Research Note

The Effect of Intensified Language Exposure on Accommodating Talker Variability

Mark Antoniou,^a Patrick C. M. Wong,^a and Suiping Wang^b

Purpose: This study systematically examined the role of intensified exposure to a second language on accommodating talker variability.

Method: English native listeners ($n = 37$) were compared with Mandarin listeners who had either lived in the United States for an extended period of time ($n = 33$) or had lived only in China ($n = 44$). Listeners responded to target words in an English word-monitoring task in which sequences of words were randomized. Half of the sequences were spoken by a single talker and the other half by multiple talkers.

Results: Mandarin listeners living in China were slower and less accurate than both English listeners and Mandarin

listeners living in the United States. Mandarin listeners living in the United States were less accurate than English natives only in the more cognitively demanding mixed-talkers condition.

Conclusions: Mixed-talkers speech affects processing in native and nonnative listeners alike, although the decrement is larger in nonnatives and further exaggerated in less proficient listeners. Language immersion improves listeners' ability to resolve talker variability, and this suggests that immersion may automatize nonnative processing, freeing cognitive resources that may play a crucial role in speech perception. These results lend support to the active control model of speech perception.

Human talkers differ markedly in the way they produce the same word, and although listeners are sensitive to these differences in talker identity, they nevertheless understand speech quite effectively and achieve phonetic constancy—stable recognition of the phonetic structure of utterances (Bradlow, Torretta, & Pisoni, 1996; Shankweiler, Strange, & Verbrugge, 1977). The process by which listeners accommodate differences across talkers has been referred to as *talker normalization* and, according to some authors, crucially depends on the efficiency of spoken language processing (Nusbaum & Magnuson, 1997; Nusbaum & Morin, 1992). We thus might expect listeners to differ in their ability to accommodate talker variability as a function of their language proficiency. In this article, we posit a relationship between language exposure and the efficiency of spoken language processing by investigating the talker normalization abilities of language learners who systematically differ from young adult native speakers in terms of the amount of time spent in the target language environment.

Numerous studies have demonstrated that there is a processing cost when there is talker variability in a set of utterances, which decreases the efficiency of speech perception. When listeners are presented with multiple talkers compared with only a single talker, recognition of vowels, consonants, syllables, and words is slower (Lively, Logan, & Pisoni, 1993; Nusbaum & Morin, 1992; Summerfield & Haggard, 1973); speech perception in noise is less accurate and slower (Mullennix, Pisoni, & Martin, 1989); word shadowing is delayed to a higher degree (Goldinger, 1998); serial recall of spoken word lists is poorer (Martin, Mullennix, Pisoni, & Summers, 1989); and recognition of spoken sentences is poorer when embedded in multitalker babble (Van Engen & Bradlow, 2007). Although there is considerable evidence that listeners are sensitive to talker variability, there are few theoretical accounts of why this occurs. A phonetic-acoustic account might be that mixed-talkers speech is simply acoustically more variable than that of a single talker. However, acoustic differences result in increased recognition times only when they signal a change in talker, whereas recognition times are unaffected when an acoustic difference of the same magnitude does not signal a different talker (Magnuson & Nusbaum, 2007). Exemplar models have argued that the costs incurred by processing multiple-talkers speech are evidence that indexical information is not merely treated as noise and discarded during speech perception, but rather affects episodic memory for spoken words (Bradlow, Nygaard, & Pisoni, 1999; Goldinger, 1998;

^aThe Chinese University of Hong Kong

^bSouth China Normal University, Shipai, Guangzhou, People's Republic of China

Correspondence to Suiping Wang: suiping@scnu.edu.cn

Editor: Jody Kreiman

Associate Editor: Ewa Jacewicz

Received September 15, 2014

Revision received December 21, 2014

Accepted March 19, 2015

DOI: 10.1044/2015_JSLHR-S-14-0259

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Pisoni, 1997). The active control model (Magnuson & Nusbaum, 2007) goes beyond the exemplar account by explaining that the deleterious effects of talker variability on task performance are due to the resource-demanding remapping between the incoming signal and known phonetic categories. According to the active control model, the flow of computation during perception is dependent on the outcome of certain comparisons that occur in a closed feedback loop that monitors and modifies the phonetic categories that subserve human speech communication in a context-sensitive way. A change in talker triggers normalization procedures that operate until a stable mapping between the talker and internal phonetic categories is achieved, and this is maintained until a subsequent talker change occurs (Ladefoged & Broadbent, 1957; Nearey, 1989). Note that the listener does not need to be aware of these mechanisms. If the active control model is correct, then accommodating talker variability would depend crucially on the efficiency of spoken language processing and the cognitive resources that underlie it (e.g., working memory).

If the active control model's assertion is correct—that talker variability requires listeners to distribute attention over more cues in the signal than when there is a single talker (Nusbaum & Schwab, 1986)—then its greatest effects should be observed in those individuals with fewer available resources. One group that allows for testing this hypothesis is nonnative listeners. A large body of research has demonstrated that whereas the processing of the native language is automatic and seemingly effortless, information processing is less efficient in a nonnative language (Ardila, 2003; Bradlow & Pisoni, 1999; Chincotta, Hyönä, & Underwood, 1997; Cutler, Garcia Lecumberri, & Cooke, 2008; Mayo, Florentine, & Buus, 1997; Service, Simola, Metsänheimo, & Maury, 2002; Takayanagi, Dirks, & Moshfegh, 2002). For example, perceiving nonnative speech in noise results in a greater performance decrement than for native listeners (Cutler et al., 2008; Garcia Lecumberri & Cooke, 2006). Completing linguistic tasks also places a greater cognitive load on nonnative than native listeners, realized as longer reaction times for word recognition, both in noise and quiet (Cooke, Garcia Lecumberri, Scharenborg, & van Dommelen, 2010). In a recent study, van Dommelen and Hazan (2012) found that proficient nonnative listeners showed poorer mixed-talker speech comprehension than native English listeners, and this was exacerbated under a cognitive load condition (recognition of isolated words vs. triplets). These findings suggest that processing speech and accommodating talker variability in a nonnative language is more cognitively demanding than in a native language.

Given that nonnative listeners have less language experience than native speakers, it follows that they should also have fewer available cognitive resources, and we hypothesized that nonnative listeners will exhibit exaggerated talker variability effects relative to natives (consistent with van Dommelen & Hazan, 2012). Indeed, nonnative listeners typically perform more poorly on word recognition tests than native listeners (Bent & Bradlow, 2003; Bradlow & Pisoni, 1999; Takayanagi et al., 2002). However, with practice and

experience, performance on cognitively demanding tasks becomes automatic, the amount of required resources decreases, and the task becomes less demanding (Navon & Gopher, 1979). Consistent with this view, Japanese listeners with high, medium, or low experience with English differed in their learning of English consonant contrasts (Guion, Flege, Akahane-Yamada, & Pruitt, 2000), and thus we might expect a similar effect to be observed in resolving talker variability. Therefore, the present study we tested the hypothesis that language experience contributes to the efficiency of processing multitalker speech. Specifically, we expected to see a performance decrement when comparing nonnative listeners who have not lived in their second language environment with those who have undergone sustained language exposure. Individuals who have undergone intensified and extended exposure to a nonnative language are expected to have more cognitive resources available for speech processing (e.g., greater working memory availability). To test this, we compared word recognition in blocked-against mixed-talker presentations between populations that differ in their language proficiency, namely native against nonnative speakers. Specifically, we compared native English speakers to Mandarin learners of American English who had either lived in the United States for a prolonged period of time or had lived only in China, in order to determine if the intensity of nonnative language exposure affects the efficiency of spoken language processing in a word-monitoring task.

Methods

Participants

We compared groups of English monolinguals with two groups of nonnative learners of English from China. Thirty-seven native English monolinguals ($M_{\text{age}} = 19.9$ years) were students at Northwestern University and will henceforth be referred to as the *native English group*. These participants had received some training in a second language, usually in high school, but none reported proficiency in a second language or its continued use on a daily basis. Two groups of Mandarin listeners were recruited, both of whom had begun learning English during primary or middle school. The first group comprised 33 native Mandarin participants living in the United States and studying at Northwestern University ($M_{\text{age}} = 20.4$ years), and will be referred to as the *nonnative U.S. group* (M age of learning = 7.1 years; $SD = 3.2$). These participants had been living in the United States for a prolonged period of time (M length of residence = 6.5 years, range = 2–11 years), during which they had practiced and developed their English skills so as to meet the academic and social demands that come with studying in an American university, including acting as teaching assistants. We originally had 34 nonnative U.S. participants, but one was removed due to an excessive number of false alarms. The second Mandarin group consisted of 44 Chinese participants living in China and studying at South China Normal University ($M_{\text{age}} = 19.0$ years). These participants

had never lived in an English-speaking country (defined as any period greater than 3 months) and will be referred to as the *nonnative Chinese group*. The nonnative Chinese participants acquired English in a classroom-based setting (M age of learning = 8.6 years, $SD = 1.6$). Two of the 44 participants used English on a daily basis, 10 used English when listening to music and media, and the remaining 32 participants primarily used English when at school. None of the participants reported any audiologic or neurologic deficits. All passed a pure-tone screening at 25 dB HL at 500, 1000, 2000, and 4000 Hz. The groups were matched for age and sex. Demographics of the three groups are presented in Table 1.

Stimulus Materials

Twenty-four words were stimuli for the word-spotting task in which listeners monitored a random sequence of words and pressed a button when a target word was heard. There were eight target words: *apple, bear, bin, cat, corn, duck, fish, and horse*. There were also 16 filler words: *chalk, clock, lettuce, mushrooms, notebook, onion, orange, parrot, pencil, rabbit, scissors, squirrel, stapler, strawberries, tape, and watermelon*. All were concrete nouns, and this list has been used in past research on Spanish–English listeners (Sommers & Barcroft, 2011). The stimuli were digitally recorded to a computer (16-bit, 44.1 kHz) from eight native English talkers (four men, four women) in a sound-attenuated booth, using a Shure SM-10A headset cardioid microphone. The stimuli were excised and acoustically analyzed in Praat. Final tokens were selected to have comparable durations across talkers and were root-mean-square normalized to an output level of about 72 dB SPL using a sound level meter.

Procedure

Participants were instructed that they would be completing a word-spotting task in which they would hear a random sequence of words. The procedure followed closely that employed by Nusbaum and Morin (1992). At the beginning of each sequence, presented on the screen was a printed target word, which participants were required to remember. They could take time to memorize the word, and once they were ready, they would press a button to begin. The target word would then disappear and a random sequence of words would be heard, lasting for approximately

Table 1. Demographic information of native English and Chinese participant groups demonstrating that the groups were matched for age and sex.

Group	n	Age		Sex			
		M	SD	Women	%	Men	%
Native English	37	19.9	3.3	24	64.9	13	35.1
Nonnative U.S.	33	20.4	2.8	22	66.7	11	33.3
Nonnative Chinese	44	19.0	0.7	29	65.9	15	34.1

30 s. Participants were asked to push a single response button whenever that particular target word was heard and to respond as quickly as possible. At the end of the sequence, the screen flashed, a new target word was displayed onscreen, and participants repeated the procedure.

Sixteen sequences were presented in total, half of which were of a single talker; the other half were spoken by mixed talkers (eight target words $\times$ two conditions [blocked vs. mixed talkers]). The order of sequences was randomized, with the restriction that blocked and mixed sequences alternated during the course of the experiment. Each sequence comprised 27 words (four presentations of the target word + 23 fillers), and the order of the words within each sequence was also randomized, with the restriction that target words could not occur in the first or last position, and they were always separated by at least one filler. Blocked-talker sequences contained four presentations of the target word and 23 fillers spoken by a single talker, whereas mixed-talker sequences contained two presentations of the target word produced by one male talker and one female talker (four target presentations in total), and distractors from the other six talkers. The interstimulus interval between words in a sequence was 250 ms. Given that there was no practice, the first sequence was omitted from analyses to ensure that the data reflected participants' performance once they were familiar with the experimental procedure.¹ We computed sensitivity scores (d') for each participant, using the formula $z(\text{hit rate}) - z(\text{false alarm rate})$, as well as participants' reaction times for correctly spotted words. Hits were defined as registered button presses following the presentation of the target word, and false alarms were button presses following the presentation of a filler.

Results

The English and Chinese groups' word-monitoring sensitivity scores (d') are illustrated in Figure 1. Upon first inspection, it appears that listeners in the nonnative Chinese group missed more targets than the native English and nonnative U.S. groups. A 3×2 analysis of variance was conducted on the word-monitoring sensitivity scores (d'), with the between-subject factor of group (native English vs. nonnative U.S. vs. nonnative Chinese) and the within-subject factor of talker variability (blocked vs. mixed talkers). No outliers were detected, and assumptions of normality, homogeneity of variance, and sphericity were satisfactory. There was a significant main effect of group, $F(2, 111) = 31.76$, $p < .001$, $\eta_p^2 = .364$. There was also a significant main effect of condition, $F(1, 111) = 16.11$, $p < .001$, $\eta_p^2 = .127$, driven by the fact that overall, signal detection (d') was poorer in the mixed condition ($M = 3.45$) than in the blocked ($M = 3.23$).

It is important to note that there was a significant Group $\times$ Condition interaction, $F(2, 111) = 5.73$, $p = .004$,

¹Note that we also conducted all statistical analyses with all data points included and the pattern of results was the same as those reported.

Figure 1. Sensitivity to detect target words in word monitoring when each trial consists of speech from a single talker (blocked) or eight different talkers (mixed) in native English speakers or Chinese nonnative learners of English living in the United States (NN US) or in China (NN China). Error bars depict standard error of the mean.

$\eta_p^2 = .094$. A Bonferroni-adjusted alpha level of .0083 was employed for post hoc analyses (.05/6 = .0083). Post hoc *t* tests revealed that listeners in the nonnative Chinese group were less accurate than the native English listeners in both the blocked-talker, $t(79) = 4.23, p < .001$, and mixed-talker conditions, $t(79) = 8.26, p < .001$. The nonnative Chinese participants were also less accurate than the nonnative U.S. group in the blocked-talker, $t(75) = 2.83, p = .006$, and mixed-talker conditions, $t(75) = 4.14, p < .001$. The native English group outperformed the nonnative U.S. group in the mixed-talker condition, $t(68) = 3.20, p = .002$, but the two groups did not differ in the blocked-talker condition, $p = .273$.

A separate 3×2 mixed factorial ANOVA was conducted on the participants' recognition times in word monitoring, which are shown in Figure 2. There was a significant main effect of group, $F(2, 111) = 38.10, p < .001, \eta_p^2 = .407$. Post hoc Sidak tests revealed that the nonnative Chinese group was slower to recognize words than both the native English $p < .001$, and nonnative U.S. groups, $p < .001$, whereas the English and nonnative U.S. groups did not differ, $p = .166$. There was also a significant main effect of condition, $F(1, 111) = 128.50, p < .001, \eta_p^2 = .537$, indicating that,

Figure 2. Target recognition time in blocked- against mixed-talker word monitoring. NN US = Chinese nonnative learners of English living in the United States; NN China = Chinese nonnative learners of English living in China. Error bars depict standard error of the mean.

overall, responses to target words in the mixed-talker condition were slower than in the blocked-talker condition (blocked response time = 451 ms vs. mixed response time = 484 ms). There was no significant Group $\times$ Condition interaction, $p = .859$.

Discussion

The findings demonstrate that intensified and extended exposure to a second language affects performance on a word-monitoring task involving blocked- against mixed-talker presentations, compatible with past research (Bradlow & Pisoni, 1999; Takayanagi et al., 2002; van Dommelen & Hazan, 2012). The results also replicate past findings that talker variability increases cognitive load (e.g., Nusbaum & Morin, 1992) and lend support to the active control model of speech perception, which specifies that cognitive processing increases to resolve talker variability (Magnuson & Nusbaum, 2007).

Nonnative learners of English were more affected by talker variability in terms of signal detection accuracy and recognition time in a word-monitoring task. The native English listeners were consistently more accurate and faster to identify target words than nonnative Chinese listeners. Native English listeners were also more accurate than the nonnative U.S. group in the cognitively demanding mixed-talker condition. Differences were also observed between the nonnative U.S. and nonnative Chinese groups based on second language proficiency (as indexed by having lived in an English-speaking country for a prolonged period of time). In particular, Chinese learners of English who had not been immersed in an English-speaking country (nonnative Chinese group) were less accurate and slower to identify words than Chinese learners of English who had lived in the United States for a prolonged period (nonnative U.S. group). The findings suggest that intensified and extended exposure to a nonnative language within an immersion context improves the efficiency of nonnative speech processing and may reduce the cognitive effort required to attend to spoken word targets both when perceiving the speech of a single as well as multiple talkers.

The results complement past findings suggesting that the effects of talker variability are exaggerated in individuals with low second-language proficiency (van Dommelen & Hazan, 2012). This lends support to the active control model of speech perception (Magnuson & Nusbaum, 2007), according to which talker variability increases cognitive demands, thus decreasing response accuracy and prolonging reaction time. This view is consistent with work demonstrating that memory span is reduced when testing occurs in a nonnative language (Brown, 1992). Here, we demonstrated that accommodating talker variability becomes more efficient as a result of intensified exposure to a second language, and we speculate that the underlying factor responsible is the reduced availability of cognitive resources (e.g., working memory) in the Chinese groups. Spending time in a second language immersion environment is known to improve proficiency (Cummins, 1980) and speech production (Flege,

Yeni-Komshian, & Liu, 1999; Piske, MacKay, & Flege, 2001), and our findings suggest that it may also reduce the cognitive effort required to process second language speech and free working memory resources to complete other tasks. The active control model specifies that the greatest decrements should be observed in those listeners who do not possess sufficient working memory resources in reserve. Our results are consistent with the active control account.

The data challenge the claims of nonanalytic exemplar models that claim that sensitivity to the indexical properties of speech relies on past instances of spoken words rather than on cognitive resource availability (Goldinger, 1998; Pisoni, 1997). According to such models, listeners preserve holistic exemplars of speech events, and statistical clustering eliminates the need for an explicit mechanism to accommodate talker variability. In order to account for the present findings, the crucial point is that a mechanism that monitors the speech signal and modifies the attentional resources devoted to speech perception as required is needed. Such a mechanism is provided by the active control model but not by nonanalytic exemplar models. However, the data may be interpreted as being compatible with analytic exemplar-based theories that incorporate notions of attentional modulation (Goldinger & Azuma, 2003; Johnson, 1997).

Future work may investigate the relationship between talker variability and second language proficiency observed here. If our assertion that second language proficiency affects the amount of available cognitive resources, it should be the case that the listeners in the nonnative China group would show less of a performance decrement in a word-monitoring task conducted in their native language, Mandarin. Performance would also be expected to improve following sustained language immersion in a second language environment. It may also be informative to examine the relationship between nonnative language exposure and resolving talker variability in a population who vary continuously in length of residence, to complement our between-groups approach here. These possibilities await to be tested in future investigations.

Acknowledgments

This research was supported by U.S. National Institutes of Health Grant R01DC013315 (awarded to P. C. M. W. and B. C.), National Science Foundation of China Grant 31271086 (awarded to S. W. and P. C. M. W.), Australian Research Council Discovery Early Career Research Award DE150101053 (awarded to M. A.), and the Global Parent Child Resource Centre Limited.

References

- Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, 36, 233–240. doi:10.1016/S0021-9924(03)00022-4
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114, 1600–1610. doi:10.1121/1.1603234
- Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, 61, 206–219. doi:10.3758/BF03206883
- Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, 106, 2074–2085. doi:10.1121/1.427952
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20, 255–272. doi:10.1016/S0167-6393(96)00063-5
- Brown, G. D. A. (1992). Cognitive psychology and second language processing: The role of short-term memory. In R. J. Harris (Ed.), *Advances in psychology: Cognitive processing in bilinguals* (Vol. 83, pp. 105–121). Amsterdam, the Netherlands: North-Holland.
- Chincotta, D., Hyönä, J., & Underwood, G. (1997). Eye fixations, speech rate and bilingual digit span: Numeral reading indexes fluency not word length. *Acta Psychologica*, 97, 253–275. doi:10.1016/S0001-6918(97)00031-0
- Cooke, M., Garcia Lecumberri, M. L., Scharenborg, O., & van Dommelen, W. A. (2010). Language-independent processing in speech perception: Identification of English intervocalic consonants by speakers of eight European languages. *Speech Communication*, 52, 954–967. doi:10.1016/j.specom.2010.04.004
- Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 14, 175–187. doi:10.2307/3586312
- Cutler, A., Garcia Lecumberri, M. L., & Cooke, M. (2008). Consonant identification in noise by native and non-native listeners: Effects of local context. *The Journal of the Acoustical Society of America*, 124, 1264–1268. doi:10.1121/1.2946707
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104. doi:10.1006/jmla.1999.2638
- Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, 119, 2445–2454. doi:10.1121/1.2180210
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251–279. doi:10.1037/0033-295X.105.2.251
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The Quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 305–320. doi:10.1016/S0095-4470(03)00030-5
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, 107, 2711–2724. doi:10.1121/1.428657
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29, 98–104. doi:10.1121/1.1908694
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94, 1242–1255. doi:10.1121/1.408177
- Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human*

- Perception and Performance*, 33, 391–409. doi:10.1037/0096-1523.33.2.391
- Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V.** (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 676–684. doi:10.1037/0278-7393.15.4.676
- Mayo, L. H., Florentine, M., & Buus, S.** (1997). Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research*, 40, 686–693. doi:10.1044/jslhr.4003.686
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S.** (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, 85, 365–378. doi:10.1121/1.397688
- Navon, D., & Gopher, D.** (1979). On the economy of the human-processing system. *Psychological Review*, 86, 214–255. doi:10.1037/0033-295X.86.3.214
- Nearey, T. M.** (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, 85, 2088–2113. doi:10.1121/1.397861
- Nusbaum, H. C., & Magnuson, J. S.** (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 109–132). San Diego, CA: Academic Press.
- Nusbaum, H. C., & Morin, T. M.** (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 113–134). Tokyo, Japan: IOS Press.
- Nusbaum, H. C., & Schwab, E. C.** (1986). The role of attention and active processing in speech perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines* (Vol. 1, pp. 113–157). New York, NY: Academic Press.
- Piske, T., MacKay, I. R. A., & Flege, J. E.** (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. doi:10.1006/jpho.2001.0134
- Pisoni, D. B.** (1997). Some thoughts on “normalization” in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.
- Service, E., Simola, M., Metsänheimo, O., & Maury, S.** (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, 14, 383–408. doi:10.1080/09541440143000140
- Shankweiler, D. P., Strange, W., & Verbrugge, R. R.** (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.
- Sommers, M. S., & Barcroft, J.** (2011). Indexical information, encoding difficulty, and second language vocabulary learning. *Applied Psycholinguistics*, 32, 417–434. doi:10.1017/S0142716410000469
- Summerfield, Q., & Haggard, M. P.** (1973). Vocal tract normalization as demonstrated by reaction times. In *Report of speech research in progress*, 2 (pp. 12–23). Belfast, Northern Ireland: The Queen’s University of Belfast.
- Takayanagi, S., Dirks, D. D., & Moshfegh, A.** (2002). Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research*, 45, 585–597. doi:10.1044/1092-4388(2002/047)
- van Dommelen, W. A., & Hazan, V.** (2012). Impact of talker variability on word recognition in non-native listeners. *The Journal of the Acoustical Society of America*, 132, 1690–1699. doi:10.1121/1.4739447
- Van Engen, K. J., & Bradlow, A. R.** (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121, 519–526. doi:10.1121/1.2400666