Reading ability influences native and non-native voice recognition, even for unimpaired readers

Minal A. Kadam, Adriel John Orena, Rachel M. Theodore, and Linda Polka

Citation: The Journal of the Acoustical Society of America **139**, EL6 (2016); doi: 10.1121/1.4937488
View online: https://doi.org/10.1121/1.4937488
View Table of Contents: http://asa.scitation.org/toc/jas/139/1
Published by the Acoustical Society of America

Reading ability influences native and non-native voice recognition, even for unimpaired readers

Minal A. Kadam,[1] Adriel John Orena,[2] Rachel M. Theodore,[1,a)] and Linda Polka[2]

[1]*Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Storrs, Connecticut 06269, USA*

[2]*School of Communication Sciences and Disorders, McGill University, 2001 McGill College, 8th Floor, Montreal, Quebec H3A 1G1, Canada*

minal.kadam@uconn.edu, adriel.orena@mail.mcgill.ca, rachel.theodore@uconn.edu, linda.polka@mcgill.ca

Abstract: Research suggests that phonological ability exerts a gradient influence on talker identification, including evidence that adults and children with reading disability show impaired talker recognition for native and non-native languages. The present study examined whether this relationship is also observed among unimpaired readers. Learning rate and generalization of learning in a talker identification task were examined in average and advanced readers who were tested in both native and non-native language conditions. The results indicate that even among unimpaired readers, phonological competence as captured by reading ability exerts a gradient influence on perceptual learning for talkers' voices.

[AC]

Date Received: July 7, 2015 **Date Accepted:** November 28, 2015

1. Introduction

The acoustic speech signal provides information related to both talker identity and linguistic content. Historically, the mechanisms involved in talker learning and speech perception have been examined in separate literatures, reflecting the long-standing view that different aspects of the speech signal were used to cue talker identity and linguistic meaning. For example, indexical properties (e.g., fundamental frequency) were considered to cue talker identity with a separate set of phonetic properties (e.g., voice-onset time) serving to cue linguistic meaning. This view was challenged by findings indicating that listeners can extract both talker and linguistic properties even when these cues are absent from the speech signal (e.g., Remez *et al.*, 1997; Remez *et al.*, 1981) and by findings indicating that speech perception is influenced by experience with talkers' voices (e.g., Nygaard *et al.*, 1994; Theodore and Miller, 2010; Theodore *et al.*, 2015). Further, linguistic ability and experience influence talker learning such that listeners perform better on talker identification tasks in their native compared to a non-native language (e.g., Goggin *et al.*, 1991). Accordingly, a complete model of spoken language processing must specify how listeners integrate talker identity and linguistic content when analyzing the speech stream.

The literature suggests that multiple aspects of linguistic ability mediate talker identification. Specifically, the very presence of a native language benefit for talker recognition suggests that language comprehension facilitates talker learning. However, a native language benefit for talker discrimination has been shown with time-reversed speech (Fleming *et al.*, 2014), a signal that precludes comprehension, pointing to a role for sub-lexical language structure on talker identification tasks. Consistent with this account, infants (Johnson *et al.*, 2011) and adults (Orena *et al.*, 2015) show a native-language benefit for talker discrimination and identification prior to language comprehension, suggesting that emerging phonetic and phonological knowledge may promote talker learning even in the absence of language comprehension. Indeed, recent research indicates that phonological knowledge (i.e., knowledge of the sound structure of language) is one critical factor in linking language processing and talker learning abilities (e.g., Perrachione *et al.*, 2011; Perea *et al.*, 2014). Support for this hypothesis comes from Bregman and Creel (2014) who examined the effects of language ability as measured by age of acquisition on non-native talker identification. Korean-English

[a)]Also at: Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, 337 Mansfield Rd., Storrs, CT 06269, USA.

bilinguals who learned English early in life showed faster learning rates for English talkers compared with Korean-English bilinguals who learned English relatively later.

This relationship has also been examined by considering talker identification performance in individuals with reading disability (i.e., developmental dyslexia). It is widely recognized that reading ability is linked to performance on speech perception and phonological tasks for both individuals with dyslexia (Bradley and Bryant, 1983; Shankweiler *et al.*, 1995) and unimpaired readers (Burnham, 2003). Perrachione *et al.* (2011) examined talker identification for native and non-native languages in adult listeners with and without dyslexia. Their results showed that while typical readers demonstrated the expected native-language benefit, adults with dyslexia did not. For participants with dyslexia, talker identification was equally poor in both languages. Moreover, the degree of reading impairment for the dyslexic participants was negatively correlated with talker identification performance. Perea *et al.* (2014) extended these findings to include examination of both children and adults with developmental dyslexia. Consistent with Perrachione *et al.* (2011), Perea *et al.* found that talker identification in individuals with dyslexia was poorer compared to unimpaired control participants and that there was a gradient influence of reading ability on talker identification. However, unlike Perrachione *et al.* (2011), Perea *et al.* found that both children and adults with dyslexia exhibited a native language benefit for talker identification, similar in magnitude to the effect observed in control participants (albeit with overall lower accuracies). The difference between the results of these two studies may lie in the methods used to form the experimental groups, or it may be due to interactions between phonological ability and variability in the stimulus materials implemented across studies.

Relevant to the current work, these studies have demonstrated that talker learning is modulated by language ability and, moreover, that stability of phonological processing is one important factor in modulating this relationship. However, it is not known whether phonological processing, as captured by reading ability, influences talker identification not only in the impaired population but also across the range of values that comprise unimpaired readers. Further, it is not clear whether the effects of reading ability on talker identification reflect persistent deficits in talker learning abilities or, alternatively, if they reflect more short-term differences in the associative learning tasks used to examine talker identification. Previous examinations of the effect of reading ability on talker identification have provided limited exposure or training to the talkers' voices prior to test (e.g., Perrachione *et al.*, 2011). It may be that the effect of reading ability on talker identification is attenuated when listeners are given extended opportunity to learn the talkers' voices.

To this end, here we examine talker identification in two groups of unimpaired readers: average readers and advanced readers who perform near the middle of and top of the normal distribution on standardized reading assessments, respectively. We used a modified version of the paradigm reported in Bregman and Creel (2014) to examine readers' ability to learn native and non-native talkers and to generalize that learning to novel utterances.

2. Methods

2.1 Participants

Thirty-six English-monolingual speakers between the ages of 18 and 24 years ($M = 20$, $SD = 2$) were recruited. Reponses to questionnaires developed in our laboratories confirmed that participants had no history of speech, language, or hearing disorders, no knowledge of or systematic exposure to French, and no previous history of dyslexia or reading impairment. To confirm that differences in reading ability (described in the following text) were not attributable to differences in nonverbal intelligence, all participants completed the Test of Nonverbal Intelligence–fourth edition (TONI). Two participants were excluded because they scored below the tenth percentile on this measure, leaving 34 participants who scored within normal limits for inclusion in the study (see Table 1). Mean TONI percentile was 40 ($SD = 21$) for the average readers and 47 ($SD = 22$) for the advanced readers, which were statistically equivalent ($t_{32} = -0.946$, $p < 0.351$, $d = -0.325$). All participants passed a pure tone hearing screen on the day of testing, administered at 20 dB for octave frequencies between 500 and 4000 Hz.

Participants were assigned to either the average or advanced reading group based on performance for a standardized assessment battery of reading sub-skills and reading comprehension. The battery included measures of phonological awareness [Comprehensive Test of Phonological Processing–second edition (CTOPP)], rapid digit and letter naming [Rapid Automatized Naming (RAN), Rapid Alternating Stimulus

Table 1. Mean percentile, standard deviation (in parentheses), and test statistics of the average and advanced readers for each measure of the standardized assessment battery and the composite reading score.

Assessment	Average readers	Advanced readers	t	p	d
CTOPP—— Elision	51 (22)	66 (10)	−2.780	0.009	−0.878
CTOPP——Blending	59 (29)	75 (14)	−1.927	0.063	−0.703
CTOPP——Nonword repetition	65 (23)	72 (17)	−0.971	0.339	−0.346
RAN——Rapid digit naming	74 (7)	83 (6)	−3.942	<0.001	−1.381
RAN——Rapid letter naming	67 (9)	81 (8)	−4.714	<0.001	−1.644
RAS——2 set	72 (10)	86 (7)	−4.694	<0.001	−1.622
TOWRE——Sight words	61(22)	86 (13)	−4.004	<0.001	−1.384
TOWRE——Decoding	62 (15)	84 (12)	−4.767	<0.001	−1.620
WRMT——Word identification	65 (24)	86 (14)	−3.121	0.004	−1.069
WRMT——Word attack	56 (23)	75 (21)	−2.596	0.014	−0.863
WRMT——Comprehension	65 (22)	84 (12)	−3.170	0.003	−1.072
Composite reading score	63 (10)	80 (5)	−5.946	<0.001	−2.150

(RAS)], timed and untimed word and nonword reading [Test of Word Reading Efficiency–second edition (TOWRE), Woodcock Reading Mastery Test–third edition (WRMT)], and reading comprehension (WRMT). A composite reading score was calculated for each participant defined as the mean percentile score across the diagnostic measures and a median split of the composite reading score determined the average and advanced reading groups. Mean percentile for the average reading group was 63 ($SD = 10$) and mean percentile for the advanced reading group was 80 ($SD = 5$), which represent statistically distinct distributions ($t_{32} = -5.946$, $p < 0.001$, $d = -2.102$). As shown in Table 1, the grouping based on composite reading score patterned with performance between the two groups on each component of the assessment battery.

2.2 Stimuli

Auditory stimuli consisted of 12 English and 12 French sentences that were matched in number of syllables and are described in detail by Valji (2004) and Orena *et al.* (2015). Each sentence was produced by four native speakers of their respective languages. Acoustic analyses confirmed that talkers of the two languages were equally discriminable on the basis of sentence duration, fundamental frequency, and variation in fundamental frequency (Orena *et al.*, 2015). Previous testing of the stimuli confirmed that the English and French sentences were equally learnable for native speakers of each language (Orena *et al.*, 2015). Two sentences of each language were used during familiarization, five sentences were used during both the training and the test phases, and the remaining five sentences were presented only during the test phase. Visual stimuli consisted of eight cartoon faces, one for each talker.

2.3 Procedure

All participants completed familiarization, training, and test phases in both English and French. The three phases were blocked by language, which was counter-balanced across participants. All were tested individually in a sound-attenuated booth. Auditory stimuli were presented via headphones at a uniform listening level. Stimuli presentation and data collection were controlled with the SUPERLAB software on a Mac OS X system. Visual stimuli were presented via computer monitor and participants made responses using a button box.

During familiarization, participants heard two sentences produced by each of the talkers. Each sentence was paired with the appropriate cartoon face. Participants were asked to attend to each sentence and face to learn the voices. No responses were collected. The training phase consisted of blocks of 60 randomized sentences [5 sentences × 4 talkers × 3 repetitions]. On each trial, listeners heard one sentence and were asked to identify the talker by choosing from one of two cartoon faces using a button box. Feedback was provided during training in the form of "CORRECT" or "INCORRECT," which was visually displayed after each trial along with the correct face. In each block, each face appeared equally often with every other face; and for each pair of faces, each face appeared equally often as the left or right face. Participants completed successive blocks of training until they met the learning criterion, defined as 85% correct or higher in a single block, or until they completed eight training blocks. For all participants, the test phase began following the end of training.

During test, listeners heard 120 randomized sentences [4 talkers × 10 sentences (5 trained and 5 novel) × 3 repetitions] and were asked to indicate which talker produced each sentence. They made their choice from a fixed array of the four cartoon faces for each language. No feedback was provided at test. The experimental session lasted approximately 90 min.

3. Results

3.1 Training

Performance during training was calculated separately for each language in terms of (1) percentage correct talker identification during the first block of training and (2) number of training blocks required to reach the learning criterion. For both metrics, performance was calculated separately for the English and French voices. Examining performance in the first training block provides a dependent measure that is more directly comparable to studies that did not provide extensive talker identification training (e.g., Perrachione *et al.*, 2011; Perea *et al.*, 2014), whereas number of training blocks provides a measure of learning rate.

Mean performance for these two metrics is shown in Fig. 1 for each reading group. Consider first the performance during the first training block. Visual inspection of Fig. 1(a) suggests that both groups of readers showed increased talker identification for the English compared to the French talkers, consistent with previous findings showing a native language advantage for talker identification (e.g., Goggin *et al.*, 1991). However, with both English and French stimuli, the advanced readers show heightened talker identification compared to the average readers. To confirm this pattern statistically, mean talker identification in the first training block was submitted to analysis of variance (ANOVA) with the within-subjects factor of language (English vs French) and the between-subjects factor of reading ability (average vs advanced). The results confirmed a main effect of language ($F_{1,32} = 83.077$, $p < 0.001$, $\eta_u^2 = 0.722$), with

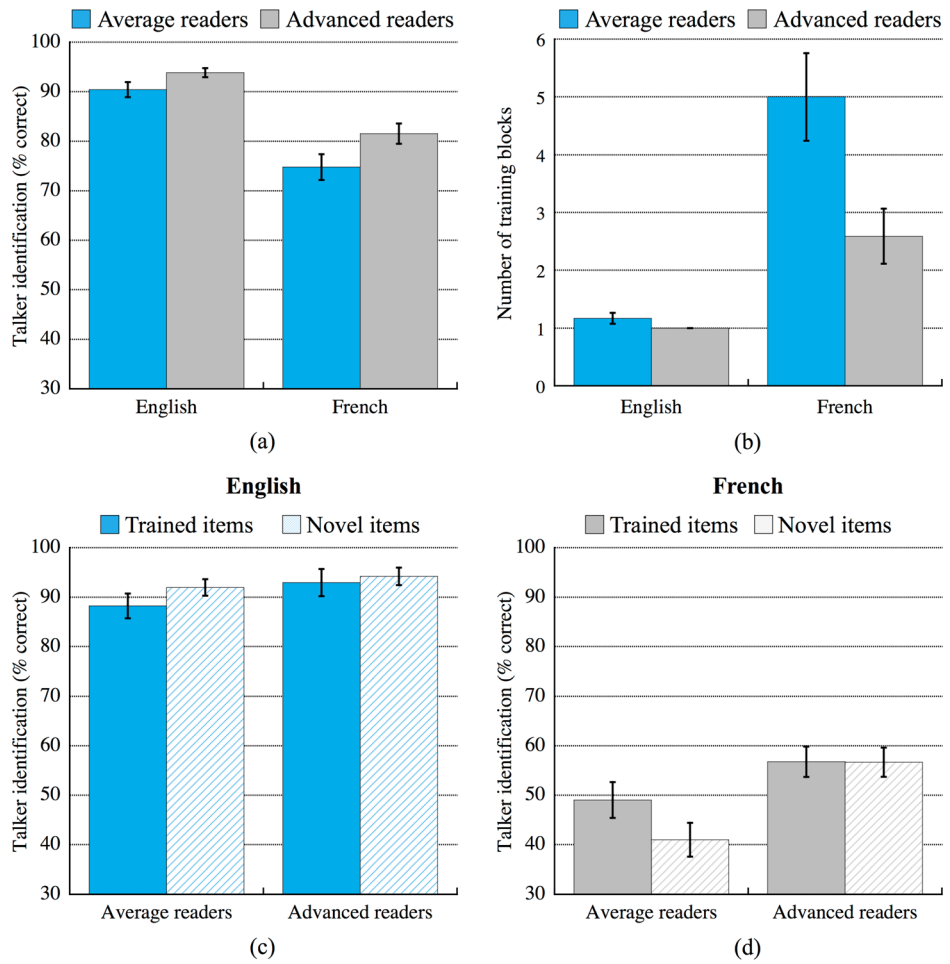

Fig. 1. (Color online) (a) shows mean talker identification in the first training block and (b) shows mean number of training blocks. Performance at test is shown in (c) for the English voices and in (d) for the French voices. Error bars indicate standard error of the mean. Chance performance was 50% in (a) and 25% in (c) and (d).

performance higher for the English ($M = 92.11$, $SD = 5.42$) compared to the French talkers ($M = 78.19$, $SD = 10.09$). There was also a main effect of reading ability ($F_{1,32} = 5.493$, $p = 0.025$, $\eta_u^2 = 0.146$) with performance in the advanced reading ability ($M = 87.70$, $SD = 4.88$) higher compared to the average reading group ($M = 82.60$, $SD = 7.52$). There was no interaction between language and reading ability ($F_{1,32} = 1.191$, $p = 0.283$, $\eta_u^2 = 0.036$).

Now consider performance with respect to the number of training blocks required to meet the learning criterion as shown in Fig. 1(b). Recall that the learning criterion was defined as reaching 85% accuracy in a training block or the completion of eight training blocks. All participants reached 85% correct accuracy for English; for French, one advanced reader and eight average readers ended training with the completion of eight training blocks. Visual inspection shows a robust effect of language on number of training blocks, such that for both reading groups, criterion was met with fewer training blocks for English compared to French talkers. However, the native language benefit appears to be mediated by reading ability, such that the difference between the English and French voices is attenuated for the advanced readers compared to the average readers. Number of training blocks was submitted to ANOVA following the structure outlined in the preceding text. The native language benefit for talker identification was confirmed such that fewer training blocks were required to meet criterion for the English ($M = 1.09$, $SD = 0.288$) compared to the French talkers ($M = 3.79$, $SD = 2.85$) ($F_{1,32} = 37.828$, $p < 0.001$, $\eta_u^2 = 0.542$). There was also a main effect of reading ability ($F_{1,32} = 7.902$, $p = 0.008$, $\eta_u^2 = 0.198$), with fewer training blocks required for the advanced ($M = 1.79$, $SD = 0.99$) compared to the average readers ($M = 3.09$, $SD = 1.62$). Moreover, the ANOVA showed a reliable interaction between language and reading ability ($F_{1,32} = 6.454$, $p = 0.016$, $\eta_u^2 = 0.168$). Independent *t*-tests showed that for the French talkers, the advanced readers required fewer training blocks to meet the learning criterion compared to the average readers ($t_{32} = 2.693$, $p = 0.011$, $d = 0.952$). For the English talkers, this trend was numerically present but did not reach threshold for statistical significance ($t_{32} = 1.852$, $p = 0.073$, $d = 0.655$).

3.2 Test

Performance at test was measured in terms of percent correct talker identification, calculated separately for the English and French voices and for the trained and novel sentences; mean performance for the average and advanced reading groups is shown in Figs. 1(c) and 1(d). Visual inspection of these figures suggests that both groups performed similarly at test for the English voices, but that the advanced readers outperformed the average readers for the French voices. Percent correct talker identification was submitted to ANOVA with the between-subjects factor of reading ability (average vs advanced) and the within-subjects factors of language (English vs French) and sentence type (trained vs novel). There was a significant main effect of language ($F_{1,32} = 448.255$, $p < 0.001$, $\eta_p^2 = 0.933$), as expected, with performance overall higher for the English ($M = 91.84$, $SD = 8.79$) compared to the French voices ($M = 50.86$, $SD = 13.61$). There was also a significant main effect of reading ability ($F_{1,32} = 5.861$, $p = 0.021$, $\eta_p^2 = 0.155$) with performance for the advanced readers ($M = 75.15$, $SD = 9.17$) higher compared to the average readers ($M = 67.55$, $SD = 9.13$). There was no main effect of trial type ($F_{1,32} = 0.522$, $p = 0.475$, $\eta_p^2 = 0.016$). There was no interaction between trial type and reading ability ($F_{1,32} = 1.601$, $p = 0.215$, $\eta_p^2 = 0.048$), but there was an interaction between language and reading ability ($F_{1,32} = 4.525$, $p = 0.041$, $\eta_p^2 = 0.124$) and between language and trial type ($F_{1,32} = 12.110$, $p < 0.001$, $\eta_p^2 = 0.275$).

Critically, the three-way interaction between language, reading ability, and trial type was reliable ($F_{1,32} = 7.576$, $p = 0.010$, $\eta_p^2 = 0.191$). Results of independent *t*-tests revealed that there was no difference between average and advanced readers for the trained English sentences ($t_{32} = -1.269$, $p = 0.214$, $d = -0.448$), the novel English sentences ($t_{32} = -0.931$, $p = 0.359$, $d = -0.329$), or the trained French sentences ($t_{32} = -1.627$, $p = 0.113$, $d = -0.575$), but that there was a striking difference between the two reading groups for the novel French sentences ($t_{32} = -3.474$, $p < 0.001$, $d = -1.228$), with performance higher for the advanced readers ($M = 56.67$, $SD = 12.18$) compared to the average readers ($M = 40.98$, $SD = 14.08$).

4. Discussion

There is a growing body of evidence indicating that listeners integrate talker identity and linguistic content in the course of speech processing. Specifically, research suggests that stability in phonological processing exerts a gradient influence on talker identification (e.g., Bregman and Creel, 2014; Perrachione *et al.*, 2011). Earlier work showed

decreased talker identification abilities in individuals with impaired reading ability (Perrachione *et al.*, 2011; Perea *et al.*, 2014), and the current data extend these findings to include an influence of phonological processing on talker identification within the unimpaired range of reading ability. The results from the training phase indicate that reading ability influenced the degree to which listeners could learn to identify the talkers' voices. The advanced readers showed increased talker identification accuracy in the first training block compared to the average readers for both the native and non-native voices, and the advanced readers required fewer training blocks to meet learning criterion for the non-native voices compared to the average readers. Composite reading score was significantly negatively correlated with number of training blocks required to learn the French talkers ($r = -0.384$, $p = 0.025$), consistent with other findings indicating a gradient role of reading ability on talker learning.

Reading-related differences in talker identification were also observed at test. The advanced readers showed overall better performance compared to the average readers. However, the robust interaction between reading ability, language, and trial type suggests that the locus of that main effect is that the average readers did not generalize learning of talkers' voices to novel French sentences to the same degree as the advanced readers. Indeed, there was a reliable correlation between composite reading score and performance on the novel French sentences ($r = -0.388$, $p = 0.024$) but not for the trained French sentences or either type of English sentence ($p > 0.160$ in all cases). Thus for the cases where the group analyses revealed influences of reading ability on talker identification, these influences were confirmed to be gradient in nature. The equivalent performance between the reading groups for the English sentences is consistent with the Bregman and Creel (2014) finding that even though bilingual listeners required increased time to learn voices in their second language, once learning had been achieved, listeners were able to generalize to the same degree as native listeners. Here the gradient effects of reading ability on talker identification were attenuated at test compared to training, which points toward a larger role of phonological knowledge in *learning* talkers' voices rather than in retention or generalization of talker learning.

The current work provides critical data in moving toward a principled account of the integration between talker identification and language ability. The finding that native language reading ability influenced non-native talker learning raises the possibility that the locus of the reading ability effect——and perhaps even the native language benefit for talker identification——is not limited to phonology (Perea *et al.*, 2014). Other potential contributing factors may be general auditory deficits (Ahissar, 2007) or a reduced ability to access and analyze pitch information (Xie and Meyers, 2015). That is, here we attribute the observed differences in the reading groups to an underlying difference in phonological ability that presents as differences in reading ability. An alternative possibility is that there is an underlying cognitive, auditory, or neural difference that drives performance on both the reading and talker identification measures. Future work is aimed at examining this possibility.

Acknowledgments

Research was supported by start-up funds to R.M.T. (University of Connecticut). Part of this project was completed as a Master's thesis by M.A.K. under the direction of R.M.T. Preliminary data were submitted to the 169th meeting of the Acoustical Society of America and the18th International Congress of Phonetic Sciences. We extend gratitude to William Baron and Jenean Jourdan for assistance with administering the assessment battery.

References and links

Ahissar, M. (**2007**). "Dyslexia and the anchoring-deficit hypothesis," Trends Cog. Sci. **11**, 458–465.

Bradley, L., and Bryant, P. E. (**1983**). "Categorizing sounds and learning to read: A causal connection," Nature **301**, 419–421.

Bregman, M. R., and Creel, S. C. (**2014**). "Gradient language dominance affects talker learning," Cognition **130**, 85–95.

Burnham, D. (**2003**). "Language specific speech perception and the onset of reading," Read. Writ. **16**, 573–609.

Fleming, D., Giordano, B. L., Caldara, R., and Belin, P. (**2014**). "A language-familiarity effect for speaker discrimination without comprehension," Proc. Natl. Acad. Sci. U.S.A. **111**, 13795–13798.

Goggin, J. P., Thompson, C. P., Strube, G., and Simental, L. R. (**1991**). "The role of language familiarity in voice identification," Mem. Cogn. **19**, 448–458.

Johnson, E. K., Westrek, E., Nazzi, T., and Cutler, A. (**2011**). "Infant ability to tell voices apart rests on language experience," Dev. Sci. **14**, 1002–1011.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**, 42–46.

Orena, A. J., Theodore, R. M., and Polka, L. (**2015**). "Language exposure facilitates talker learning prior to language comprehension, even in adults," Cognition **143**, 36–40.

Perea, M., Jimenez, M., Suarez-Coalla, P., Fernandez, N., Vina, C., and Cuetos, F. (**2014**). "Ability for voice recognition is a marker for dyslexia in children," Exp. Psychol. **61**, 480–487.

Perrachione, T. K., Del Tufo, S. N., and Gabrieli, J. D. (**2011**). "Human voice recognition depends on language ability," Science **333**, 595–595.

Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identification based on phonetic information," J. Exp. Psychol. Hum. Percept. Perform. **23**, 651–666.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–979.

Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., Brady, S. A., Thornton, R., Lundquist, E., Dreyer, L., Fletcher, J. M., Stuebing, K. K., Shaywitz, S. E., and Shaywitz, B. A. (**1995**). "Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax," Psychol. Sci. **6**, 149–156.

Theodore, R. M., and Miller, J. L. (**2010**). "Characteristics of listener sensitivity to talker-specific phonetic detail," J. Acoust. Soc. Am. **128**, 2090–2099.

Theodore, R. M., Myers, E. B., and Lomibao, J. A. (**2015**). "Talker-specific influences on phonetic category structure," J. Acoust. Soc. Am. **138**, 1068–1078.

Valji, A. (**2004**). "Language preference in monolingual and bilingual infants," Master's thesis, McGill University, Montréal, Québec, Canada.

Xie, X., and Myers, E. (**2015**). "The impact of musical training and tone language experience on talker identification," J. Acoust. Soc. Am. **137**, 419–432.