COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2017) 1–13

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12520

Abstraction and the (Misnamed) Language Familiarity Effect

Elizabeth K. Johnson,[a] Laurence Bruggeman,[b,c,d] Anne Cutler[d,e,f]

[a]*Department of Psychology, University of Toronto*
[b]*Department of Linguistics, Macquarie University*
[c]*ARC Centre of Excellence in Cognition and its Disorders*
[d]*The MARCS Institute, Western Sydney University*
[e]*ARC Centre of Excellence for the Dynamics of Language*
[f]*Max Planck Institute for Psycholinguistics*

Received 20 May 2016; received in revised form 25 February 2017; accepted 10 April 2017

Abstract

Talkers are recognized more accurately if they are speaking the listeners' native language rather than an unfamiliar language. This "language familiarity effect" has been shown not to depend upon comprehension and must instead involve language sound patterns. We further examine the level of sound-pattern processing involved, by comparing talker recognition in foreign languages versus two varieties of English, by (a) English speakers of one variety, (b) English speakers of the other variety, and (c) non-native listeners (more familiar with one of the varieties). All listener groups performed better with native than foreign speech, but no effect of language variety appeared: Native listeners discriminated talkers equally well in each, with the native variety never outdoing the other variety, and non-native listeners discriminated talkers equally poorly in each, irrespective of the variety's familiarity. The results suggest that this talker recognition effect rests not on simple familiarity, but on an abstract level of phonological processing.

Keywords: Talker recognition; Language familiarity effect; Native language; Dialect; Abstraction

1. Introduction

Although listeners readily identify familiar talkers in predictable situations, such as when a friend says "hello" on the telephone, under less ideal circumstances talker recognition accuracy is variable (e.g., Ladefoged, 2003; Yarmey, 1995). Listeners use acoustic cues to

Correspondence should be sent to Elizabeth K. Johnson, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada. E-mail: elizabeth.johnson@utoronto.ca

talker identity, such as fundamental frequency range and speaking rate (see Kreiman & Sidtis, 2011, for review), and also linguistic factors, such as phonetic cues (Andics, McQueen, & Turennout, 2007; Remez, Fellowes, & Rubin, 1997; Sheffert, Pisoni, Fellowes, & Remez, 2002). In particular, a crucial factor in predicting how accurately listeners identify talkers has been shown to be the identity of the language spoken: English speakers identify English-speaking talkers more accurately than they identify Spanish- or German-speaking talkers (Goggin, Thompson, Strube, & Simental, 1991). This has become known as the language familiarity effect (LFE) and it is highly robust (e.g., Bregman & Creel, 2014; Köster & Schiller, 1997; Levi & Schwartz, 2013; Perrachione, Pierrehumbert, & Wong, 2009; Thompson, 1987; Winters, Levi, & Pisoni, 2008). Its underlying cause, however, is not well understood. Here, we examine the level of processing involved, by testing listeners' recognition of talkers speaking in a familiar or unfamiliar regional variant of the native language.

Early evidence for the LFE came from cross-linguistic studies showing that identifying a talker in a forensic-style "voice line-up" is more difficult if the listener does not know the language the talkers speak. In Goggin et al.'s (1991) influential study, six German-English bilinguals were recorded in both German and English. Listeners were exposed to a short sample of speech by a target talker in one language, then, after a 5-min delay, to all six talkers (speaking the same language heard in the initial exposure). Their task was to pick the target talker. Native English speakers performed better when the bilinguals spoke English, whereas native German speakers performed better when the same individuals spoke German. Talker identification training studies have also shown that listeners can learn to identify talkers faster if they understand the language spoken by the to-be-learned talkers (Bregman & Creel, 2014; Winters et al., 2008).

An initial interpretation of the LFE (see, e.g., Köster & Schiller, 1997) was that comprehending a spoken message makes talker recognition easier. However, this now seems less likely, given that the LFE holds both for 7.5-month-old infants, with limited comprehension ability (Johnson, Westrek, Nazzi, & Cutler, 2011), and for late bilinguals with high comprehension proficiency, who nevertheless identify talkers less well than early bilinguals do (Bregman & Creel, 2014). Moreover, strongly accented though clearly comprehensible L2 speakers can be more difficult to identify than talkers with no accent (Goggin et al., 1991; Goldstein, Knight, Bailis, & Conver, 1981; McGhee, 1937; Thompson, 1987). Further, listeners rate Chinese and English speakers as less similar (i.e., can better distinguish them) when they speak in a familiar than in an unfamiliar language, even if the speech content is rendered unrecognizable by distortion (Fleming, Giordano, Caldara, & Belin, 2014). Thus, comprehension cannot explain the LFE.

Both Johnson et al. (2011) and Fleming et al. (2014) proposed instead that listeners draw on language-specific phonological knowledge to help discriminate talkers. In this view, talker recognition in unknown languages suffers because the phonology is unfamiliar. Note that this explanation is also compatible with the data suggesting that language proficiency is related in a gradient fashion to the strength of the LFE, as L2 phonology is acquired in a gradient manner (Archibald, 1998).

Unsurprisingly, most previous studies have confounded knowing how talkers in a given language community sound and knowing a phonological system. One approach that might tease these factors apart is to examine listeners' ability to distinguish talkers speaking in different variants of the native language, in particular variants that share an underlying abstract phonology (e.g., common syllable structure, common rhythm, common phonotactic constraints, largely common phoneme inventory). Regional accents differ in this respect from non-native accents, since the speech of second-language talkers is often colored by their native-language phonology (Archibald, 1998). If phonological processing indeed underlies the LFE, then listeners hearing a regional variant of their native language, even an unfamiliar one, should be able to distinguish between talkers (and, by extension, the F in LFE is an inaccurate description of the effect). If familiarity is all that matters, rarely heard variants will lead to a LFE just as unknown languages do.

Some studies have attempted to test the regional accent question. Impaired ability to distinguish speech in a different dialect compared to speech in one's own dialect has been claimed for comparisons of The Hague Dutch versus standard received Dutch (Kerstholt, Jansen, Van Amelsvoort, & Broeders, 2006), African-American Vernacular English (AAVE) versus general Midwestern American English (Perrachione, Chiao, & Wong, 2010), and for Glasgow Scottish English versus standard Southern British English (Stevenage, Clarke, & McNeill, 2012). Unfortunately all three of these comparisons pit a standard variant against a highly marked variant with lower socioeconomic status (The Hague Dutch, AAVE, Glasgow Scottish), a factor that has long been known to affect judgments of speaker identity (Lambert, Hodgson, Gardner, & Fillenbaum, 1960).

Furthermore, no study produced strong results. In the Dutch study, the difference in accuracy for the non-standard and standard target talkers did not reach significance, and no listeners from the Hague were tested. In the British study, listeners from both dialect areas were tested, but while southern British listeners showed poorer performance for the Scottish speakers compared with speakers of their own dialect, the Scottish listeners showed only small and inconsistent differences. The U.S. study also tested listeners from each dialectal group; the results for the two groups appear similar in a figure, but are not reported in the text. Thus not only were these investigations of the regional accent issue confounded with relative social status of the regional variant, but the evidence they provide is far from conclusive.

In this study, we compare two English variants without status difference: Australian English (of Sydney) and North American English (of Toronto, Canada). These two dialects are never mistaken for one another as they are phonetically distinct (many vowels and consonants are realized differently). However their underlying phonology is largely identical, including syllable structure, phonotactic constraints, consonant inventory, and rhythm. Simple familiarity is tested in that speakers of each variant are more familiar with their native variant than with the other variant.

2. Experiment 1: North American listeners

In Experiment 1, we test native North American English listeners on their ability to recognize individual talkers using North American English, Australian English, and

Dutch. We predict a LFE, in that performance with Dutch (which is unknown to all of these listeners) should, on any account, be worse than with English. Performance with Australian English (a variant of the native language, but an infrequently encountered one) should resemble performance in the native variant (i.e., North American English) if knowledge of abstract phonological structure is crucial, but it should resemble performance in the unfamiliar language (i.e., Dutch) if the LFE simply reflects accrued familiarity with a particular input type.

2.1. Participants

Forty-eight monolingual North American English speakers (29 females; age range: 18–51, mean: 22.3) with little exposure to Australian English and no knowledge of Dutch were tested at the University of Toronto. Participants received course credit or $10 for their participation.

2.2. Materials, design, and procedure

Sentence recordings by 12 female talkers were used (four Australian English talkers, four North American English [NAE] talkers, and four Dutch talkers). The Dutch and NAE recordings were those used in Johnson et al. (2011), and the Australian English recordings were modeled after the NAE set. The sentences were neutral statements such as *Artists have always been attracted by the life in the capital*, or *Een gevoel van enorme opluchting maakte zich van hem meester* ("he was overcome by a feeling of great relief") and were balanced across languages for length in phonemes and syllables (all sentences also fell in the range 12–16 syllables). All North American talkers were from the Greater Toronto Area; all Australian talkers from around Sydney, New South Wales.

The talkers were selected for similar voice quality, and we measured the duration, mean F0, and standard deviation in F0 of all utterances. Average measures are shown in the first 12 rows of Table 1. To assess relative variability across each talker set, for each measure and each language pair we compared variance across Set A versus Set B. The ratio of these two variances allows derivation of an F-value with degrees of freedom ($n-1$, $n-1$); here these degrees of freedom are (3, 3), yielding a critical F-value of 9.1 for significance at $p < .05$. The first three rows of Table 2 show the F-values for the three Experiment 1 language pairs; it can be seen that Fs for all comparisons across the two English variants are far below 9.1. The same is true for all F0 comparisons. In duration, the Dutch set was somewhat more uniform than the two English sets (specifically, one NAE talker and two Australian talkers spoke slightly more slowly than the others).

On each of 12 separate trials, participants heard two repetitions of one sentence spoken by one talker and were instructed to remember the talker. A 1-min distractor video clip with music but no speech followed. Participants were then presented with a voice line-up containing the target talker and three distractor talkers. The distractor talkers always spoke the same language (variant) as the target, and each of the four talkers uttered one (different) sentence. Participants judged which of the four talkers was the target. They

Table 1
Mean acoustic measures, with standard deviations in parentheses, for the 12 talkers used in all experiments (North American English [NAE], Australian English [AusE], and Dutch), plus the four Japanese talkers used in Experiment 3

Talkers	Duration (s)	Mean F0 (Hz)	sd F0
NAE 1	3.20 (0.32)	211 (6.7)	44 (5.7)
NAE 2	3.25 (0.25)	213 (5.8)	42 (4.0)
NAE 3	3.35 (0.38)	213 (9.6)	39 (9.2)
NAE4	3.62 (0.54)	233 (11)	48 (7.9)
AusE 1	3.66 (0.40)	209 (13.3)	44 (10.5)
AusE 2	3.83 (0.53)	199 (4.3)	46 (13.4)
AusE 3	3.74 (0.51)	193 (9.0)	50 (13.7)
AusE 4	3.26 (0.43)	214 (12.7)	54 (5.9)
Dutch 1	3.16 (0.35)	244 (15)	39 (5.0)
Dutch 2	3.06 (0.31)	201 (7.3)	41 (14)
Dutch 3	3.09 (0.32)	198 (7.6)	30 (8.0)
Dutch 4	3.14 (0.28)	225 (13)	46 (9.0)
Japanese 1	3.00 (0.33)	249 (5.8)	48 (3.0)
Japanese 2	3.00 (0.31)	214 (11)	45 (8.0)
Japanese 3	3.00 (0.28)	249 (14)	47 (9.0)
Japanese 4	3.10 (0.37)	230 (14)	47 (3.5)

Table 2
Comparison of acoustic variability of talkers for each language pair used in the experiments, in *F*-values calculated as the ratio of the two variances

	Duration	Mean F0	sdF0
NAE-AusE	1.77	1.13	1.88
NAE-Dutch	17.14*	4.34	2.98
AusE-Dutch	30.29*	4.92	2.51
NAE-Japanese	38.77*	2.46	16.93*
AusE-Japanese	68.53*	2.79	20.11*
Dutch-Japanese	2.26	1.76	50.5*

Critical *F*-values for significance at $*p < .05$ are >9.1. The comparisons in the top three lines occurred in all experiments, and those in the last three lines were in Experiment 3 only.

were instructed to guess if they were unsure. Each of the 12 talkers served once as target. All participants received the same trials. Trial order was randomized within four blocks, each containing one NAE, one Australian, and one Dutch trial. Individual testing sessions lasted 30 mins, including a break halfway through.

2.3. Results

Mean performance per Trial Type (NAE, Australian English, Dutch) was calculated for each participant (see Fig. 1). Performance across trial types was compared in a repeated-measures ANOVA followed by *t* tests.

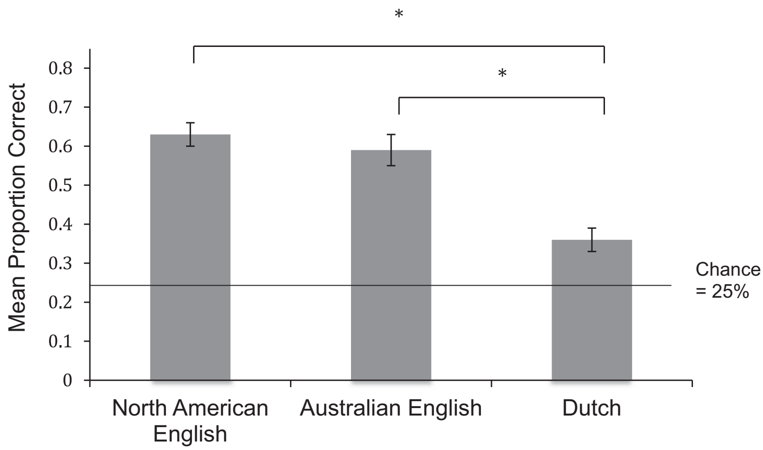

Fig. 1. Mean proportion correct in Experiment 1 by trial type (error bars indicate SE). NAE listeners performed significantly worse in Dutch voice line-ups than any other type of line-up. Performance did not differ between NAE and Australian English line-ups.

The ANOVA revealed a significant effect of Trial Type, $F(2, 94) = 16.01$, $p < .001$, $\eta_p^2 = .254$, and t tests showed significantly worse performance for the Dutch trials than for either NAE: $t(47) = 6.08$, $p < .001$, or Australian English: $t(47) = 3.88$, $p < .001$. Performance on NAE and Australian English trials did not differ; however, $t(47) < 1$, n.s.. Thus, the predicted LFE was found, but there was no effect of English variant, suggesting that listeners performed the task as easily in the non-native as in their native variant.

3. Experiment 2: Australian listeners

Although we endeavored to ensure comparability across talker sets, the two English-speaker sets could potentially have differed in some way that rendered the NAE talkers intrinsically hard to distinguish, or the Australian talkers relatively easy to distinguish, thus masking an inter-variant difference for the Experiment 1 listeners. Experiment 2 thus reversed the nativeness factor for the listeners, by testing Australian listeners. These listeners would be less familiar with NAE than with their native variant, although NAE is commonly heard in the media (and their perceptual skills for NAE phonemes even in noise are high; Cutler, Smits, & Cooper, 2005).

3.1. Method

Twenty-four monolingual Australian English speakers (20 females; age range: 18–49, mean: 22.9) without knowledge of Dutch were tested at Macquarie University in Sydney, and paid course credit or $15. Materials, design, and procedure were as in Experiment 1.

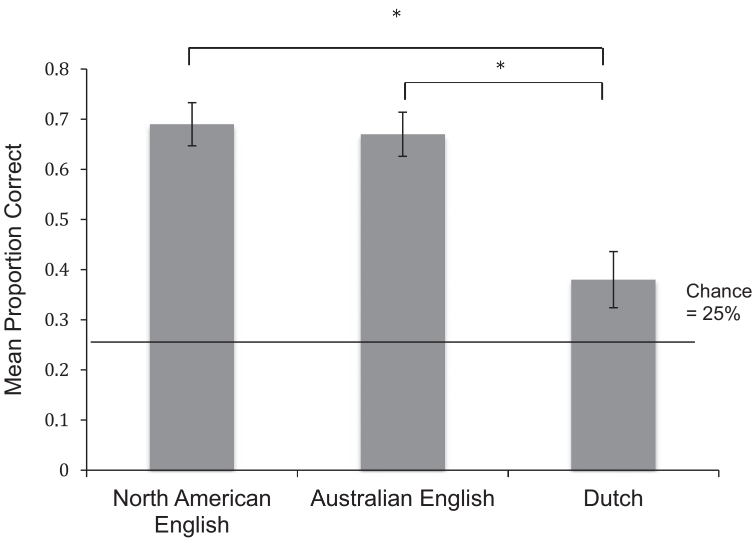

Fig. 2. Mean proportion correct in Experiment 2 by trial type (error bars indicate SE). Australian listeners performed significantly worse in Dutch voice line-ups than any other type of line-up. Performance did not differ between the two English line-ups.

3.2. Results

The results (see Fig. 2) were analyzed as for Experiment 1. The ANOVA showed a significant effect of Trial Type, $F(2, 46) = 12.671$, $p < .001$, $\eta_p^2 = .355$; t tests revealed the predicted LFE, in that participants performed significantly less accurately in the Dutch trials than in the Australian English trials, $t(23) = 4.16$, $p < .001$, or the NAE trials, $t(23) = 4.22$, $p < .001$. However, performance in the Australian English and NAE trials again did not differ, $t(47) = 0.37$, n.s..

4. Experiment 3: Dutch listeners

In Experiment 3, we test Dutch listeners on recognition of talkers in our three talker sets plus Japanese. For these listeners, Japanese is an unknown language; the Dutch-Japanese comparison thus tests for a LFE, which should again appear on any account. The NAE and Australian conditions allow us to test whether these two variants pattern similarly in relation to the native language Dutch (N.B. no prior LFE study of dialects has tested this factor in non-native stimuli), to the foreign language Japanese, and to each other.

Dutch listeners are proficient in English (being required to understand university lectures in the language). The variety used as target in schools is mainly British, but recognition skills for British and NAE are closely comparable, and high, even at high school age (Van der Haagen, 1998). Exposure to the present two English variants would, though, be asymmetrical. As foreign-language TV programs in the Netherlands are subtitled, not dubbed, Dutch listeners have many opportunities to hear NAE; but Dutch media rarely present Australian programs. In the only study we could find in which Dutch listeners heard Australian English, their scores on a repetition task were below their scores for Scottish-accented English (Mitterer & McQueen, 2009). If relative familiarity affects the LFE, NAE will produce better performance than Australian English. If only abstract phonological structure is relevant, performance on the two English variants will be equivalent.

4.1. Participants

Thirty-six native speakers of Dutch (29 females; age range: 18–27, mean: 20.8) with little exposure to Australian English or Japanese were tested, in return for a small payment, at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands. No participant used English at home or began learning it before fifth grade.

4.2. Materials, design, and procedure

The NAE, Australian, and Dutch materials were as in Experiment 1; Japanese materials were those used in Johnson et al. (2011). Acoustic comparisons (Table 2) showed that the Japanese talker set, like the Dutch set, had somewhat more uniform duration than the two English sets; but F0 standard deviation comparisons of Japanese and any other language were significant, because the Japanese measures, though higher, were more similar than in the other talker sets (see Table 1). Recall here that Japanese is a pitch accent language, in which pitch movements convey lexical distinctions independently of sentence contours. The four Japanese talkers were more similar than the other sets in how much pitch movement they used per sentence.

Design and procedure were as for Experiment 1, except that there were now 16 trials. Also, participants were asked to judge the mood of the character in the distractor music-only video clip preceding each voice line-up. In the line-ups participants heard two sentences spoken by each of four talkers and judged which talker was the target. Trial order was pseudo-randomized so that no language occurred more than twice in succession. Each individual test session lasted 40 min.

4.3. Results

Mean performance per Trial Type was again calculated for each participant (Fig. 3). Again the overall ANOVA revealed a significant effect of Trial Type, $F(3, 105) = 4.152$, $p = .008$, $\eta_p^2 = .106$. As predicted by the LFE, *t* tests revealed significantly better performance in the Dutch trials than in all other trial types [Dutch and Japanese: $t(35) = 1.99$, $p = .027$; Dutch and Australian English: $t(35) = 4.02$, $p < .001$; Dutch and NAE: $t(35) = 2.52$, $p = .008$]. Importantly, however, no other significant differences in performance were observed across trial types. Thus, Dutch listeners performed equivalently with all three of the non-native languages: Japanese, which they cannot understand, NAE,

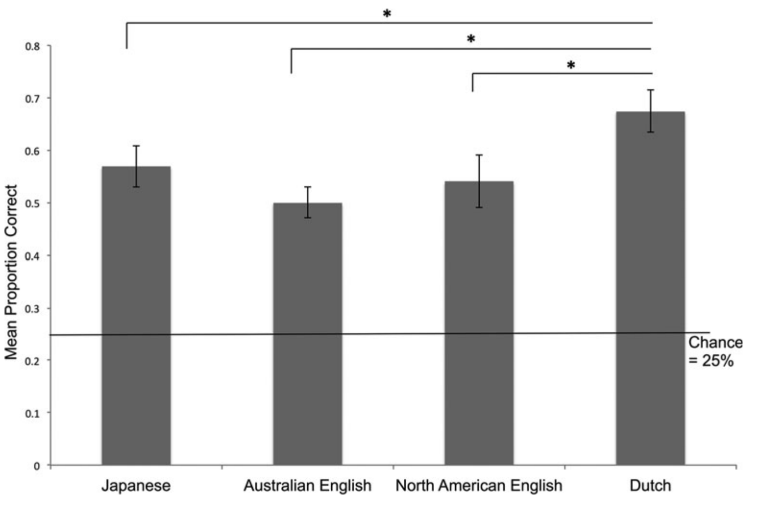

Fig. 3. Mean proportion correct in Experiment 3 by trial type (error bars indicate SE). Dutch listeners performed significantly better in Dutch voice line-ups than in any other line-up type. Performance did not differ significantly across the three foreign trial types.

which they can understand, and also hear relatively often, and Australian English, which they can understand but rarely hear. Again these results are as predicted by an explanation of the LFE in terms of abstract phonological structure: The LFE applies to the "foreign" versus "native" distinction, and neither comprehensibility nor accrued familiarity modifies this primary distinction.

5. Cross-experiment analyses

To assess the strength of the LFE, we conducted two analyses in which this factor was central, in that the Dutch listeners (native Dutch, foreign English) were compared to the English-speaking listeners (native English, foreign Dutch), and a language (averaged across variety) by listener group interaction was predicted. The items in these analyses were identical, as only Dutch and English data of the Dutch listeners were used. The predicted interaction was highly significant in each comparison: Experiments 3 versus 1, $F(1, 86) = 39.01$, $p < .001$, $\eta_p^2 = .322$; Experiments 3 versus 2, $F(1, 86) = 38.39$, $p < .001$, $\eta_p^2 = .398$.

6. General discussion

The language familiarity effect refers to the robust and consistent body of evidence that listeners find it easier to identify talkers speaking their native language than a non--native language. Our results suggest that it is misnamed, in that it is not driven by simple familiarity. A convincing explanation for the LFE has been lacking since an initial account in terms of listener comprehension (e.g., Köster & Schiller, 1997) was counter-indicated by evidence of an LFE even when the non-native input could be well comprehended, and also even when neither language could be comprehended at all. The present study has replicated the presence of an LFE even when the foreign language can be comprehended (Experiment 3) and thus has further strengthened the evidence against this account. However, our study overall has also moved the LFE debate closer toward resolution, by testing whether or not the LFE appears with speech in a dialectal variant of the native language, with no social status difference from the native variety, and whether or not the presence of an LFE is affected by familiarity with a non-native variant. In all three experiments, we observed a strong LFE, but it was quite unaffected by dialectal variation, either in the native or the non-native case. The two varieties of English used in this study shared the same underlying phonology, though they varied in how familiar they were to the listeners. Thus the results conform to the prediction whereby the "familiarity" in the LFE has a very narrow referent: it concerns only abstract phonological structure. The results do not fit with an account in terms of relative familiarity in the sense of frequency of exposure to the two varieties, since this factor produced no effect in any experiment. The driving force behind the LFE seems to be neither language comprehension nor the frequency of encountering talkers from different linguistic communities, but access to the phonological structure of a talker's utterance. It would be more appropriately termed the phonological attunement effect.

It is important that equivalence between performance with the two varieties of English was observed both with participants for whom English was their native language and those for whom English was a non-native second language. Previous results by Bregman and Creel (2014), which demonstrated that the LFE could not be viewed as a side-effect of speech comprehension, included a difference in talker identification success between early bilinguals and high-performing late bilinguals (with the latter group performing less well). Our present findings offer an elaboration of the account of this finding, by suggesting that the difference between the early and late bilinguals that was crucial for their performance in talker recognition is likely to have been their relative mastery of the target language phonology. Over-reliance on the phonological structures that characterize the native language is indeed a feature of both production and perception in a second language (Archibald, 1998); the earlier a listener's experience with a language, the more likely it is that knowledge of the phonology is fully in place.

Our findings offer some practical implications for the forensic sciences, suggesting that phonological structure might be included among the factors relevant to determining the credibility of ear witness testimony in court (Edmond, Martire, & San Roque, 2011), as well as theoretical implications for the development of talker recognition skills. The ability to recognize talkers improves between early and late infancy (e.g., Friendly, Rendall, & Trainor, 2014), and it continues to improve well into middle to late childhood (e.g., Creel & Jimenez, 2012; Levi & Schwartz, 2013; Mann, Diamond, & Carey, 1979). Children's increasing understanding of native-language phonology may be causally linked to this improvement. Furthermore, difficulty in recognizing talkers in infancy could potentially predict subsequent speech processing and reading skills (see Perea et al., 2014; Perrachione, Del Tufo, & Gabrieli, 2011, for related discussion).

The goal of our study was to shed further light on the underlying cause of the LFE in talker identification. Our results do not at all counter-indicate a role for familiarity in dealing with talkers; listeners' skills in recognizing friends and family from a syllable or less of input are real and useful. The results concern only the basis of recognition differences across languages. The results also do not imply that when variants of the same language display phonological differences (such as the non-stress-based rhythm of Singapore English [Deterding, 2001], or the reduplication processes and syllable structure simplifications of Jamaican English [Gooden, 2003]), talker recognition should remain unaffected; mapping the boundaries of phonological attunement is an important task for future research. Our findings certainly fit well with much evidence from learning tasks (Eisner & McQueen, 2005; Theodore & Miller, 2010) and from discrimination and decision making (Allen & Miller, 2004; Cutler et al., 2011), and within the larger framework proposed by Creel and Bregman (2011), all indicating that talker recognition and speech processing are tightly coupled operations. Speech perception is in fact always an interplay of effects at many different levels, from phonological knowledge to short-term memory; here we have located the basis of the so-called language familiarity effect in attunement to abstract phonology.

Acknowledgments

We acknowledge funding from SSHRC and NSERC (EKJ), from the Australian Research Council (grant CE140100041; AC) and from Macquarie University (grant MQSIS9201501719; LB). We thank Josh Romano for testing the Toronto participants.

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *115*, 3171–3183. https://doi.org/10.1121/1.1701898.

Andics, A., McQueen, J., & van Turennout, M. (2007). Phonetic context influences voice discriminability. In J. Trouvain & W. Barry (Eds.), *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1829–1832). Germany: Saarbrucken.

Archibald, J. (1998). *Second language phonology*. Amsterdam: John Benjamins.

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, *130*, 85–95. https://doi.org/10.1016/j.cognition.2013.09.010.

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Linguistics and Language Compass*, *5*, 190–204. https://doi.org/10.1111/j.1749-818X.2011.00276.x.

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*, 487–509. https://doi.org/10.1016/j.jecp.2012.07.007.

Cutler A., Andics A. & Fang Z. (2011). *Inter-dependent categorization of voices and segments*. In W. S. Lee & E. Zee (Eds.), *Proceedings of the Seventeenth International Congress of Phonetic Sciences* (pp. 552–555). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language versus non-native dialect. *Speech Communication*, *47*, 32–42. https://doi.org/10.1016/j.specom.2005.02.001.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, *29*, 217–230. https://doi.org/10.1006/jpho.2001.0138.

Edmond, G., Martire, K. A., & San Roque, M. S. (2011). “Mere guesswork”: Cross-lingual voice comparisons and the jury. *Sydney Law Review*, *33*, 395–425.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*, 224–238. https://doi.org/10.3758/BF03206487.

Fleming, D., Giordano, B. L., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences*, *111*, 13795–13798. https://doi.org/10.1073/pnas.1401383111.

Friendly, R. H., Rendall, D., & Trainor, L. J. (2014). Learning to differentiate individuals by their voices: Infants’ individuation of native- and foreign-species voices. *Developmental Psychobiology*, *56*, 228–237. https://doi.org/10.1002/dev.21164.

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*, 448–458. https://doi.org/10.3758/BF03199567.

Goldstein, A. G., Knight, P., Bailis, K., & Conover, J. (1981). Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society*, *17*, 217–220. https://doi.org/10.3758/BF03333718.

Gooden, S.A. (2003). The phonology and phonetics of Jamaican reduplication. PhD dissertation, Ohio State University.

Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*, 1002–1011. https://doi.org/10.1111/j.1467-7687.2011.01052.x.

Kerstholt, J. H., Jansen, N. J. M., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention, and telephone. *Applied Cognitive Psychology*, *20*, 187–197. https://doi.org/10.1002/acp.1175.

Köster, O., & Schiller, N. O. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, *4*, 18–28.

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444395068.

Ladefoged, P. (2003). Validity of voice identification. *The Journal of the Acoustical Society of America*, *114*, 2403. https://doi.org/10.1121/1.4778312.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, *60*, 44–51. https://doi.org/10.1037/h0044430.

Levi, S. V., & Schwartz, R. G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research*, *56*, 913–925. https://doi.org/10.1044/1092-4388(2012/12-0095).

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of Voice Recognition: Parallels with Face Recognition. *Journal of Experimental Child Psychology*, *27*, 153–165. https://doi.org/10.1016/0022-0965(79)90067-5.

McGhee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology*, *16*, 249–271.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, *4*(11), e7785. https://doi.org/10.1371/journal.pone.0007785.

Perea, M., Jiménez, M., Suárez-Coalla, P., Fernández, N., Viña, C., & Cuetos, F. (2014). Ability for voice recognition is a marker for dyslexia in children. *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)*, *61*, 480–487. https://doi.org/10.1027/1618-3169/a000265.

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. M. (2010). Asymmetric cultural effects on perceptual expertise underlie own-race bias for voices. *Cognition*, *114*, 42–55. https://doi.org/10.1016/j.cognition.2009.08.012.

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, *333*, 595. https://doi.org/10.1126/science.1207327.

Perrachione, T. K., Pierrehumbert, J. B., & Wong, P. C. M. (2009). Differential neural contributions to native- and foreign-language talker identification. *Journal of Experimental Psychology. Human Perception and Performance*, *35*, 1950–1960. https://doi.org/10.1037/a0015869.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 651–666. https://doi.org/10.1037/0096-1523.23.3.651.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology. Human Perception and Performance*, *28*, 1447–1469. https://doi.org/10.1037/0096-1523.28.6.1447.

Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The 'other-accent' effect in voice recognition. *Journal of Cognitive Psychology*, *6*, 647–653. https://doi.org/10.1080/20445911.2012.675321.

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, *128*, 2090–2099. https://doi.org/10.1121/1.3467771.

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*, 121–131. https://doi.org/10.1002/acp.2350010205.

Van der Haagen, M.J. (1998). Caught between norms: The English pronunciation of Dutch learners. PhD dissertation, Katholieke Universiteit Nijmegen.

Winters, S. J., Levi, S. V., & Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *The Journal of the Acoustical Society of America*, *123*, 4524–4538. https://doi.org/10.1121/1.2913046.

Yarmey, D. A. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, *1*, 792–816. https://doi.org/10.1037/1076-8971.1.4.792.