How pronunciation distance impacts word recognition in children and adults[a]

Tessa Bent,[1,b] Rachael F. Holt,[2] Kristin J. Van Engen,[3] Izabela A. Jamsek,[2] Lian J. Arzbecker,[2] Laura Liang,[2] and Emma Brown[1]

[1]Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, Indiana 47408, USA

[2]Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210, USA

[3]Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, Missouri 63130, USA

ABSTRACT:

Although unfamiliar accents can pose word identification challenges for children and adults, few studies have directly compared perception of multiple nonnative and regional accents or quantified how the extent of deviation from the ambient accent impacts word identification accuracy across development. To address these gaps, 5- to 7-year-old children's and adults' word identification accuracy with native (Midland American, British, Scottish), nonnative (German-, Mandarin-, Japanese-accented English) and bilingual (Hindi-English) varieties (one talker per accent) was tested in quiet and noise. Talkers' pronunciation distance from the ambient dialect was quantified at the phoneme level using a Levenshtein algorithm adaptation. Whereas performance was worse on all non-ambient dialects than the ambient one, there were only interactions between talker and age (child vs adult or across age for the children) for a subset of talkers, which did not fall along the native/nonnative divide. Levenshtein distances significantly predicted word recognition accuracy for adults and children in both listening environments with similar impacts in quiet. In noise, children had more difficulty overcoming pronunciations that substantially deviated from ambient dialect norms than adults. Future work should continue investigating how pronunciation distance impacts word recognition accuracy by incorporating distance metrics at other levels of analysis (e.g., phonetic, suprasegmental).

I. INTRODUCTION

From infancy through adolescence, children's awareness of phonetic variation increases, as do their abilities to recognize words with unfamiliar pronunciations. Within the first few years of life, they improve in their ability to handle phonetic variation arising from many sources including idiolect, gender, emotion, unfamiliar regional dialects, and nonnative accents (Best et al., 2009; Houston and Jusczyk, 2000; Singh et al., 2004; van Heugten and Johnson, 2012; van Heugten et al., 2015; van Heugten et al., 2018). Recent work suggests that some of the fundamental cognitive skills supporting perception of these variations are in place relatively early in development, including phoneme remapping and lexically guided retuning (McQueen et al., 2012; White and Aslin, 2011). However, the ability to process and cognitively represent variation arising from some sources, such as nonnative accents and regional dialects, appears to take many years to reach adult-like levels for both word recognition tasks and tasks tapping into sociolinguistic competence (e.g., dialect categorization) (Bent, 2018; Jones et al., 2017; McCullough et al., 2019a).

Many of the studies in this area have focused on young children's recognition of words produced with unfamiliar regional accents (Best et al., 2009; Kitamura et al., 2013; Mulak et al., 2013; Potter and Saffran, 2017; van der Feest and Johnson, 2016; van Heugten and Johnson, 2014, 2016; van Heugten et al., 2015), while fewer have used either constructed accents or nonnative accents (Paquette-Smith et al., 2020; van Heugten et al., 2018; Weatherhead and White, 2016). Due to the age of the children in these studies (primarily infants and toddlers), variations on the visual fixation paradigm, Preferential Looking Procedure, or Headturn Preference Procedure have been used most frequently. In these tasks, children's eye gaze or head turns are measured for either lists of words (e.g., high- vs low-frequency words) or pictures that match or mismatch auditorily presented words. Taken together, these studies suggest that children can recognize familiar words produced with unfamiliar accents by late in the second year of life (see summary in van Heugten et al., 2018).

Although children's abilities to cope with phonetic variation improve during the first few years of life, other lines of work have emphasized their continued development during the early school-aged years. For both regional dialects and nonnative accents, studies have found that children's social inferences and preferences (Creel, 2018; Dossey *et al.*, 2020; Kinzler and DeJesus, 2013a, 2013b; Weatherhead *et al.*, 2018) and word recognition abilities (Bent, 2014; Bent and Atagi, 2015, 2017; Bent and Holt, 2018; Bent *et al.*, 2019; Creel *et al.*, 2016; Dossey *et al.*, 2020; Holt and Bent, 2017; McDonald *et al.*, 2018; Nathan *et al.*, 1998) are continuing to develop during these years.

A range of social inference, decision-making, and sociolinguistic competence tasks have been used to study regional dialect and nonnative accent perception in children. In friendship preference tasks, children are presented with speakers who have different accents (e.g., native vs nonnative) and are asked with whom they would prefer to be friends. Children from approximately 5 years of age prefer to be friends with a native speaker over a nonnative speaker and this preference strengthens through 7 years (Creel, 2018; Kinzler *et al.*, 2009; Kinzler and DeJesus, 2013a), although evidence that a speaker is "nice" or "mean" can change these preferences (Kinzler and DeJesus, 2013a). With emotionally neutral content, 5- to 6-year-old children also rate native speakers as nicer than nonnative speakers (Kinzler and DeJesus, 2013a). Performance in tasks tapping into geography knowledge has shown somewhat mixed results. Kinzler and DeJesus (2013a) found that 5- to 6-year-old children identified a native speaker as more likely to be "living around here" or "American" compared to a French-accented speaker, but performance in Creel (2018) was poor in a similar location judgment task for 3- to 7-year-old children with some improvement over the age range tested. With American regional dialects, accurate locality judgments and discrimination abilities between dialects also begin to appear around age 5 and continue to improve throughout the teenage years (McCullough *et al.*, 2019b) and even into early adulthood (Dossey *et al.*, 2020). The ability to accurately group speakers by regional dialect in a free classification task also shows a long developmental trajectory stretching through adolescence (Jones *et al.*, 2017).

A number of studies have tested the word recognition abilities of 3- to 7-year-old children with unfamiliar accents (Bent, 2014; Bent and Atagi, 2015, 2017; Bent and Holt, 2018; Creel *et al.*, 2016; Nathan *et al.*, 1998), finding that children and adults show better word recognition for familiar than unfamiliar accents and dialects in open-set tasks. However, children continue to experience greater challenges from unfamiliar accents and dialects than adults throughout this age range. These difficulties have been observed with both word- (Bent, 2018; Nathan *et al.*, 1998) and sentence-length stimuli (Bent and Atagi, 2017; Bent and Holt, 2018; McDonald *et al.*, 2018), but are typically much larger in open-set than closed-set tasks (Creel *et al.*, 2016), and are exacerbated by background noise (Bent and Atagi, 2015; Bent and Holt, 2018). In addition to accuracy differences,

children tend to be slower to process speech produced by nonnative talkers than native talkers (McDonald *et al.*, 2018). Compared to adults, who are generally able to overcome variability from unfamiliar accents in quiet, children can struggle even in good listening conditions and show severe challenges for some unfamiliar accents in even moderate levels of noise (Bent and Atagi, 2015; Bent and Holt, 2018). The reason for children's reduced abilities to overcome variation stemming from unfamiliar accents has not been definitively determined, but there is evidence that some of their difficulties derive from a less robust use of contextual cues (Bent *et al.*, 2019), smaller vocabulary sizes (Bent, 2018; Levy *et al.*, 2019), and underdeveloped phonological processing skills (Bent and Atagi, 2017).

The studies that have directly compared sensitivity to differences among and between unfamiliar native and nonnative varieties show that 5- to 7-year-old children typically have more difficulty categorizing or discriminating between their home dialect and a different native variety than between their home dialect and a nonnative variety (Evans and Lourido, 2019; Floccia *et al.*, 2009a; Girard *et al.*, 2008; Paquette-Smith *et al.*, 2019; Wagner *et al.*, 2014). Additionally, children may rate nonnative accents as more distinct from their home variety than a regional variety (Weatherhead *et al.*, 2019). However, the sensitivity to these distinctions can be influenced by linguistic experience; in Evans and Lourido (2019), bilingual children showed higher accuracy than monolinguals in an accent categorization task for three accent comparisons tested (i.e., home-foreign, foreign-regional, and home-regional).

In studies comparing word recognition for nonnative and unfamiliar regional varieties, the results are mixed. Some work shows that children can overcome the unfamiliar pronunciation patterns present in regional dialects more readily than nonnative varieties (Bent and Holt, 2018), while other research has found greater difficulty in word identification for a regional variety than a nonnative one (Levy *et al.*, 2019). Another study that explicitly compared native and nonnative varieties showed no difference among the home, regional, and nonnative varieties, likely because children were near ceiling in the four-alternative forced-choice word identification task (Evans and Lourido, 2019). Similarly, an investigation of much younger children's word recognition with one regional and one nonnative variety showed no difference between the two when the participants were not provided with an adaptation phase (Paquette-Smith *et al.*, 2020).

The vast majority of studies on children's perception of unfamiliar accents or dialects have used one regional dialect or one nonnative accent (e.g., Best *et al.*, 2009; Nathan *et al.*, 1998; van Heugten *et al.*, 2018) as the test case for evaluating the impact of unfamiliar accents on perception. The relative dearth of studies investigating how multiple accents (including both regional and nonnative) impact children's speech processing has limited our understanding in this area. That is, claims about the developmental trajectory for children's abilities to overcome phonetic variation

4104   J. Acoust. Soc. Am. **150** (6), December 2021

Bent *et al.*

are very likely influenced by the acoustic-phonetic distance between the home variety and the unfamiliar variety employed in the study. Thus, studies employing accents that are more similar to one another may show earlier emergence of abilities to handle variation than those employing regional or nonnative varieties that are more distinct. Even within these varieties and categories (e.g., native vs nonnative), the specific talker's production patterns will strongly influence perceptual patterns: talkers with stronger accents or less proficiency in a second language are frequently more difficult to understand (Bent and Bradlow, 2003). A related body of research on the effects of mispronunciation distance (in toddlers and adults) further motivates including pronunciation distance metrics into studies of unfamiliar accent perception. Work in this area has shown that increased distance from a target pronunciation leads to less successful lexical access. For example, White and Morgan (2008) demonstrated that children as young as 19 months of age show gradient sensitivity to words with mispronunciations such that they more readily look to words with a smaller degree of phonological mismatch with the target word than words with greater degrees of mismatch. These types of findings from the mispronunciation literature suggest that including pronunciation distance metrics for naturally produced speech from talkers with different regional and nonnative accents may provide insight into why lexical access is difficult for particular listener groups or under specific listening conditions.

In addition to difficulties comparing studies that use a single accent or talker without quantification of pronunciation distance, across-study comparisons also can be challenging when different tasks are employed and studies rarely include more than one task, with some recent notable exceptions (Creel *et al.*, 2016; Dossey *et al.*, 2020; McCullough *et al.*, 2019a). For example, differences between children's processing of familiar vs unfamiliar accents were found to be quite small for closed-set tasks but substantially larger in open-set word recognition (Creel *et al.*, 2016). Additionally, children may be able to succeed in a Headturn Preference Procedure, Preferential Looking Procedure, or visual fixation task with an unfamiliar variety by age 18 months, but with a different variety or procedure they may fail to successfully recognize unfamiliar pronunciations.

The methodological differences in the talkers and accents tested as well as the tasks used make across-study comparisons challenging. Therefore, it is imperative that studies include multiple talkers and accents within the same procedure, as well as the same talker and accent with different procedures to understand how both variables affect performance. Last, the acoustic-phonetic differences from the child's native dialect must be quantified and related to perceptual patterns. This approach will lead to a deeper understanding of the cognitive-linguistic mechanisms underlying the development of word recognition.

Here, we take the approach of incorporating multiple native and nonnative accents, with one talker representing each accent, while keeping the task constant. Specifically,

we present new data testing 5- to 7-year-old children and adults' sentence recognition abilities with four different unfamiliar accents (some native, some nonnative) under various listening conditions (quiet and noise). We also include data from an additional two unfamiliar accents from previous work (Bent and Holt, 2018), which were collected under the same task conditions.

This study also serves to advance the field by examining how specific pronunciation patterns impact word recognition for different age groups and in different listening environments. As the initial step in the work, we incorporate Levenshtein distance, which quantifies the difference between two sequences or strings, in this case the extent to which unfamiliar accents differ from the ambient dialect at the phoneme level. These distances are calculated by comparing phonemic transcriptions of the familiar and unfamiliar accented productions of the same words or sentences. Talkers whose productions differ to a greater extent in the phonemic domain from the ambient dialect (in this case Midland American English) receive higher Levenshtein distances (see Sec. II for more detail). These scores allow for a quantification of difference among talkers without *a priori* assumptions about the extent to which specific accents or speakers may differ from the ambient dialect (e.g., it does not assume that nonnative talkers will be more distinct than native talkers with an unfamiliar regional accent). Furthermore, we can investigate how pronunciation distance influences word recognition at both a global talker level as well as at a more fine-grained word or sentence level. The investigation of how the extent and type of deviation from the ambient dialect impact word recognition accuracy across development is an essential next step for advancing our understanding of the mechanistic changes leading to improvements in children's abilities to map unfamiliar productions onto words in their lexicons. Critically, this approach can provide a crucial link between studies investigating fundamental mechanisms underlying listeners' abilities to cope with phonetic variation (e.g., lexically guided phonetic retuning) and those testing word recognition and sentence comprehension.

Although Levenshtein distances have been compared with foreign accent or dialect strength ratings (Bartelds *et al.*, 2020; Gooskens and Heeringa, 2004; Wieling *et al.*, 2014a; Wieling *et al.*, 2014b) and intelligibility across related Scandinavian languages (Beijering *et al.*, 2008), this metric has only been used once before to our knowledge in relation to recognition of unfamiliar accents by children (Levy *et al.*, 2019). In Levy *et al.* (2019), a different target language (German) was employed with two unfamiliar accents (one regional and one nonnative). Levenshtein distances were reported at the talker level as an indicator of overall distance from the home standard, but the distances were not incorporated into the statistical analyses in terms of their relation to intelligibility scores. The study also did not include adults in the analyses or measure the impact of noise. By incorporating Levenshtein distances into the statistical analyses of word identification accuracy and including six unfamiliar accents, the data presented here allow for

J. Acoust. Soc. Am. **150** (6), December 2021

Bent *et al.*   4105

a broader assessment of how specific talker characteristics impact spoken word recognition at two different points in the lifespan. While our study is still limited to one talker per accent, it is an initial step towards a more comprehensive view of the impact of unfamiliar pronunciations on word recognition and will allow for novel assessments of how pronunciation distance impacts word recognition.

II. METHOD

A. Participants

Listeners included 292 monolingual American English-speaking adults and children. Adults ($n = 112$; 60 female) were between the ages of 18 and 35 years with an average age of 23.5 years (standard deviation, SD = 4.1). Children ($n = 180$; 90 female) were between the ages of 5 and 7 years with equal numbers of 5-, 6-, and 7-year-old children. We also include data from 90 children and 96 adults (reported in Bent and Holt, 2018), who were tested under the same conditions but were presented with different talkers. All adults and parents of child participants reported typical speech, language, and hearing. An additional 48 participants were tested, but their data were excluded because they did not fit the language background criteria, including bilingual or multilingual language backgrounds (four adults; two children); reported high exposure to one or more of the accents or dialects included in their assigned condition (15 adults; 4 children); reported atypical speech/hearing (one adult; six children); did not meet age inclusion criteria (three adults); technical/equipment error (one adult; one child); refused to assent to the project (one child); or did not complete the procedure (ten children). Ratings of exposure to a range of dialects and accents were obtained from adults *via* self-report and for children *via* parental report using a 1–5 scale, where $1 = $ no exposure and $5 = $ frequent daily exposure. For the ambient regional dialect (Midland American English), both adults and children had high ratings with the adult average of 4.8 (range $= 1$–5) and children with an average of 4.6 (range $= 1$–5). Ratings for the accents and dialects included in the participant's condition (see more detail below) were also collected. The exposure ratings were much lower for all non-ambient accents (German-accented English $= 1.2$ for adults and 1.0 for children; Scottish English $= 1.2$ for adults and 1.0 for children; Hindi-accented/Indian English $= 1.5$ for adults and 1.2 for children; Mandarin-/Chinese-accented English $= 1.5$ for adults and 1.1 for children). Participants also identified their home dialect by self-or parental report. Most respondents indicated that their home dialect was Midland ($n = 205$), a combination of Midland and another dialect ($n = 3$), or North Central ($n = 35$). Other home dialects included Appalachian ($n = 9$), Southern ($n = 8$), West ($n = 5$), Western Pennsylvania ($n = 5$), and New York City ($n = 3$). The remaining respondents included other dialects ($n = 7$) or did not provide a response ($n = 12$). For the children, parents indicated that English was the only ($n = 174$) or primary ($n = 6$) language spoken in the home. Most adults had studied another language, but none of the included participants reported fluency in any language other than English.

B. Stimuli

The sentence stimuli were 60 sentences selected from the Hearing In Noise Test for Children (HINT-C) (Nilsson *et al.*, 1996). These sentences are simple declaratives composed of words (including three to four keywords per sentence) that should be highly familiar to children. Five female talkers were recorded reading the sentences. One speaker was a 23-year-old native monolingual English speaker from the Midland dialect region. The other four speakers included a Scottish English speaker, a native German speaker, a native Mandarin speaker, and a Hindi-English bilingual speaker. The German and Mandarin speakers were selected from the Hoosier Database of Native and Nonnative Speech for Children (Bent, 2014). Recordings of the nonnative speakers and the Midland speaker are available on SpeechBox (Bradlow, n.d.). The native German speaker was 29 years of age, had been living in the U.S. for 1.7 years, and started studying English at age 10. The native Mandarin speaker was 23 years of age, had been living in the U.S. for three years, and started learning English at age 8. The Scottish and Hindi speakers were recorded specifically for this study at Ohio State University and Indiana University, respectively. The Scottish speaker was 24 years of age and had lived in the U.S. for two years prior to the recording. The Hindi speaker was a simultaneous Hindi-English bilingual, was 18 years of age, and had been in the U.S. for 4 months at the time of recording.

To quantify the segmental characteristics of the talkers included in the study, an adapted version of the Levenshtein Distance Algorithm was used (Levy *et al.*, 2019). The traditional Levenshtein algorithm (Levenshtein, 1966) uses a binary coding method to compare two pronunciations of a word across speakers, dialects, or languages. In these comparisons, two phonemic transcriptions of a word are aligned so that the number of operations (i.e., substitutions, deletions, additions) to change one word into another word are determined. The alignment is optimized to find the alignment that results in the fewest operations. Each of these operations is given the same penalty (i.e., 1) and these penalties are summed to determine the distance between the two pronunciations of the word. The adapted method used here utilizes a similar approach but provides for more gradual variation depending on the type of error. This scoring method was used by Levy *et al.* (2019) under the assumption that not all phonemic changes carry equal weight in perception. The penalty weights in the adapted algorithm were derived from concepts presented in Pettersson *et al.* (2013), in which lower weights are assigned for more frequently occurring variations. Greater penalties for consonant than vowel errors are consistent with the literature (Gao, 2019) showing that consonant deviations have greater consequences for perception of nonnative-accented speech than vowel errors. This metric quantifies the phonemic differences between the speakers of

non-ambient dialects/accents compared to speakers from the ambient dialect.

To make these comparisons, each of the 60 sentences from the talkers used in this study and in Bent and Holt (2018) was phonemically transcribed. In addition, transcriptions were made for three additional Midland talkers from the Hoosier Database. All sentences were phonemically transcribed by two transcribers, at least one of whom was from the Midland region. Transcriptions were based primarily on perceptual analysis but were supported by observing the acoustic characteristics of the speech including waveform and spectrogram visualizations. After the two transcribers independently completed their transcriptions, the two transcriptions were compared. Where there were disagreements, the initial two transcribers met with a third transcriber (first author T.B.) to determine the final transcription. The transcriptions for the talkers from the non-ambient dialects/accents were compared to transcriptions for the four Midland productions (i.e., the Midland speaker used here as well as the three other Midland speakers in the Hoosier database). If a non-ambient dialect speaker's production matched any of the four Midland speakers' productions, they were not penalized, but if the production was not observed in any of the four Midland speakers' productions, they received a penalty. Differences between the non-ambient dialect speakers and Midland speakers' productions were then calculated based on the following from Levy et al. (2019):

- Vowel substituted by another vowel = 0.5
- Consonant substituted by another consonant = 0.75
- Phoneme insertion = 1.0
- Change to word length = $1/\log10(\max(\text{length}(\text{Word1}), \text{length}(\text{Word2})))$ (where Word1 = number of phonemes in the non-ambient dialect speaker's production and Word2 = number of phonemes for the ambient dialect production)
- Other (e.g., deletions, vowel to consonant substitution, consonant to vowel, etc.) = 0.4

The score for each word was calculated by summing the penalties described above resulting in a single score per word. Note that words with insertions or deletions do not necessarily receive two penalties (e.g., if one phoneme was deleted and one was inserted total word length would not change and thus no additional penalty would be assigned). Higher numbers indicate that the speaker deviates further from the Midland speakers, while a score of 0 indicates that their productions were the same as one or more of the Midland speakers. Based on an average across all words, the speakers received the following scores: Japanese (0.663), Hindi (0.403), Mandarin (0.308), German (0.286), British (0.263), and Scottish (0.136). A Levenshtein calculation example for a single sentence is shown in Table I.

C. Procedure

All participants were recruited and tested in the Language Sciences Lab at the Center for Science and

TABLE I. Example of transcriptions for one sentence and associated Levenshtein scores at the word level.

Accent	The	House	Has	Nine	Bedrooms
Midland 1 (used in intelligibility tests)	ðə	haʊs	hæz	naɪn	bɛdɹʊmz
Midland 2	ðə	haʊs	hæz	naɪn	bɛdɹʊmz
Midland 3	ðə	haʊs	hæz	naɪn	bɛdɹʊmz
Midland 4	ðə	haʊs	hæz	naɪn	bɛdɹʊms
Levenshtein score			n/a		
British	ðə	haʊs	hæs	naɪn	bɛdɹʊmz
Levenshtein score	0	0	0.75	0	0
German	də	haʊs	hɛs	naɪn	bɛdɹʊmz
Levenshtein score	0.75	0	1.25	0	0
Scottish	ðɛ	haʊs	hæz	naɪn	bɛʒɹʊmz
Levenshtein score	0.5	0	0	0	0.75
Mandarin	ðə	haʊs	hɛs	naɪ	bɛdʐɹums
Levenshtein score	0	0	1.25	2.50	0.75
Hindi	ðəʔ	haʊs	hæz	naɪn	bədtʃɹums
Levenshtein score	3.10	0	0	0	2.61
Japanese	dɛ	haʊs	hɛz	laɪm	bɛlomz
Levenshtein score	1.25	0	0.5	1.5	2.83

Industry (COSI) in Columbus, Ohio. Participants were assigned to one of two accent conditions. In each condition, they were presented with sentences from three talkers representing three different accents/dialects. In one condition, listeners were presented with the Midland, German-accented, and Scottish talkers. In the other condition, the listeners were presented with the Midland, Hindi/Indian English, and Mandarin-accented talkers. For the data from Bent and Holt (2018), listeners were presented with Midland, Japanese-accented, and British talkers. Each talker contributed 20 sentences to each condition and the specific sentences assigned to each talker were counterbalanced across listeners. Within the accent conditions, participants were assigned to a noise or quiet condition. The noise condition consisted of sentences that were mixed with an 8-talker babble (Van Engen et al., 2014) at a signal-to-noise ratio (SNR) of +4 dB. For the trials with babble, sentences were mixed with a random selection from the babble that was one second longer than the sentence, so that there was 500 ms of babble before the sentence began as well as a 500-ms babble tail after the sentence ended.

Testing took place in a quiet lab in the museum. All stimuli were presented at a comfortable listening level. Before the start of the experimental trials, listeners were presented with nine practice trials in which they heard three sentences from each of the talkers included in their condition. No feedback other than general encouragement was provided during the practice and experimental trials. Following the practice trials, the 60 experimental trials blocked by accent were presented with the order of the accents counterbalanced across listeners. During all trials, listeners were presented with one sentence at a time binaurally over Audiotechnica headphones (model 8TH-770COM). They were instructed to repeat the sentence out loud to the best of their ability. An experimenter, who was

trained by the second author (R.F.H.), recorded the participant's response in real time and scored the responses offline. If the experimenter was unsure of a participant's response, they would ask follow-up questions to clarify, ask the participant to repeat, point, or describe the word in question. Because children with typical speech development in this age range are highly intelligible (Flipsen, 2006) and earlier work using very similar methods showed very few discrepancies between initial and second orthographic transcriptions (Bent and Atagi, 2017), we did not make audio recordings of the participants' responses for reliability checking.

Stimulus presentation was controlled with E-Prime v. 2.0 (Psychology Software Tools, 2007) on a Dell Optiplex 790 desktop computer. All testing procedures were approved by the local institutional review board (IRB) and, as is customary in museum laboratory settings, participants were not compensated.

III. RESULTS

A. Adults and children in quiet and noise

Word recognition accuracy was analyzed using generalized linear mixed effects models with a logit link function to account for the binomial outcome measure (i.e., correct or incorrect). All words in the sentences were entered into the model. For most words, a strict scoring criterion was applied so that words were coded as incorrect if they included added or deleted morphemes. The only exceptions to this rule were for a(n)/the, has/had, have/had, is/was, and are/were. These

alternations were counted as correct per the original scoring criteria for the HINT-C. Figure 1 displays individual means and group means (adults/children) for each accent in both quiet and noise conditions.

Fixed effects for this analysis included age group (adults vs children), listening conditions (quiet vs noise), and accent (seven levels; dummy coded with Midland as the reference level). Random intercepts were included for participant and item (models that included random slopes were not able to converge).

To assess whether our three fixed effects significantly affected word recognition in general, we first built a model including all three of them (output in Table II) and compared it to nested models that omitted each factor individually.

The model comparisons showed significant effects of accent ($\chi^2 = 9716.80$, $p < 0.001$), age ($\chi^2 = 265.53$, $p < 0.001$), and noise ($\chi^2 = 371.59$, $p < 0.001$). Adults performed better than children and performance in quiet was better than in noise. Furthermore, the model output (Table II) shows that performance on each accent differed significantly from performance on the Midland accent (p-values in the output are based on asymptotic Wald tests). The negative parameter estimate for each accent indicates that, in every case, participants were less accurate on the non-Midland accent.

Next, we built models to investigate two- and three-way interactions among age, listening condition, and accent (Table III). We first compared the fit of a model that
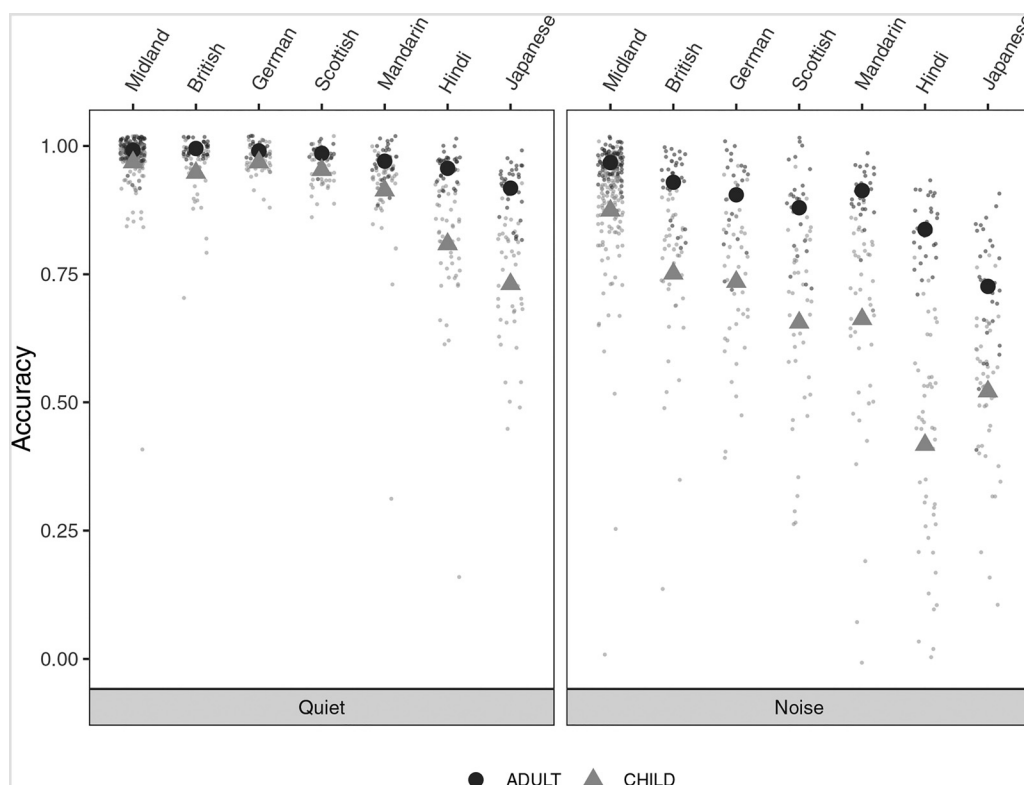

FIG. 1. Mean word recognition accuracy for adults (dark circles) and children (light triangles) in quiet (left) and in noise (right). Small dots represent individual participant means in each condition.

TABLE II. Output of the mixed effects model that includes the fixed effects of Age, Condition, and Accent (without interactions). Estimates correspond to log odds of a correct response. The intercept corresponds to performance in the Midland condition averaged over noise and age; the age estimate indicates that the odds of a correct response increase by 1.566 for adults compared to children; the condition estimate indicates that those odds increase 1.913 for quiet compared to noise. Estimates for all accents are negative; that is, the odds of a correct response are lower for all accents than for the Midland dialect.

	Estimate	Standard error	z-value	p-value
Intercept (Midland)	4.085	0.068	60.49	<0.001
Age (Adult vs Child)	1.566	0.082	19.05	<0.001
Condition (Quiet vs Noise)	1.913	0.080	24.00	<0.001
Accent – British	−0.957	0.046	−20.62	<0.001
Accent – German	−1.013	0.049	−20.81	<0.001
Accent – Scottish	−1.445	0.047	−30.66	<0.001
Accent – Mandarin	−1.349	0.044	−30.98	<0.001
Accent – Hindi	−2.461	0.042	−58.61	<0.001
Accent – Japanese	−2.619	0.043	−61.58	<0.001

included all two-way interactions among age, noise, and accent to that of a model that also included the three-way interaction. The three-way interaction did not yield a statistically significant improvement to model fit ($\chi^2 = 8.72$, $p = 0.19$). To investigate the significance of the two-way interactions, we compared the model with all three interactions to nested models that excluded each one. These comparisons showed that accent x noise ($\chi^2 = 253.47$, $p < 0.001$) and accent x age ($\chi^2 = 112.69$, $p < 0.001$) significantly improved model fit, but age x noise ($\chi^2 = 0.61$, $p = 0.44$) did not. The output of the model that included age,

TABLE III. Output of the mixed-effects model that included fixed effects for age group, listening condition, and accent, along with the significant two-way interactions (age group x accent and listening condition x accent).

	Estimate	Standard error	z-value	p-value
Intercept	3.921	0.070	55.675	<0.001
Age Group (Adult vs Child)	1.415	0.098	14.473	<0.001
Condition (Quiet vs Noise)	1.553	0.093	16.780	<0.001
Accent - British	−0.595	0.065	−9.138	<0.001
Accent - German	−0.707	0.069	−10.207	<0.001
Accent - Scottish	−1.102	0.065	−16.992	<0.001
Accent - Mandarin	−1.227	0.058	−21.253	<0.001
Accent - Hindi	−2.145	0.054	−39.719	<0.001
Accent - Japanese	−2.618	0.051	−50.651	<0.001
Age x Accent (British)	0.495	0.114	4.359	<0.001
Age x Accent (German)	0.023	0.119	0.191	0.849
Age x Accent (Scottish)	0.154	0.114	1.349	0.177
Age x Accent (Mandarin)	0.069	0.108	0.637	0.524
Age x Accent (Hindi)	0.642	0.102	6.284	<0.001
Age x Accent (Japanese)	−0.101	0.095	−1.059	0.289
Condition x Accent (British)	0.663	0.110	6.051	<0.001
Condition x Accent (German)	0.975	0.120	8.096	<0.001
Condition x Accent (Scottish)	1.007	0.112	8.994	<0.001
Condition x Accent (Mandarin)	0.342	0.095	3.592	<0.001
Condition x Accent (Hindi)	0.601	0.090	6.695	<0.001
Condition x Accent (Japanese)	−0.288	0.090	−3.186	0.001

listening condition, accent and the significant two-way interactions is displayed in Table III.

Inspection of this output shows that the interaction between age and accent is driven by the British and Hindi accents. In each case, children showed larger deficits for the non-ambient accent (relative to the Midland accent) than adults. The interaction between listening condition and accent is driven by differences in the effects of noise for all accents relative to the Midland accent. That is, noise is generally more detrimental to the recognition of non-ambient accents than the ambient one.

B. Children: Developmental patterns

In a second analysis, we investigated the effect of age on children's performance for these accents across listening conditions. The children in this study ranged in age from 5 to 7 years. For this analysis, their age in months was centered and scaled, and included in the statistical models as a fixed effect. Random effects were the same as in the first analysis. We assessed the main effect of age by first fitting a model that included age and listening condition (averaged over accents) and compared it to a model that included listening condition only. The addition of age significantly improved model fit ($\chi^2 = 36.575$, $p < 0.001$); older children generally performed better than younger children (Fig. 2).

To assess interactions among age, accent, and noise condition, we then fit a model that included their three-way interaction along with all two-way interactions and compared it to a model that omitted the three-way interaction. The three-way interaction significantly improved model fit ($\chi^2 = 16.037$, $p = 0.014$).

To assess this three-way interaction, we ran separate analyses of the data collected in quiet and noise. In quiet, the interaction between age and accent significantly improved model fit relative to a model that included age and accent only ($\chi^2 = 23.299$, $p < 0.001$). Additionally, the model output indicated that age significantly interacted with performance on the Japanese accent ($p = 0.003$) and marginally on the Mandarin accent ($p = 0.079$), with older kids outperforming younger kids. In noise, the interaction between age and accent also significantly improved model fit compared to a model that included age and accent only ($\chi^2 = 19.483$, $p = 0.003$). Model output indicated that the interaction was driven by performance on the Hindi accent ($p = 0.003$), again with older children outperforming younger children. Model outputs for the analyses of the child data are included in the Appendix.

C. Levenshtein distance

To investigate the effect of Levenshtein distance on recognition accuracy, we analyzed the word recognition accuracy data with the distance metric included for each token *rather* than information about the particular accent of a speaker. It is worth noting that the distance variable is highly skewed: of the 1860 individual word tokens for the speakers of the non-ambient dialect, 1326 of them had a

J. Acoust. Soc. Am. **150** (6), December 2021

Bent *et al.* 4109
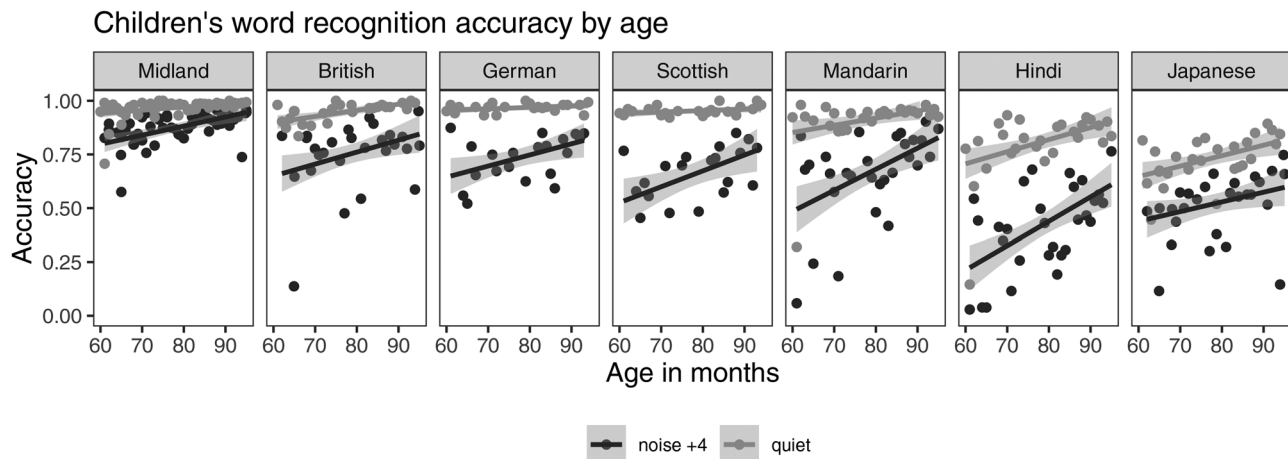
Children's word recognition accuracy by age

FIG. 2. Each panel shows the relation between children's age in months and word identification accuracy in quiet (light) and in noise (dark) for one of the accents. Each dot represents a single child. Linear fits with standard errors are included to aid in visualization.

distance score of 0. That is, these words did not differ from the Midland dialect with respect to their phonemic transcriptions. In addition, of course, all words produced by the Midland speaker also had scores of 0 by definition. To first assess the main effect of Levenshtein distance, we compared a model that included fixed effects for age group (adult vs child), noise condition (quiet vs noise), and distance to one that omitted distance. Distance contributed significantly to model fit ($\chi^2 = 2030.4$, $p < 0.001$), indicating that greater distance (i.e., higher Levenshtein score) reduced the likelihood of correct word identification (Fig. 3).

To investigate interactions among distance, age group, and listening condition, we next built a model including all

three of these fixed effects along with their two- and three-way interactions. This model was a better fit to the data than one that omitted the three-way interaction ($\chi^2 = 3.857$ $p = 0.0495$). To assess this significant three-way interaction, we next fit separate models for quiet and noisy conditions with age group, distance, and their interaction as fixed effects.

In quiet, the model including the interaction between age and Levenshtein distance did not provide a significantly better fit to the data ($\chi^2 = 0.4559$, $p = 0.4996$); indicating that although the adults performed better than the children overall, the effect of distance did not differ across the two age groups in quiet (or, equivalently, the effect of age did not differ across distance). In noise, the inclusion of the
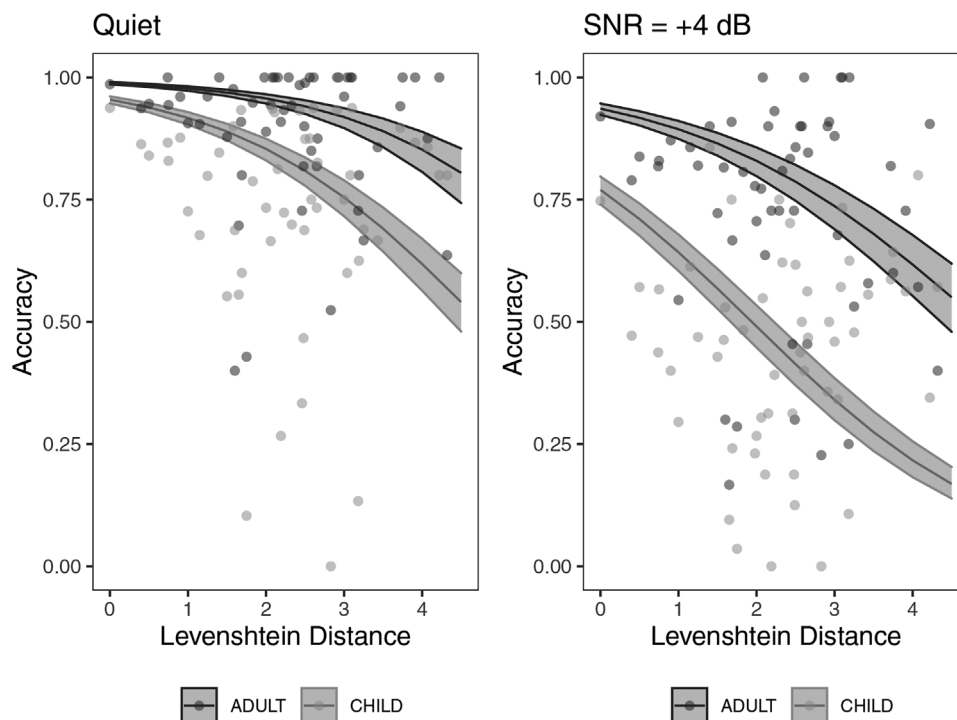

FIG. 3. Word recognition accuracy in quiet (left) and in noise (right) as a function of Levenshtein distance. Individual data points represent mean accuracy for each age group at each distance. Lines represent model predictions with 95% confidence intervals.

interaction effect significantly improved model fit (χ^2 = 5.3652, $p = 0.021$): the effect of distance was steeper for the children than for the adults.

D. Distance scores by accent

Across speakers, the proportion of words that received Levenshtein scores above zero (i.e., deviated from the Midland at the segmental level) varied considerably, ranging from only 14% of words for the Scottish speaker (44/310) to 51% for the Japanese speaker (157/310). The distribution Levenshtein scores for each talker are presented in Fig. 4.

To statistically investigate whether distance and accent made independent contributions to accuracy, we built a regression model that included fixed effects of age group, accent, and condition (no interactions) and another one that also included distance. The comparison of these models revealed that distance significantly improved fit ($\chi^2 = 186.25$, $p < 0.001$). That is, a model that included both accent and distance fit the data better than one that included accent only. The effects of distance and accent (in quiet and noise for both age groups) are visualized in Fig. 5.

IV. DISCUSSION

This study examined word identification accuracy for school-aged children and adults in both quiet and noise-added conditions for seven different accents with one talker representing each accent. These accents included the ambient native accent (Midland American English), non-native accents (Japanese-, Mandarin-, and German-accented English), less familiar native accents (British and Scottish English), and a bilingual accent (Hindi/Indian English). Thus, this dataset represents one of the broadest

investigations into how different accents impact word recognition at two points in development. Further, this study is the first to incorporate Levenshtein distances at the word level into statistical models of word recognition.

The results showed the expected main effects of accent, listening environment, and listener age on word recognition accuracy: accuracy was lower for all non-ambient accents than the ambient accent, for the noise-added condition than the quiet condition, and for children than adults. Further, all non-ambient accents were significantly more adversely impacted by noise compared to the ambient accent. Some interaction effects, however, diverge from prior work. Specifically, we did not observe a statistically significant interaction between age and listening condition (i.e., quiet vs noise), which may seem surprising because there are many reports in the literature of children having more difficulty in noise compared to adults (Elliott et al., 1979; Fallon et al., 2000). On average, adults correctly identified 98% of the words in quiet (SD = 0.15) and 90% in noise (SD = 0.30), while the children correctly identified 91% in quiet (SD = 0.28) and 70% in noise (SD = 0.46). The high degree of variability in the children's performance, therefore, rendered the interaction non-significant even though they dropped by 21% points from quiet to noise while the adults dropped by only eight points. It is important to note that, while the previous studies used analyses of variance (ANOVAs) for data analysis, the current study used generalized linear mixed effects models. Because individual variability can be modeled with random effects, this analytical approach controlled for individual differences in performance that arise from general difficulty with speech in noise or other cognitive and linguistic factors that depress performance for children (e.g., immature cognitive abilities,

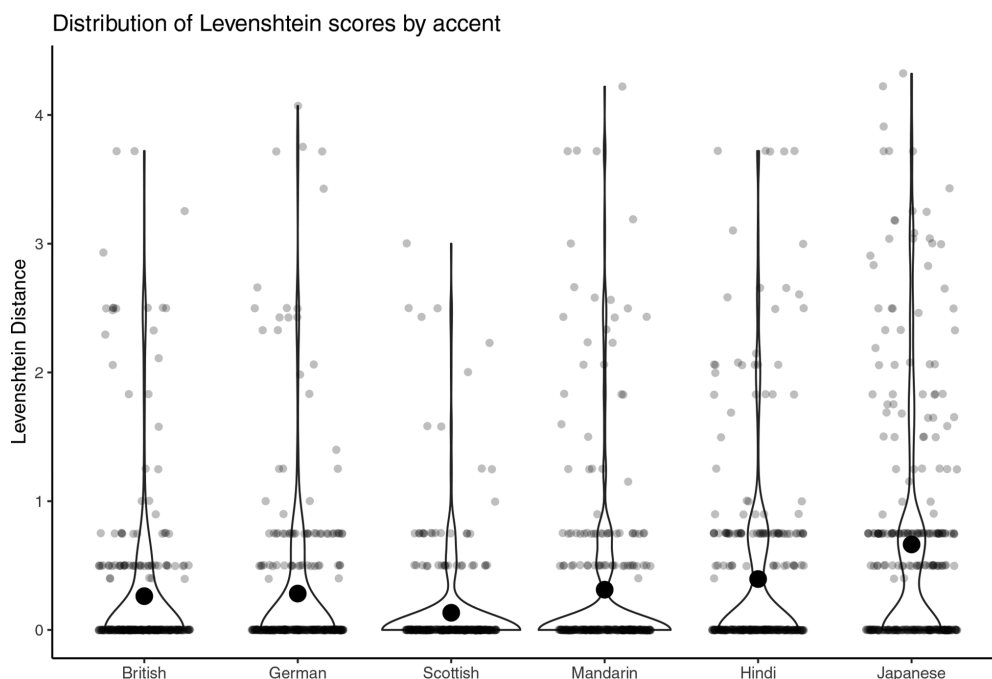Distribution of Levenshtein scores by accent

FIG. 4. Violin plots showing the distribution of distance metrics for all words in each of the non-ambient accents. Large dots indicate means, small dots individual words. The total number of words per speaker was 310. All 310 have a distance of 0 for the Midland speaker by definition.
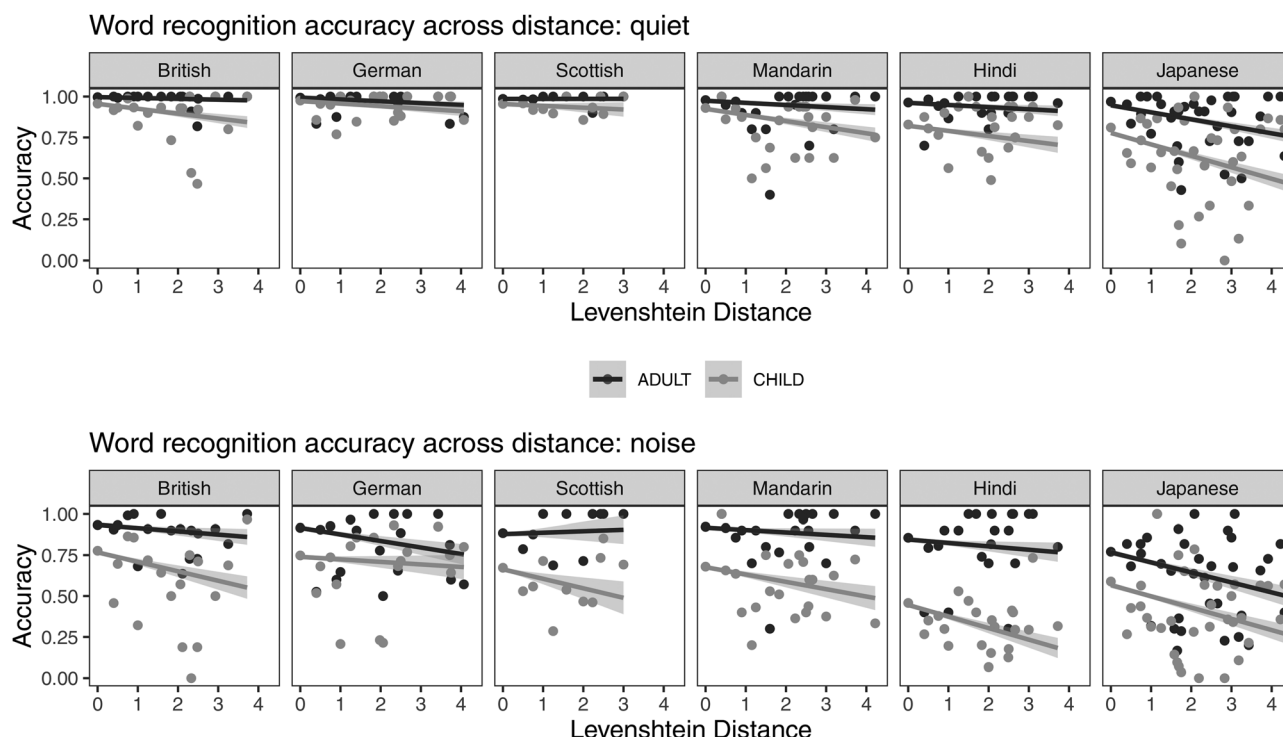
J. Acoust. Soc. Am. **150** (6), December 2021

Bent et al.   4111

FIG. 5. Word recognition accuracy (*y* axis) as a function of Levenshtein distance (*x* axis) in each accent for children and adults in quiet (top panel) and noise (bottom panel). Each dot represents the average accuracy for the children (light) or adults (dark) for the associated Levenshtein score. Linear fits with standard errors are included to aid in visualization.

including decreased attention and memory; or linguistic factors, such as smaller vocabulary sizes). Similarly, the inclusion of random effects allowed us to control for differences in item intelligibility that are not due to accent or noise (e.g., lexical frequency). Finally, this analysis allowed us to set the ambient accent (Midland American English) as the reference level to which each other accent was compared (obviating the need for *post hoc* comparisons).

When controlling for individual listener differences in word recognition, we only observed interactions between listener age (children vs adults) and talker accent for two of our accents, British English and Hindi-accented English. These interactions arose because children showed more word recognition difficulty for these two accents relative to Midland than adults. For the British talker, this difference appears to be primarily driven by children's performance in the noise condition; that is, both children and adults showed near-ceiling performance on the British talker in quiet; similarly, adults were also highly accurate in noise, whereas children showed substantial decrements with the British talker in noise. In contrast, the interaction between age and accent for the Hindi speaker can be seen in both the quiet and noise conditions. Children have more difficulty than adults with the Hindi speaker in quiet and show the lowest performance for this talker in noise. The Hindi accent was also the accent that demonstrated the largest age effect for the child-only developmental analysis; children made substantial gains in understanding this speaker across the age range tested here (5–7 years of age). These results

demonstrate that word recognition patterns and developmental trends will crucially depend on the specific talkers and accents included in the study. Therefore, claims made about children's relative success or difficulty at overcoming divergences from the ambient dialect at specific ages may not always generalize to other accents or talkers. To obtain a more complete understanding of children's development of coping with phonetic variability in unfamiliar accents, a wider range of talker accents and listener ages are necessary.

In many previous studies, advantages in word recognition for unfamiliar native over nonnative varieties were observed (Adank *et al.*, 2009; Bent and Holt, 2018; Bent *et al.*, 2016; Floccia *et al.*, 2006; Goslin *et al.*, 2012), while others have shown different patterns (Evans and Lourido, 2019; Floccia *et al.*, 2009b; Levy *et al.*, 2019; Paquette-Smith *et al.*, 2020). The results here caution against strong claims regarding the impact of talker native language status on intelligibility. For example, although there was some variation across age groups and listening conditions, the Hindi-English speaker frequently had the lowest word recognition accuracy scores, even though she is a native English speaker. Similarly, listeners frequently showed similar or better performance on the German- or Mandarin-accented talkers than the Scottish talker. Although the Hindi-English bilingual talker's productions are influenced both by her native variety (Indian English) as well as her status as a bilingual speaker who grew up speaking both English and Hindi, the Scottish talker was a monolingual English

speaker. Therefore, prior claims about nonnative accents being more difficult or requiring different processing strategies than native varieties for adults and/or children may have arisen due to the specific varieties or talkers selected; they should likely not be generalized to claims about differences in perception for native vs nonnative talkers more broadly. In fact, even though the language learning situations are clearly different between native and nonnative speakers, there may not be fundamental differences in how listeners process regional vs nonnative varieties. Rather, from the listener's perspective, nonnative speech may frequently be more difficult to understand due to substantial phonemic deviations from the ambient dialect.

Rather than making assumptions about how particular accents or classes of accents (e.g., native vs nonnative) may impact word identification, these results suggest that it is essential to characterize the accents included with objective measures of distance from the ambient dialect. The inclusion of the Levenshtein scores in this investigation is a first step towards providing insight into how production distance impacts word recognition accuracy. Here, we demonstrated that Levenshtein distances were significant predictors of word recognition accuracy. That is, as productions deviated further from the ambient dialect norms, word recognition accuracy decreased. Moreover, the impact of these distance scores was different across the quiet and noise-added conditions. In quiet, although adults performed better than the children, the effect of distance did not differ for the two age groups. In the noise-added condition, however, children showed greater decrements in word recognition accuracy with increasing distance from the ambient dialect than adults. These results suggest that even in optimal listening conditions, productions that diverged more substantially from the ambient accent were harder for children and adults to identify. In more effortful conditions, however, the mapping between the less familiar pronunciations and words in their lexicons becomes increasingly difficult for children. One explanation for this interaction is that the addition of noise may decrease the cognitive resources available for making these particularly challenging mappings. Another possibility is that in addition to quantitative differences that have been previously observed between adults and children (Bent, 2018; Bent and Atagi, 2015; Bent and Holt, 2018), there may be qualitative changes across development. That is, children's abilities to handle specific types of deviations from the ambient dialect may change throughout development. Although the data presented here cannot determine whether there are qualitative changes across development, future work should investigate how not only how the general distance between the ambient and non-ambient productions impact word recognition accuracy but also how specific types of production differences may impact children and adults differently. For example, future work could select specific items that include pronunciation differences of various types (e.g., consonant changes, differences in word length, non-ambient phonemes, etc.) and compare children's and adults' abilities to identify these items. Furthermore, we

used a binary scoring method in which incorrect responses could have been of multiple types (e.g., no response, incorrect word, nonword). A more detailed analysis of the types of responses provided by children and adults could provide insight into whether there are qualitative differences in their perception and/or response strategies. Other scoring metrics, such as fuzzy string matching (Bosker, 2021), could also be employed. These scoring methods could provide a more gradient view into children's and adults' responses. That is, it is possible that although a child and an adult both incorrectly perceived a word spoken in an unfamiliar accent, the word retrieved by the adult could be closer to the target item than the child's percept. Finally, the current procedures only measure word recognition accuracy but do not necessarily tap into word or sentence comprehension. Future studies could incorporate tasks such as word definitions (e.g., Nathan et al., 1998) or questions tapping comprehension following the presentation of speech samples.

The quantifications of accent distance may be useful metrics to add into models and frameworks of speech understanding. For example, "accented speech" is one of the input-related factors in the Framework for Understanding Effortful Listening (Pichora-Fuller et al., 2016). Likewise, these quantifications could be incorporated into the Ease of Language Understanding model to predict whether implicit or explicit processing of language input would be required (Rönnberg et al., 2013). The results here suggest that incorporating specific distance metrics may be fruitful for precise modeling of word recognition along with the other quantified factors in the models. The use of Levenshtein distances also builds upon and complements other approaches quantifying pronunciation distance among native and nonnative speakers. For example, Floccia and colleagues have attempted to equate accent strength across native and nonnative talkers by incorporating listener accent ratings (e.g., Floccia et al., 2009a), but as they note, the strategies used by listeners to make these ratings may differ depending on whether the talker is perceived as a native or a nonnative speaker (Floccia et al., 2009b). Other work has directly measured the similarity of the talkers' and the listeners' accents to predict intelligibility, showing that acoustic similarities in vowel spectral and duration characteristics (as measured by ACCDIST; Huckvale, 2004) predict intelligibility (Pinet et al., 2011; Stringer and Iverson, 2019). However, they require explicit measurement of the specific listener's accent and only capture information about vowels.

Although Levenshtein scores are a first step in quantifying distance from the ambient dialect, they only capture distance at the phoneme level. Words receiving zero scores are not necessarily produced as a Midland speaker would produce them. Differences at the sub- or supra-segmental levels are not captured by this metric. More fine-grained metrics including quantification of narrow transcription or specific acoustic-phonetic measurements should be incorporated into perception studies as well. It could be particularly fruitful to

J. Acoust. Soc. Am. **150** (6), December 2021

Bent et al.   4113

investigate word recognition accuracy across words that received Levenshtein scores of zero (i.e., did not differ phonemically from the Midland speakers). The characteristics of words that were difficult to identify even without phonemic deviations could then be more precisely determined. This approach would build upon work using word learning paradigms showing that toddlers' abilities to generalize newly learned words across talkers are hindered by productions in which productions cross a phoneme boundary compared to productions with sub-phonemic differences (Newman *et al.*, 2018).

The Levenshtein scoring method employed here also did not account for the frequency with which certain types of errors were present (Wieling *et al.*, 2014a). That is, deviations that are more common may be more easily overcome during word recognition than productions that occur less frequently or are less familiar. Here, for example, if a talker substituted an alveolar stop for an interdental, they would receive the same penalty as if they substituted a fricative with a different place of articulation (e.g., /z/ for /ð/). The stop substitution may be familiar to listeners due to its frequency in a number of dialects spoken within (e.g., African American Language, New York City dialects) and outside the U.S. as well as in children's speech. There is also evidence that children are better and faster at identifying words with more common than less common misarticulations (Krueger *et al.*, 2018). A larger corpus than that used in the current study would be needed to obtain measures of frequency across different production differences. In addition, the Levenshtein measure used here does not capture information about variability within a speaker. Although assumptions about nonnative speakers' productions being more variable—and consequently yielding a greater challenge for the listener—have recently come into question (Vaughn *et al.*, 2019), the exploration of how production variability, in addition to distance, would be valuable. Finally, future research into adaptations of the Levenshtein algorithm should investigate whether penalty weights should be adjusted depending on listening conditions. For example, the adaptation used here from Levy *et al.* (2019) assigned higher penalties for consonant substitutions than vowel substitutions, but it is possible that vowel substitutions may have a greater impact on intelligibility in noisy environments while consonant substitutions could be more detrimental in quiet conditions.

The different relations between Levenshtein scores and word recognition accuracy across the accents suggest that some accents or talkers may have other aspects of their productions that are leading to reductions in word recognition accuracy that are not related to phonemic distance. Although all talkers had words that substantially deviated from the Midland dialect, children and adults were still able to overcome these deviations for some talkers showing high word recognition accuracy across Levenshtein scores (e.g., for the German-accented talker). In contrast, some talkers (e.g., the Hindi- and Japanese-accented talkers) showed very steep declines with increasing Levenshtein scores. There are

several possible explanations for this result. First, in addition to the segmental deviations captured by the Levenshtein scores, these talkers may have produced suprasegmental deviations that, in combination with the segmental deviations, made word recognition more difficult. The Levenshtein metric does not capture differences in suprasegmentals, including possible deviations in stress, intonation, or rhythm. Future work should continue to quantify distances in these other dimensions to determine how individually or in combination these differences from the ambient dialect impact word recognition across development and in different listening conditions. Second, scoring in this study was done at the word level without reference to word type (e.g., function vs content word), word characteristics (e.g., word frequency) or word position within a sentence, all factors that could impact word recognition (e.g., Howes, 1957). It is possible that the same types of deviations from Midland norms would have different effects depending on these lexical and sentential factors. For example, a substitution such as /ɛ/ for /ə/ in "the" may have a quite different impact on word identification than a similar substitution pattern in a content word (e.g., /ɛ/ for /ʌ/ in "funny"). Third, factors about the preceding or following words also may impact the level of difficulty beyond the production characteristics of the target word. For example, two talkers could have a word with an identical Levenshtein score, but one of the talkers may have had overall higher Levenshtein scores and more words with non-zero scores than the other talker. Thus, the talker with more phonemic deviations overall would likely have multiple words within a single sentence that deviated from Midland norms. In contrast, a talker with a lower overall Levenshtein score (e.g., the German-accented talker) could have sentences with only a single word that deviated from Midland norms. The listeners' abilities to identify a word that deviated from native norms in sentences where all other words were produced in a more familiar way would be easier than the same production where many of the other words are also produced in ways that challenged the lexical mapping process. Further, because listeners use context to aid in word recognition in difficult listening conditions (e.g., Holt and Bent, 2017; Kalikow *et al.*, 1977), a word that was misidentified early in a sentence because of its phonemic deviations may impact word recognition later in the sentence even for words with similar Levenshtein scores.

Although this study provided a broad investigation into perception of unfamiliar accents by adults and children, there are still several limitations to the design. First, there was only one talker representing each accent. Therefore, claims about the accent cannot necessarily be generalized beyond the specific talker included in this study. For example, our German-accented talker was relatively easy for our listeners to understand, but, of course, a German-accented talker with a stronger accent whose productions diverged further from Midland norms would presumably be more difficult for listeners to understand. Similarly, British or Scottish talkers with different residential histories or socio-economic statuses may have different production patterns

4114   J. Acoust. Soc. Am. **150** (6), December 2021

Bent *et al.*

than the talkers we employed here, again impacting word recognition in different ways. Future work should include multiple talkers representing each accent with quantification of their segmental deviations from the ambient dialect to determine which aspects of the patterns observed here can be generalized more broadly to an accent and which are more idiosyncratic and reside at the talker level. The stimuli were also limited to short, read sentences with simple, early acquired vocabulary. With spontaneous speech, longer sentences, or speech with more complex vocabulary, production differences across native and nonnative speakers (e.g., differences in fluency or speech rate) may become more apparent and have larger impacts on listener perception. In particular, it will be essential to assess materials that may be more challenging for adults to determine how lexical and syntactic aspects of the stimuli would interact with pronunciation distance and/or noise for materials that increase task difficulty. Therefore, future studies should also incorporate a wider range of materials.

V. CONCLUSION

The results from this study point to the need for further development of objective methods for quantifying pronunciation distance and relating these measures to perceptual patterns in children and adults. This study provided a step in this direction by investigating the relation between Levenshtein distances, a measure of phonemic distance, and school-aged children's and adults' perception of multiple native and nonnative accents. The results showed that Levenshtein distances related to word recognition accuracy for both adults and children. Furthermore, Levenshtein distance and listener age interacted in the noise condition, such that children had increasing difficulty identifying words compared to adults as pronunciation distance from the ambient accent increased. Future studies of adults' and children's perception of less familiar accents should continue to incorporate measures of phonemic distance and begin to quantify divergences from the ambient dialect at the sub- and supra-segmental levels.

ACKNOWLEDGMENTS

APPENDIX

Fixed effects output for analysis of children's performance:

1. Model with three-way interaction among age, accent, and condition (quiet vs noise):

accuracy $\sim$ age_scaled * accent * condition
+ (1 | participant) + (1 | list_sent_word)
Fixed effects:

	Estimate	Std.Error	z value	p-value
(Intercept)	3.214512	0.073682	43.627	< 0.001 ***
age_scaled	0.396206	0.056569	7.004	< 0.001 ***
British	−0.827762	0.065645	−12.610	< 0.001 ***
Mandarin	−1.261099	0.055238	−22.830	< 0.001 ***
German	−0.744486	0.069116	−10.772	< 0.001 ***
Hindi	−2.464970	0.052086	−47.325	< 0.001 ***
Japanese	−2.602196	0.054604	−47.656	< 0.001 ***
Scottish	−1.192999	0.064301	−18.553	< 0.001 ***
condition	1.655129	0.114053	14.512	< 0.001 ***
age_scaled:British	0.070997	0.064054	1.108	0.0267685
age_scaled:Mandarin	−0.038741	0.053825	−0.720	0.471675
age_scaled:German	0.015642	0.066186	0.236	0.813173
age_scaled:Hindi	0.038733	0.051287	0.755	0.450111
age_scaled:Japanese	−0.175041	0.053705	−3.259	0.001117 **
age_scaled:Scottish	0.028885	0.062003	0.466	0.641313
age_scaled:condition	0.021333	0.113017	0.189	0.850279
British:condition	0.647714	0.131064	4.942	< 0.001 ***
Mandarin:condition	0.372591	0.110162	3.382	< 0.001 ***
German:condition	0.875598	0.137856	6.352	< 0.001 ***
Hindi:condition	0.609695	0.103373	5.898	< 0.001 ***
Japanese:condition	−0.408799	0.108250	−3.776	< 0.001 ***
Scottish:condition	0.903670	0.128232	7.047	< 0.001 ***
age_scaled:British:cond	−0.004809	0.128102	−0.038	0.970057
age_scaled:Mandarin:cond	−0.243765	0.107594	−2.266	0.023476
age_scaled:German:cond	0.075120	0.132178	0.568	0.569814
age_scaled:Hindi:cond	−0.290351	0.102388	−2.836	0.004571 **
age_scaled:Japanese:cond	−0.222612	0.107397	−2.073	0.038192
age_scaled:Scottish:cond	−0.077232	0.123773	−0.624	0.532642

2. Children's data in quiet only:

accuracy $\sim$ age_scaled * accent + (1 | participant)
+ (1 | list_sent_word)
Fixed effects:

	Estimate	Std.Error	z value	p-value
(Intercept)	4.09455	0.09678	42.308	< 0.001 ***
age_scaled	0.40863	0.07641	5.348	< 0.001 ***
British	−0.50240	0.11290	−4.450	< 0.001 ***
Mandarin	−1.08866	0.09237	−11.786	< 0.001 ***
German	−0.29014	0.11993	−2.419	0.01555 *
Hindi	−2.18872	0.08440	−25.933	< 0.001 ***
Japanese	−2.82900	0.08816	−32.091	< 0.001 ***
Scottish	−0.72814	0.10946	−6.652	< 0.001 ***
age_scaled:British	0.09052	0.10993	0.823	0.41027
age_scaled:Mandarin	−0.15060	0.08565	−1.758	0.07870
age_scaled:German	0.02744	0.11185	0.245	0.80620
age_scaled:Hindi	−0.10057	0.07929	−1.268	0.20467
age_scaled:Japanese	−0.25863	0.08670	−2.983	0.00285 **
age_scaled:Scottish	−0.03883	0.10281	−0.378	0.70564

J. Acoust. Soc. Am. **150** (6), December 2021

Bent *et al.*   4115

3. Children's data in noise only:

Accuracy ∼ age_scaled * accent + (1 | participant) + (1 | list_sent_word)

Fixed effects:

	Estimate	Std.Error	z value	p-value
(Intercept)	2.41507	0.09549	25.291	< 0.001 ***
age_scaled	0.38881	0.08439	4.607	< 0.001 ***
British	−1.16870	0.06259	−18.673	< 0.001 ***
Mandarin	−1.46271	0.05798	−25.229	< 0.001 ***
German	−1.19753	0.06380	−18.771	< 0.001 ***
Hindi	−2.80296	0.05871	−47.739	< 0.001 ***
Japanese	−2.42786	0.06074	−39.974	< 0.001 ***
Scottish	−1.66380	0.06277	−26.507	< 0.001 ***
age_scaled:British	0.07822	0.06199	1.262	0.20703
age_scaled:Mandarin	0.08590	0.06370	1.349	0.17749
age_scaled:German	−0.01675	0.06703	−0.250	0.80274
age_scaled:Hindi	0.19031	0.06365	2.990	0.00279 **
age_scaled:Japanese	−0.05924	0.05999	−0.987	0.32345
age_scaled:Scottish	0.07392	0.06578	1.124	0.26109

Significance codes: *** ≤ 0.001; ** ≤ 0.01; * ≤ 0.05.

Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (2009). "Comprehension of familiar and unfamiliar native accents under adverse listening conditions," J. Exp. Psychol. Human Percept. Perform. 35(2), 520–529.

Bartelds, M., Richter, C., Liberman, M., and Wieling, M. (2020). "A new acoustic-based pronunciation distance measure," Fr. Art. Int. 3, 39.

Beijering, K., Gooskens, C., and Heeringa, W. (2008). "Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm," Ling. Netherlands 25(1), 13–24.

Bent, T. (2014). "Children's perception of foreign-accented words," J. Child. Lang. 41(6), 1334–1355.

Bent, T. (2018). "Development of unfamiliar accent comprehension continues through adolescence," J. Child. Lang. 45(6), 1400–1411.

Bent, T., and Atagi, E. (2015). "Children's perception of nonnative-accented sentences in noise and quiet," J. Acoust. Soc. Am. 138(6), 3985–3993.

Bent, T., and Atagi, E. (2017). "Perception of nonnative-accented sentences by 5- to 8-year-olds and adults: The role of phonological processing," Lang. Speech 60(1), 110–122.

Bent, T., Baese-Berk, M., Borrie, S. A., and McKee, M. (2016). "Individual differences in the perception of regional, nonnative, and disordered speech varieties," J. Acoust. Soc. Am. 140(5), 3775–3786.

Bent, T., and Bradlow, A. R. (2003). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. 114(3), 1600–1610.

Bent, T., and Holt, R. F. (2018). "Shhh… I need quiet! Children's understanding of American, British, and Japanese-accented English speakers," Lang. Speech 61(4), 657–673.

Bent, T., Holt, R. F., Miller, K., and Libersky, E. (2019). "Sentence context facilitation for children's and adults' recognition of native-and nonnative-accented speech," J. Speech Lang. Hear. Res. 62(2), 423–433.

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., and Quann, C. A. (2009). "Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words," Psychol. Sci. 20(5), 539–542.

Bosker, H. R. (2021). "Using fuzzy string matching for automated assessment of listener transcripts in speech intelligibility studies," Behav. Res. Methods 53, 1945–1953.

Bradlow, A. R. (n.d.) "SpeechBox," https://speechbox.linguistics.northwestern.edu (Last viewed October 15, 2021).

Creel, S. C. (2018). "Accent detection and social cognition: Evidence of protracted learning," Dev. Sci. 21(2), e12524.

Creel, S. C., Rojo, D. P., and Paullada, A. N. (2016). "Effects of contextual support on preschoolers' accented speech comprehension," J. Exp. Child Psychol. 146, 156–180.

Dossey, E., Clopper, C. G., and Wagner, L. (2020). "The development of sociolinguistic competence across the lifespan: Three domains of regional dialect perception," Lang. Learn Develop. 16(4), 330–350.

Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (1979). "Children's understanding of monosyllabic nouns in quiet and in noise," J. Acoust. Soc. Am. 66(1), 12–21.

Evans, B. G., and Lourido, G. T. (2019). "Effects of language background on the development of sociolinguistic awareness: The perception of accent variation in monolingual and multilingual 5-to 7-Year-Old children," Phonetica 76(2–3), 142–162.

Fallon, M., Trehub, S. E., and Schneider, B. A. (2000). "Children's perception of speech in multitalker babble," J. Acoust. Soc. Am. 108(6), 3023–3029.

Flipsen, P., Jr. (2006). "Measuring the intelligibility of conversational speech in children," Clin. Linguist. Phonet. 20(4), 303–312.

Floccia, C., Butler, J., Girard, F., and Goslin, J. (2009a). "Categorization of regional and foreign accent in 5-to 7-year-old British children," Int. J. Behav. Dev. 33(4), 366–375.

Floccia, C., Butler, J., Goslin, J., and Ellis, L. (2009b). "Regional and foreign accent processing in English: Can listeners adapt?," J. Psycholing. Res. 38(4), 379–412.

Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (2006). "Does a regional accent perturb speech processing?," J. Exp. Psychol. Human Percept. Perform. 32(5), 1276–1293.

Gao, Z. (2019). "Weighing phonetic patterns in non-native english speech," Ph.D. thesis, George Mason University, Fairfax, VA.

Girard, F., Floccia, C., and Goslin, J. (2008). "Perception and awareness of accents in young children," Br. J. Dev. Psychol. 26, 409–433.

Gooskens, C., and Heeringa, W. (2004). "Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data," Lang. Var. Change 16(3), 189–207.

Goslin, J., Duffy, H., and Floccia, C. (2012). "An ERP investigation of regional and foreign accent processing," Brain Lang. 122(2), 92–102.

Holt, R. F., and Bent, T. (2017). "Children's use of semantic context in perception of foreign-accented speech," J. Speech Lang. Hear. Res. 60, 223–230.

Houston, D. M., and Jusczyk, P. W. (2000). "The role of talker-specific information in word segmentation by infants," J. Exp. Psychol. Human Percept. Perform. 26(5), 1570–1582.

Howes, D. (1957). "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. 29(2), 296–305.

Huckvale, M. (2004). "ACCDIST: A metric for comparing speakers' accents," in Proceedings of the International Conference on Spoken Language Processing, October 4–8, Jeju, Korea, pp. 1669–1672.

Jones, Z., Yan, Q. Y., Wagner, L., and Clopper, C. G. (2017). "The development of dialect classification across the lifespan," J. Phon. 60, 20–37.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. 61(5), 1337–1351.

Kinzler, K. D., and DeJesus, J. M. (2013a). "Children's sociolinguistic evaluations of nice foreigners and mean Americans," Dev. Psychol. 49(4), 655–664.

Kinzler, K. D., and DeJesus, J. M. (2013b). "Northern = Smart and Southern = Nice: The development of accent attitudes in the United States," Q. J. Exp. Psychol. 66(6), 1146–1158.

Kinzler, K. D., Shutts, K., DeJesus, J., and Spelke, E. S. (2009). "Accent trumps race in guiding children's social preferences," Soc. Cogn. 27(4), 623–634.

Kitamura, C., Panneton, R., and Best, C. T. (2013). "The development of language constancy: Attention to native versus nonnative accents," Child Dev. 84(5), 1686–1700.

Krueger, B. I., Storkel, H. L., and Minai, U. (2018). "The influence of misarticulations on children's word identification and processing," J. Speech Lang. Hear. Res. 61(4), 820–836.

Levenshtein, V. I. (1966). "Binary codes capable of correcting deletions, insertions, and reversals," Sov. Phys. Dokl. 10, 707–710.

Levy, H., Konieczny, L., and Hanulíková, A. (2019). "Processing of unfamiliar accents in monolingual and bilingual children: Effects of type and amount of accent experience," J. Child Lang. 46(2), 368–392.

McCullough, E. A., Clopper, C. G., and Wagner, L. (2019a). "Regional dialect perception across the lifespan: Identification and discrimination," Lang. Speech 62(1), 115–136.

McCullough, E. A., Clopper, C. G., and Wagner, L. (**2019b**). "The development of regional dialect locality judgments and language attitudes across the life span," Child Dev **90**(4), 1080–1096.

McDonald, M., Gross, M., Buac, M., Batko, M., and Kaushanskaya, M. (**2018**). "Processing and comprehension of accented speech by monolingual and bilingual children," Lang. Learn Develop. **14**(2), 113–129.

McQueen, J. M., Tyler, M. D., and Cutler, A. (**2012**). "Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds," Lang. Learn. Develop. **8**(4), 317–339.

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., and Irwin, J. R. (**2013**). "Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent," Child Dev. **84**(6), 2064–2078.

Nathan, L., Wells, B., and Donlan, C. (**1998**). "Children's comprehension of unfamiliar regional accents: A preliminary investigation," J. Child Lang. **25**(2), 343–365.

Newman, R. S., Morini, G., Kozlovsky, P., and Panza, S. (**2018**). "Foreign accent and toddlers' word learning: The effect of phonological contrast," Lang. Learn. Develop. **14**(2), 97–112.

Nilsson, M., Soli, S. D., and Gelnett, D. J. (**1996**). *Development of the Hearing in Noise Test for Children (HINT-C)* (House Ear Institute, Los Angeles, CA).

Paquette-Smith, M., Buckler, H., White, K., Choi, J., and Johnson, E. (**2019**). "The effect of accent exposure on children's sociolinguistic evaluation of peers," Dev. Psychol. **55**, 809–822.

Paquette-Smith, M., Cooper, A., and Johnson, E. (**2020**). "Targeted adaptation in infants following live exposure to an accented talker," J. Child Lang. **48**, 325–349.

Pettersson, E., Megyesi, B., and Nivre, J. (**2013**). "Normalisation of historical text using context-sensitive weighted Levenshtein distance and compound splitting," in *Proceedings of the 19th Nordic Conference of Computational Linguistics (Nodalida 2013)*, May 22–24, Oslo, Norway, pp. 163–179.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K., and Wingfield, A. (**2016**). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)," Ear Hear. **37**, 5S–27S.

Pinet, M., Iverson, P., and Huckvale, M. (**2011**). "Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity," J. Acoust. Soc. Am. **130**(3), 1653–1662.

Potter, C. E., and Saffran, J. R. (**2017**). "Exposure to multiple accents supports infants' understanding of novel accents," Cognition **166**, 67–72.

Psychology Software Tools (**2007**). "E-Prime" (Version 2.0).

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström,, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller,, M. K., and Rudner, M. (**2013**). "The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances," Front. Syst. Neurosci. **7**, 31.

Singh, L., Morgan, J. L., and White, K. S. (**2004**). "Preference and processing: The role of speech affect in early spoken word recognition," J. Mem. Lang. **51**(2), 173–189.

Stringer, L., and Iverson, P. (**2019**). "Accent intelligibility differences in noise across native and nonnative accents: Effects of talker–listener pairing at acoustic–phonetic and lexical levels," J. Speech Lang. Hear. Res. **62**(7), 2213–2226.

van der Feest, S. V., and Johnson, E. K. (**2016**). "Input-driven differences in toddlers' perception of a disappearing phonological contrast," Lang. Acquist. **23**(2), 89–111.

Van Engen, K. J., Phelps, J. E. B., Smiljanic, R., and Chandrasekaran, B. (**2014**). "Enhancing speech intelligibility: Interactions among context, modality, speech style, and masker," J. Speech Lang. Hear. Res. **57**(5), 1908–1918.

van Heugten, M., and Johnson, E. K. (**2012**). "Infants exposed to fluent natural speech succeed at cross-gender word recognition," J. Speech Lang. Hear. Res. **55**(2), 554–560.

van Heugten, M., and Johnson, E. K. (**2014**). "Learning to contend with accents in infancy: Benefits of brief speaker exposure," J. Exp. Psychol. Gen. **143**(1), 340–350.

van Heugten, M., and Johnson, E. K. (**2016**). "Toddlers' word recognition in an unfamiliar regional accent: The role of local sentence context and prior accent exposure," Lang. Speech **59**(3), 353–363.

van Heugten, M., Krieger, D. R., and Johnson, E. K. (**2015**). "The developmental trajectory of toddlers' comprehension of unfamiliar regional accents," Lang. Learn Develop. **11**(1), 41–65.

van Heugten, M., Paquette-Smith, M., Krieger, D. R., and Johnson, E. K. (**2018**). "Infants' recognition of foreign-accented words: Flexible yet precise signal-to-word mapping strategies," J. Mem. Lang. **100**, 51–60.

Vaughn, C., Baese-Berk, M., and Idemaru, K. (**2019**). "Re-examining phonetic variability in native and non-native speech," Phonetica **76**(5), 327–358.

Wagner, L., Clopper, C. G., and Pate, J. K. (**2014**). "Children's perception of dialect variation," J. Child Lang. **41**(5), 1062–1084.

Weatherhead, D., Friedman, O., and White, K. S. (**2018**). "Accent, language, and race: 4–6-year-old children's inferences differ by speaker cue," Child Dev. **89**(5), 1613–1624.

Weatherhead, D., Friedman, O., and White, K. S. (**2019**). "Preschoolers are sensitive to accent distance," J. Child Lang. **46**(6), 1058–1072.

Weatherhead, D., and White, K. S. (**2016**). "He says potato, she says potahto: Young infants track talker-specific accents," Lang. Learn Develop. **12**(1), 92–103.

White, K. S., and Aslin, R. N. (**2011**). "Adaptation to novel accents by toddlers," Dev. Sci. **14**(2), 372–384.

White, K. S., and Morgan, J. L. (**2008**). "Sub-segmental detail in early lexical representations," J. Mem. Lang. **59**(1), 114–132.

Wieling, M., Bloem, J., Mignella, K., Timmermeister, M., and Nerbonne, J. (**2014a**). "Measuring foreign accent strength in English," Lang. Dyn. Change **4**(2), 253–269.

Wieling, M., Nerbonne, J., Bloem, J., Gooskens, C., Heeringa, W., and Baayen, R. H. (**2014b**). "A cognitively grounded measure of pronunciation distance," PLoS One **9**(1), e75734.

J. Acoust. Soc. Am. **150** (6), December 2021

Bent *et al.*   4117