

Our research contrasts similar and new phones across English and Dutch using native and non-native speakers of English as well as native and non-native listeners. Given that the acoustic-phonetic realization of a word spoken with a non-native accent is altered, the question is how this affects comprehension by native as well as non-native listeners. Two auditory lexical decision experiments were conducted. A female native speaker of English and a female native speaker of Dutch judged to have a moderately strong accent produced a set of 160 English stimuli, half of which were words and half nonwords (the nonwords did not exist in either English or Dutch). Half of the words and nonwords contained a majority of phonemes that are new or unique to English and do not occur in Dutch (e.g., /θ/, /æ/). We will refer to these stimuli as the *unique phoneme* stimuli. The other half of the words and nonwords consisted only of phonemes that are similar in English and Dutch (e.g., /s/, /i/). These we will refer to as the *common phoneme* stimuli. All stimuli were monosyllabic and matched for a number of relevant variables, including word frequency, number of phonemes, and number of letters. These stimuli were then presented to both American and Dutch listeners. A group of 40 monolingual American college students were tested at the

University of Kansas in the U.S. A group of 40 Dutch college students were tested at the University of Nijmegen in The Netherlands. These Dutch listeners all had some proficiency in English.

All participants heard both the native and non-native speech stimuli in counterbalanced order using a randomized blocked design. They were to indicate whether a given stimulus was an English word or not by pressing one of two response buttons (W or NW). Reaction times and error rates were collected. For the reaction time data, stimulus duration was subtracted from the response latencies to control for differences in word length.

Overall results for both reaction times and errors are shown in Figure 7. Focusing first on the reaction time data, a significant Listener by Speaker interaction was obtained. Post-hoc tests revealed that, as expected, American listeners respond significantly faster to native English speech than to Dutch-accented speech. In contrast, the Dutch listeners showed the opposite pattern, with significantly faster reaction times to the Dutch-accented speech. A similar significant interaction was also found for the error data. American listeners made significantly more errors on the Dutch-accented speech than on the native speech while no difference is observed across native and non-native speech for the Dutch listeners. It seems that in a speeded lexical decision task, American listeners preferred the native speech while the Dutch listeners preferred their own Dutch-accented variety.

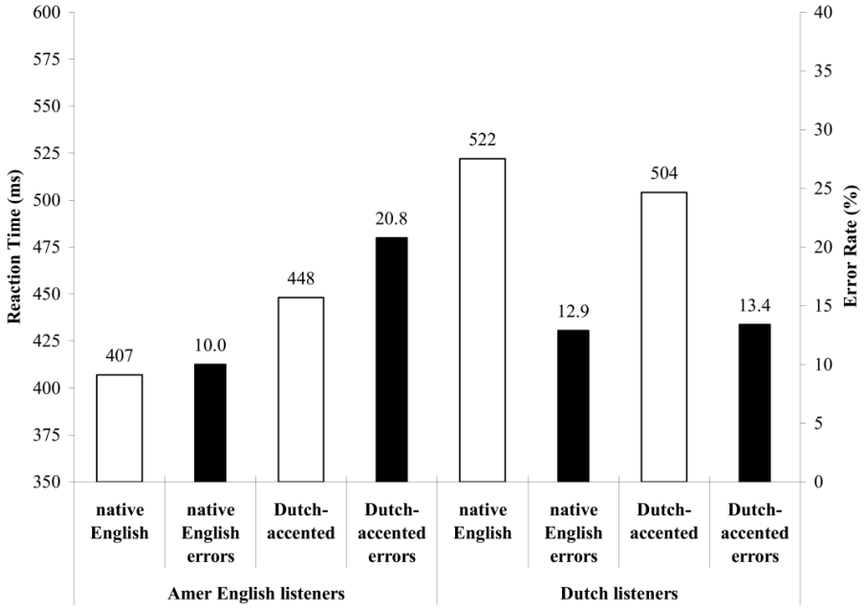

Figure 7. Lexical decision reaction times and error rates to native and Dutch-accented English by American and Dutch listeners

A second set of results involves the contrast between stimuli containing common versus unique phonemes. When listening to native English, American listeners did not make any distinction between common and unique phoneme stimuli, neither in terms of reaction time nor error rate. This should not come as a surprise since this common/unique distinction is of course meaningless to monolingual English speakers. However, when American listeners hear Dutch-accented speech, they make significantly more errors on unique phoneme stimuli (21%) than on common-phoneme stimuli (12%). The unique-phoneme accented stimuli, as heard by the native listeners, resulted in many more errors.

A different pattern obtains for the Dutch listeners listening to Dutch-accented speech: Unlike the Americans, they have comparable reaction times and error rates for unique phoneme stimuli (14%) and common phoneme stimuli (10%). The Dutch do not seem to be hindered by the accented unique stimuli. However, when the Dutch listeners listen to native English, a difference between common and unique phonemes shows up in terms of reaction time, as shown in Figure 8. The Dutch listeners have more trouble with the unique than with the common phoneme stimuli when produced by a native speaker: they respond more slowly to the unique phoneme stimuli. Presumably, this reflects the fact that the native English stimuli with unique phonemes mismatch the Dutch listeners' internal representations for those sounds.

In sum, we found that native speakers of English, as expected, have difficulty with Dutch-accented speech as compared to "unaccented" native English speech. Most intriguing, however, is that native speakers of Dutch prefer the Dutch-accented speech. These results are in agreement with those obtained by Bent & Bradlow (2003) in a sentence transcription task and suggest that second language learners process accented speech more efficiently, both in terms of latency and accuracy, than native speech, at least when non-native speaker and listener share the same native language. Second, the distinction between common and unique phonemes is relevant in terms of accuracy when listening to accented speech. This is true for both American and Dutch listeners. The American listeners make many more errors on the unique phoneme words when listening to non-native English. For the Dutch, this effect is much smaller. Overall, accent affects the unique phoneme stimuli more than the common phoneme stimuli. Finally, Dutch listeners differentiate common from unique phonemes even when listening to a native speaker of English. Word recognition accuracies decrease and response latencies increase when processing English stimuli with phonemes that have no counterpart in the first language. A Dutch listener is less efficient in processing unique phoneme words even when listening to a native English speaker. Even though the stimulus is not degraded, perception by Dutch listeners is more difficult for the unique phoneme words.

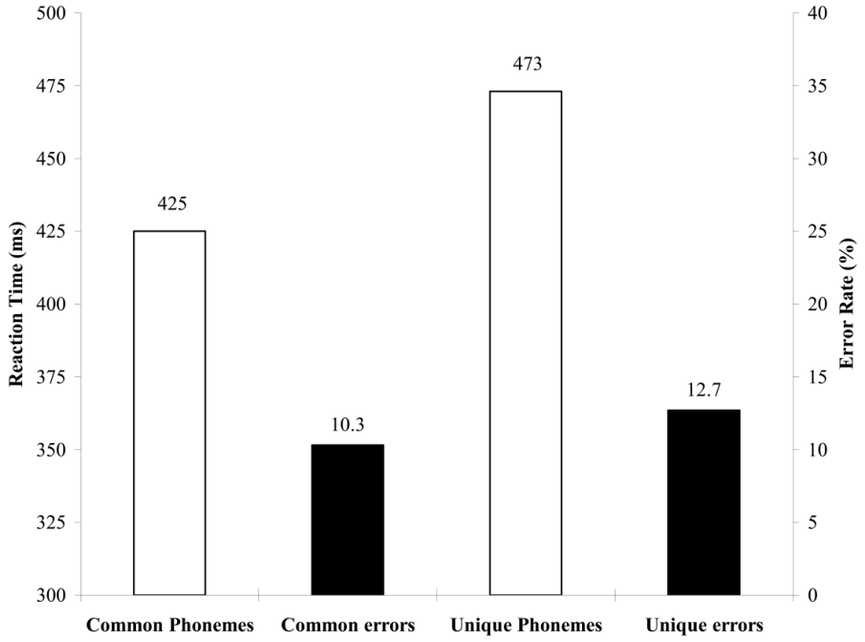

Figure 8. Lexical decision reaction times and error rates to native English speech by Dutch listeners

Conclusion

Our research has applied models developed to understand the acquisition of native-language contrasts to non-native accented productions. Taken together, the research discussed in this chapter has documented and quantified the high degree of acoustic variability that seems characteristic of non-native speech. Non-native productions generally show more vowel variability and overlapping adjacent vowel categories. Results from our training study in which degree of acoustic variability was systematically manipulated suggest that it is this great acoustic variability that prevents trainees from improving their perception of non-native accented speech. However, greater acoustic variability does not always result in impaired comprehension. Results from our lexical decision experiments indicated that while Americans prefer native English speech, the Dutch prefer the Dutch-accented stimuli. Although we did not conduct any acoustic comparisons of the native and Dutch-accented stimuli, our results for Spanish-accented English would predict that the Dutch-accented speech exhibits greater acoustic variability than the native English speech. This increased acoustic variability, however, did not harm comprehension by listeners sharing the same native

language as the speaker. This is presumably due to the fact that non-native speakers of a language share knowledge of both their native language as well as (aspects of) the second language. Because of a shared knowledge of the way the native language and the second language interact, the non-native listener has an advantage over the native listener in interpreting acoustic information that may deviate substantially from the native norm.

We thus return to one of the core themes in Jim Flege's work, namely the interaction between the sound systems of the native and second language in acquisition. A thorough understanding of the phonetic and phonological structure of both the learner's native language and the target language to be acquired is required as the basis for specific predictions about the kinds of training that may be most beneficial to learning certain contrasts as well as about the kinds of problems that non-native listeners may have in comprehension. Our recent research suggests that, for certain distinctions, training with prototypes may be more beneficial than training with high variability stimuli. In addition, future research should investigate whether our finding that Dutch listeners prefer Dutch-accented English is due to the fact that Dutch and English share certain phonetic and phonological characteristics or reflects a more general accommodation on the part of non-native listeners.