Individual Differences in Perception of Unfamiliar Speech

Melissa Baese-Berk[1], Tessa Bent[2], Stephanie Borrie[3], and Megan McKee[2]

Affiliations: [1]University of Oregon, [2]Indiana University, [3]Utah State University
E-mail Addresses: mbaesebe@uoregon.edu, tbent@indiana.edu, Stephanie.Borrie@usu.edu, meemckee@umail.iu.edu

ABSTRACT

Substantial individual differences in speech perception emerge under adverse listening conditions, including when perceiving unfamiliar speech. In the present study, individual differences in word recognition for three varieties of unfamiliar speech—Spanish-accented English, Irish-accented English, and dysarthric speech—were investigated. We asked whether the ability to perceive these unfamiliar, and highly variable, speech signals correlated with performance on three linguistic and non-linguistic tasks. Individual word recognition scores were significantly correlated across all three types of speech. Further, cognitive and perceptual flexibility as well as receptive vocabulary were significant predictors of word recognition accuracy. The findings presented here are an important step in understanding what makes a listener successful at speech perception in adverse conditions, suggesting that both linguistic and non-linguistic cognitive abilities support word recognition with speech types that deviate from typical native dialect norms.

Keywords: Individual differences, accented speech, disordered speech, speech perception

1. INTRODUCTION

In everyday communication, listeners' encounter highly variable speech signals. However, in general, listeners are quite good at recognizing the words they are exposed to. In fact, when normal-hearing listeners are exposed to familiar dialects in quiet listening conditions, most individuals perform at ceiling. However, when listeners are exposed to speech in adverse listening conditions, either as a result of environmental or talker-related degradation, substantial individual differences begin to emerge [1-3]. In the present study, we examine individual variability in word recognition by asking two key questions: (1) Are listeners who are particularly successful at recognizing one type of unfamiliar speech also successful at recognizing other types of unfamiliar speech? (2) Which linguistic and cognitive factors predict listener success in word recognition with unfamiliar speech?

Unfamiliar speech varieties—including regional dialects, non-native accents, and disordered speech—have been shown to negatively influence speech perception, including reductions in speech intelligibility [4-7]. Listeners can improve their understanding of unfamiliar varieties throughout repeated exposure [8-12]. However, substantial individual differences are observed both pre- and post-exposure to the unfamiliar speech [13-16].

Identifying the underlying cognitive and linguistic factors that give rise to these individual differences is important theoretically and for clinical applications. Cognitive factors appear to influence individual variability in speech perception. The ability to perceive speech in noise is correlated with some cognitive skills, including working memory [17]. Further, a number of cognitive and linguistic skills including cognitive flexibility, working memory, short-term memory, selective attention, and vocabulary knowledge have been shown to correlate with perceptual learning of unfamiliar accents [16, 18]. These studies suggest that speech perception and adaptation in adverse conditions may be related to higher-level linguistic and non-linguistic cognitive functions. However, previous studies have typically examined a single type of unfamiliar speech or adverse listening condition. Whether the same cognitive-linguistic functions support accurate perception of all unfamiliar speech types and adverse listening conditions is largely unknown.

In the present study, we use three unfamiliar speech varieties as a test case for recognizing speech in adverse conditions. Specifically, we examine two types of unfamiliar accented speech (Irish English and Spanish-accented English) and one type of disordered speech (dysarthric speech). These types of speech differ from native American English along multiple segmental and suprasegmental dimensions. For example, an unfamiliar dialect, spoken by a native speaker, differs from familiar dialects primarily in terms of vowel production [19]. On the other hand, non-native speech differs in vowel and consonant production, as well as lexical stress and intonation. This variation is primarily driven by the interaction between the first and second languages (L2) of the speaker [20, 21], as well as unique properties of the L2 that challenge learners regardless of their L1 [22]. Non-native speech also tends to display more inter- and intra-talker variability than native speech varieties [23]. Finally, dysarthria, a motor speech disorder arising from neurological disease or injury, is characterized by pathological speech patterns which include imprecise articulation of vowels and consonants, disturbed prosody, and abnormal vocal quality [24]. Dysarthric speech differs from native norms in both systematic and unsystematic ways [6, 12].

Here, we examine the relationship among listeners' abilities to accurately perceive three unfamiliar speech varieties (Irish English, Spanish-accented English, and dysarthric speech). We also explore whether word recognition for unfamiliar speech is predicted by measures of linguistic and cognitive performance.

2. METHOD

2.1. Participants

Data was collected from 22 native American English speaking adults between the ages of 18 and 35 years. Participants reported no history of speech disorders and passed a hearing screening. None of the participants reported significant exposure to the speech types used in the present study.

2.2. Stimuli

Seventy-six experimental phrases [24, 25] were used. The stimuli consisted of 6 syllable phrases (ranging in length from 3-5 words) that were syntactically plausible but semantically anomalous to control for the contribution of semantic and contextual knowledge to intelligibility. The experimental phrases were elicited from six speakers (three males, three females) representing three unfamiliar speech types—native accent (Irish English), non-native accent (Spanish-accented English), and one form of disordered speech (dysarthric speech).

2.3. PROCEDURE

Participants completed a word recognition task and several cognitive-linguistic assessments. For the word recognition task, participants were seated in front of a computer monitor and presented with 78 experimental phrases (13 novel, and completely randomized, phrases per speaker) via headphones. Following the presentation of each phrase, participants were asked to type what they heard. No feedback regarding task performance was given. In addition to the word recognition task, participants completed a demographic questionnaire, Intra-Extra-Dimensional Set Shift (IED) task from the Cambridge Neuropsychological Test Automated Battery (CANTAB) [26], the Flanker test of cognitive inhibition, and the Peabody Picture Vocabulary Test (PPVT), 4th Edition [27].

In brief, the IED task was designed to test cognitive flexibility, by examining rule acquisition and reversal using a series of lines and shapes. This test assessed both intra- and extra-dimensional shifts as measured by the participant's ability to generalize previously learned rules to novel exemplars (intra-dimensional) or by shifting to a novel rule (extra-dimensional). In contrast, the Flanker test of cognitive inhibition measured the participants' ability to inhibit conflicting information. A series of arrows (e.g., > > > > > or > > < > > >) were presented and participants were instructed to indicate the direction of the central arrow while ignoring the surrounding arrows via a keyboard press. In general, response time is slower for incongruent trials than congruent trials [28]. Finally, the PPVT is a test of receptive vocabulary in which participants are required to match pictures to verbally presented words.

2.4. Scoring and Analysis

The word recognition task was analysed for proportion of key words that were correctly recognized (henceforth, proportion correct). Words correct were defined as those that matched the intended target exactly, as well as homophones and/or common misspellings. However, words with added or deleted morphemes were scored as incorrect (e.g., *supplied* for *supplies*).

The three cognitive-linguistic assessments were scored using standard scoring protocols. For the analyses, we used the log-normalized conflict cost measure generated by the Flanker test [16]. For the IED test, we used the z-scores from the normed number of total errors on the intra-dimensional shifts and the extra-dimensional shifts. For the PPVT, we used the raw scores.

3. RESULTS

Substantial variability across listeners was observed for each unfamiliar speech type. The mean scores, standard deviation, minimum and maximum scores for each speech type are reported below in Table 1.

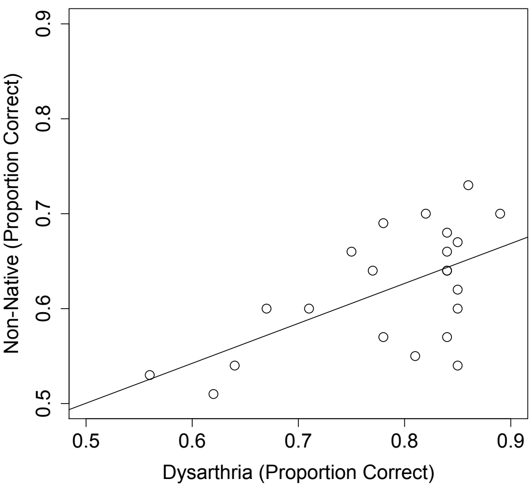
Table 1: Summary statistics for each speech type.

Comparison	Non-Native	Dialect	Dysarthria
Mean	.62	.78	.76
Std. Dev.	0.06	0.09	0.05
Minimum	0.51	0.56	0.57
Maximum	0.73	0.89	0.82

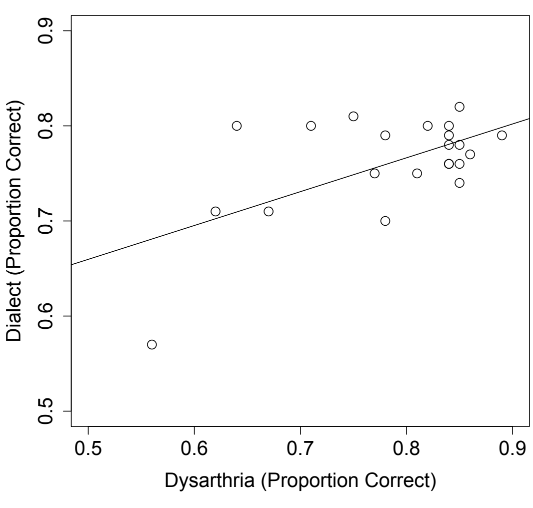
We first correlated word recognition accuracy for each of the three speech types, to examine whether perception abilities for each unfamiliar speech type are correlated with one another. When examining average performance across all items and both speakers, there are significant correlations for all three comparisons (Dysarthria and Dialect: r^2=.592, p=.003; Dysarthria and Non-Native: r^2=.597, p=.003;. Non-Native and Dialect: r^2=.23, p=.01). Thus, listeners who demonstrate successful word recognition with one type of unfamiliar speech are also adept at recognizing words produced by speakers with other unfamiliar speech types.

Figure 1a: Word recognition accuracy for the non-native (y-axis) and dysarthric (x-axis) speech.
Figure 1b: Word recognition accuracy for the unfamiliar dialect (y-axis) and dysarthric (x-axis) speech.
Figure 1c: Word recognition accuracy for the unfamiliar dialect (y-axis) and non-native (x-axis) speech.

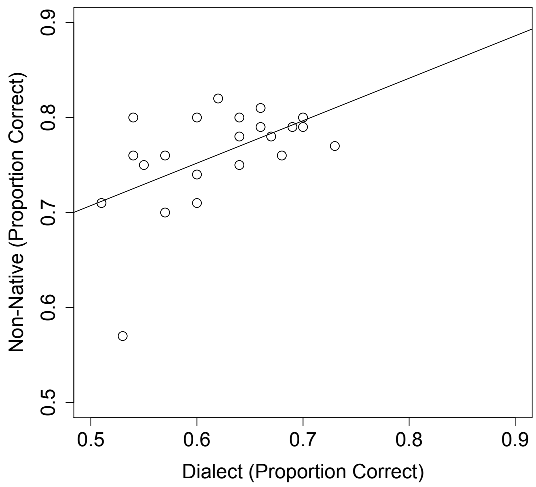
a



b



c



Next, the word recognition data were analysed using logistic mixed effects models to examine what cognitive-linguistic factors predict speech perception scores. Fixed factors entered into the regression were speech type (i.e., non-native, unfamiliar dialect, vs. dysarthria), gender (i.e., male vs. female), raw score on the PPVT, the Flanker conflict cost score, and the two z-scores from the IED test: Intra-Dimensional Shift (IDS) score and Extra-Dimensional Shift (EDS) score. Random effect structure was the maximal structure supported by the data. Subjects were not included as random items because variance in the listeners is accounted with several other measures (i.e., PPVT score and ID-ED scores). Significance was determined using model comparisons to determine whether each fixed factor contributed significantly to model fit.

The model with the best fit included fixed factors of Speech type (i.e., Dialect, Dysarthria, or Non-Native), Gender of the speaker (i.e., male or female), PPVT score, ID-ED scores, and a random effect of items. The Flanker scores were not significant predictors of performance on the word recognition task. The best fitting model is summarized below in Table 2, and the model comparisons appear in Table 3.

Table 2: Model parameters for the best-fitting linear mixed-effects model, as determined by model comparison.

Fixed Effects	Estimate	Std Err	t-value
Intercept	0.598	0.030	19.827
Speech Type (Dysarthria)	0.082	0.034	2.394
Speech Type (Non-native)	-0.063	0.034	-1.863
Gender	-0.006	0.034	0.183
PPVT	0.001	0.002	7.149
Intra-Dimensional Shift	0.155	0.060	2.579
Extra-Dimensional Shift	-0.069	0.033	-2.094

Table 3: Results of model comparisons. Predictor is the factor not included in the model.

Predictor	χ^2	p-value
Speech Type	48.3	<.001
Gender	21.2	<.001
Receptive Vocabulary	50.3	<.001
Intra-Dimensional Shift	6.65	.009
Extra-Dimensional Shift	4.39	.036

4. CONCLUSIONS AND DISCUSSION

Listeners varied widely in their ability to perceive each speech variety, replicating previous findings for the perception of non-native and disordered speech [13-16]. However, the present findings extend existing research by demonstrating that, although substantial individual differences in the perception of unfamiliar speech exist, individual performance is correlated across types of unfamiliar speech. Further, this relationship occurs even though the systematicity of the variability in each unfamiliar speech type differs. Listeners who succeed at perception of Irish English, the speech type that is most systematic in its deviation from American English norms also succeed at perceiving speech that is quite unsystematic in the ways in which it deviates from typical native norms (i.e., dysarthric speech). Further, the deviations, particularly in dysarthric speech, occur along multiple acoustic-phonetic dimensions, including both segmental and suprasegmental features.

The current results also indicate that several cognitive-linguistic factors correlate with speech perception across unfamiliar speech types. Receptive vocabulary and cognitive flexibility emerged as significant predictors of word recognition accuracy. Receptive vocabulary has been tied to better speech-in-noise perception, possibly due to greater lexical connectivity facilitating the use of top-down knowledge [2]. This study suggests that the vocabulary size effect extends to perception of unfamiliar speech varieties in quiet listening conditions. Finally, cognitive flexibility also emerged as a predictor of word recognition performance. Speech perception, especially in the face of highly variable input is a task that requires listeners to be flexible with their speech perception strategy. Listeners are required to map a noisy, continuous signal onto a set of discrete representations. Because the signal is so variable, listeners must be adaptable in terms of what they are willing to map to each of these representations. The current findings suggest that general, non-linguistic cognitive and perceptual flexibility support listeners' ability to map novel acoustic-phonetic words variants onto items in the lexicon. One open question is whether these findings, based on listener adversity due to speaker characteristics, would also extend to listener adversity arising from environment-related degradation (e.g., noise). Indeed, the Flanker task, a measure of inhibitory control and selective attention, was not a significant predictor of speech perception performance. It is possible that such measures would be more critical in noise-added conditions, in which listeners are required to segregate and supress perception of competing acoustic information while attending to the target speech.

Taken together, these results add further support to the growing body of literature that acknowledges individual variability, as well as linguistic and non-linguistic contributions to speech perception. It also demonstrates that performance on one type of unfamiliar speech correlates with performance on another. These findings provide a step in understanding what makes someone a good listener, by exploring the cognitive and linguistic skills that support speech perception in adverse listening conditions. Future work should examine how the constellation of cognitive-linguistic skills supporting speech perception varies depending on the type of adverse condition, including both talker-based and environment-based factors.

5. REFERENCES

[1] Gilbert, J., Tamati, T., Pisoni. D., 2013. Development, Reliability and Validity of PRESTO: A New High-Variability Sentence Recognition Test. *Journal of the American Academy of Audiology,* 24, 26-36.

[2] Mattys, S., Davis, M., Bradlow, A., Scott, S. 2012. Speech recognition in adverse conditions: A review. *Language and Cognitive Processes,* 27, 953–978.

[3] Tamati, T., Gilbert, J., Pisoni, D. 2013. Some Factors Underlying Individual Differences in Speech Recognition on PRESTO: A First Report. *Journal of the Acoustical Society of America,* 24, 616-634

[4] Adank, P., Evans, B., Stuart-Smith, J., Scott, S. 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance,* 35. 520–529.

[5] Bent, T., Bradlow, A. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America,* 114: 1600–1610.

[6] Borrie, S, McAuliffe, M., Liss, J. 2012. Perceptual Learning of Dysarthric Speech: A Review of Experimental Studies. *Journal of Speech, Language, and Hearing Research,* 55. 290–305.

[7] Floccia, C., Butler, J., Girard, F., Goslin, J. 2009. Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development,* 33, 366–375.

[8] Baese-Berk, M., Bradlow, A., Wright, B. 2013. Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America,* 133. EL174–EL180.

[9] Bradlow, A., Bent, T. 2008. Perceptual adaptation to non-native speech. *Cognition*, 106, 707-729.

[10] Borrie, S., McAuliffe, M., Liss, J., Kirk, C., O'Beirne, G., & Anderson, T. 2012. Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language & Cognitive Processes, 27*, 1039-1055.

[11] Clarke, C., Garrett, M. 2004. Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America* 116: 3647-3658.

[12] Sidaras, S., Alexander, J., Nygaard, L. 2009. Perceptual learning of systematic variation in Spanish accented speech 125: 3306–3316.

[13] Atagi, E., Bent, T. 2015. Relationship between listeners' nonnative speech recognition and categorization abilities. *Journal of the Acoustical Society of America* 137: EL44–EL50.

[14] Borrie, S. In Press. Visual information: A help or hindrance to perceptual processing of spastic dysarthric speech. *Journal of the Acoustical Society of America*.

[15] Derwing, T., Munro, M. 1997. Accent, intelligibility and comprehensibility: Evidence from four L1s 20, *Studies in Second Language Acquisition*, 19, 1–16.

[16] Janse, E., Adank, P. 2012. Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology* 65: 1563–1585.

[17] Akeroyd, Michael A. 2009. Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology,* 47, S53–S71.

[18] Adank, P., Janse, E. 2010. Comprehension of a novel accent by young and older listeners. *Psychology and Aging,* 25. 736–740.

[19] Clopper, C., Pisoni, D. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America* 118. 1661–1676.

[20] Best, C., McRoberts, G., Goodell, E. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system, *Journal of the Acoustical Society of America,* 109, 775–794.

[21] Weber, A., Cutler, A. 2006. First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America,* 119, 597-607.

[22] Baker, R., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K., Bradlow, A. 2011. Word durations in non-native English. *Journal of Phonetics*, 39, 1–17.

[23] Hanulíková, A., Weber, A. 2011. Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception & Psychophysics* 74. 613–629.

[24] Duffy, J. 2012. *Motor speech disorders: Substrates, differential diagnosis, and management* (3rd ed.). St. Louis, MS: Elsevier Mosby.

[25] Liss, J., Spitzer, S., Caviness, J., Adler, C., Edwards, B. 1998. Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *Journal of the Acoustical Society of America* 104. 2457–2466.

[26] Cutler, A., Butterfield, S. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31: 218–236.

[27] Sahakian, B., Morris, R., Evenden, J., Heald, A., Levy, R., Philpot, M., Robbins, T. 1988. A comparative study of visuospatial memory and learning in Alzheimer-type dementia and Parkinson's disease. *Brain* 111. Oxford University Press: 695–718.

[28] Dunn, L., Dunn, D. 2007. *PPVT-4 : Peabody picture vocabulary test*. Minneapolis, MN: Pearson Assessments.

[29] Eriksen, B., Eriksen, C.. 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics* 16. 143–149.