Finding your (child's) voice: Caregiver identification of familiar child voices

Emily Shroads (emily.shroads@mail.utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON L5L 1C6, Canada

Madeleine Yu (madeleine.yu@mail.utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON L5L 1C6, Canada

Elizabeth K. Johnson (elizabeth.johnson@utoronto.ca)
Department of Psychology, University of Toronto
Mississauga, ON L5L 1C6, Canada

Abstract

Previous research has shown that voices of unfamiliar young children are more difficult to differentiate and identify than the voices of adults. In the present study, we examine whether difficulty identifying child voices extends to cases in which those voices are highly familiar. Caregivers (n = 132) of 3.5- to 10-year-old children were presented with voice recordings of their own child amongst gender- and age-matched peers and asked to identify which voice belonged to their child. Although overall accuracy was high, voices of younger children were misidentified more often than voices of older children. In contrast with existing models of familiar voice identification, results suggest that listeners are sensitive to variability in low-level acoustic cues to speaker identity in familiar as well as unfamiliar voice processing.

Keywords: familiar voice recognition; speaker identity

Introduction

On a busy playground, a child calls out "Mom!" Multiple caregivers turn their heads, each believing the voice to belong to their child. All but one of them are wrong.

There are reasons to find this phenomenon surprising. There are clear adaptive advantages to quickly and accurately detecting and identifying the voice of one's own child, and caregivers have a high degree of familiarity with their child's voice. Such exposure should allow caregivers to develop well-specified representations of their child's voice in memory; identification, requiring the comparison of a stimulus to a stored representation, and a mapping to associated identity information, should be easy. Yet compared to faces, voice information is deprioritized in processing of identity (Stevenage, Neil, & Hamlin, 2014), and identification by voice alone is consistently shown to be more difficult than with the availability of visual cues (Barsics & Brédart, 2012).

Although research in voice identification has focused primarily on adult voices, some work has begun to explore identification of children's voices as well. Creel and Jimenez (2012) compared adults' ability to learn and identify adult-adult and child-child voice pairs. Participants were trained on videos of cartoon characters speaking short passages, then tested by hearing a voice and identifying which cartoon character it belonged to. Identification of the child voices was above chance, but poorer than for adult voices, suggesting that child voices may be more difficult to differentiate and learn to identify. However, only one child voice pair (5-year-old females) was used in the study, limiting the generalizability of this finding. In a recent study, Cooper, Fecher and Johnson (2020) further explored the challenges of identifying child voices relative to adults, utilizing a greater number of voices per age group (20 children and 20 adults). Adults were tested in a voice discrimination task, in which they were presented with pairs of words and asked to decide whether they were spoken by the same or different voices. The task was completed for both 2.5-year-old children's voices and adults'. Differentiation between adult voices was found to be much easier than between child voices. Moreover, in a voice training and identification task, voice-identity associations were learned far more quickly and accurately for adult voices compared to child voices.

To understand why it might be difficult to identify children's voices, it is useful to consider the cues listeners have been shown to use in voice recognition and identification more broadly. Work in this area has focused on the specific characteristics used by listeners to distinguish talkers and learn identity associations with previously-unfamiliar adult voices. In such contexts, listeners have been shown to utilize a variety of acoustic cues including fundamental frequency or pitch, speech rate, and nasality, among others (Murry & Singh, 1980). Listeners' access to phonological cues in speech is also important in voice identification, as shown in studies demonstrating voice discrimination advantages when speech is presented in listeners' native language, in which they have full phonological mastery, relative to an unfamiliar or nonnative language (Goggin, et al., 1991; Johnson, Bruggeman, and Cutler, 2018).

Notably, both the acoustic and phonological cues that listeners use to distinguish and identify adult voices may be more unreliable for child voices, providing a potential explanation for the observed greater difficulty of child voice identification. Compared to adult voices, children's productions feature greater within-speaker variability on a variety of acoustic measures (Gerosa et al., 2006; Lee,

Potamianos & Narayanan, 1999), yielding less consistent cues for establishing speaker identity. Additionally, phonological mismatch and lower intelligibility in young children's speech may limit the application and reliability of phonological cues to speaker identity.

However, several major considerations in characterizing the difficulty of the "playground problem" and other everyday scenarios for child voice identification remain unexplored. First is the impact of the age of the child. In their demonstrations of the difficulty of identifying children's voices, both Creel and Jimenez (2012) and Cooper et al. (2020) consider only the voices of young children. It is possible that child voices become gradually more distinguishable until productions become fully adult-like, but it is also possible that this difficulty is confined to the voices of relatively young children, and not experienced with child voices more broadly. A second consideration is whether research findings in the identification of unfamiliar child voices are applicable to situations involving familiar voices. Both Creel and Jimenez (2012) and Cooper et al. (2020) required participants to identify trained, unfamiliar voices, yet comparison of work in identification of unfamiliar and familiar adult voices suggests that these are distinct processes: while unfamiliar voice discrimination appears to rely on auditory pattern-matching, familiar voice identification seems to utilize higher level comparisons to gestalt-like stored representations (Stevenage, 2018). Thus, greater acoustic and phonological variability in children's voices may pose far less of a problem for identification when voices are familiar. However, to date, adult identification of familiar child voices remains largely unexplored. Bartholomeus (1973) tested teachers on identification of their classes of 4- and 5-year-old students by voice alone and through face-voice matching, with reasonable accuracy observed; however, only 4 teachers were included.

Here, we present a large-scale study on adult caregivers' identification of familiar child voices. Participants are tested in a *compound decision task* (see Duncan, 2006), which combines elements of both typical signal detection tasks and forced-choice identification tasks. Like the playground problem, this task involves identifying voices in the context of uncertainty as to whether or not the target voice is present. Moreover, to examine the impact of child age on voice identification performance, voice samples from a wide age range of children (3.5 – 10 years old) are used. If familiar voice identification, unlike with unfamiliar voices, is unimpeded by variability in acoustic and phonological cues to speaker identity, caregivers should readily identify their child's voice when they hear it, regardless of their child's age. However, recognizing that they have *not* heard their child's voice when it is not present may prove more difficult, and in these contexts, age of the child may play a more important role. Specifically, voices of younger children, with their more variable voice characteristics, may be more likely to attain a close enough match to representations of their own child's voice to be confused for their child's voice in its absence (false alarm), despite being easily rejected in its presence.

Methods

Caregivers and their children were recruited from a database of families local to the Greater Toronto area to complete child voice recordings and subsequently participate in the voice identification task. To be eligible, children were required to be between 3.5 and 10 years old, have normal hearing, and not be receiving speech therapy; caregivers were required to report that when interacting with their child, their child spoke to them in English at least 80% of the time.

Materials

One hundred and five children (mean age = 2517 days, approx. 6.89 years; 55 female) were recorded producing 20 isolated single-syllable, CVC words and 4 sentences. For each item, children saw an image of the item and its written form, and heard an on-screen character, voiced by an adult native speaker of North American English, produce the item. Children were prompted by the experimenter to repeat what the character said using their "normal indoor voice." Items were recorded in a fixed order of 4 blocks, consisting of six words followed by one sentence. To encourage children to engage with the task, saying each item was rewarded by revealing a new "sticker" in an on-screen sticker book. To further encourage children's comfort with the task, the first 6 items were semantically-related and well-known across the age range (animals: *dog*, *cat*, *duck*, etc.) Recordings took place during video call sessions monitored by the experimenter for audio quality and any distinctive or identifying background noise (e.g. voice of another household member or pet). Prior to the recording session, families received extensive setup instructions to help optimize recording quality and minimize background noise. Experimenters were instructed to listen carefully during children's productions for any concurrent background noise or disruptions in call audio quality, and in these cases to ask the child to repeat the item again until a higher-quality token was obtained.

Children's productions of each item were segmented from the recordings, with leading and trailing silence removed, and were normalized for total RMS amplitude. Any items that (1) contained background noise that was distinctive in nature (2) had evident spectrotemporal distortions due to bandwidth limitations, or (3) were produced in nonstandard vocal styles (whispering, shouting) were discarded. Due to monitoring for such issues at the time of recording, the rate of item exclusion was low (< 3%).

Child recordings were divided into groups of four voices each, such that the resulting voice sets (1) consisted of children of a single gender, (2) had no two children who were greater than six months apart in age at the time of recording, and (3) were matched for minor audio quality issues such that no child's recordings were distinctive from other group members' for reasons of audio quality alone. A child's voice could be included in up to two voice sets. Only caregivers of children whose voices could be matched into a valid set were invited to participate in the voice identification task.

Voice Identification Task

Participants Adult caregivers of 91 previously-recorded children were invited to take part in the voice identification study. When two caregivers living in the household with the child met eligibility criteria for the study, they were both invited to participate. In total, 138 caregivers from 84 families completed the study. Eighty-five caregivers were female (83 identified as parents, 2 as grandparents), 49 were male (47 identified as parents, 2 not specified), and 4 caregivers did not report their gender (3 parents, 1 not specified). The 84 target children of adult participants were on average 2599 (*SD* = 731.80) days old at the time of recording, or approximately 7.12 years (*SD* ≈ 2.00, range: 3.56 to 9.95); 41 were female. Caregivers were invited to participate an average of 48 days following the recording of their child's voice.

Data were excluded from participants whose personalized task versions contained a configuration error ($n = 5$) or whose responses were missing for some trials ($n = 1$), yielding a final sample of 132 participants.

Procedure The voice identification task was administered online via the Gorilla experiment platform (Anwyl-Irvine et al., 2020); participants were sent a link and password to the task version configured to present a matched voice set containing their child's voice.

Participants were instructed that on each trial, they would listen to three voices, and be asked to identify which of the three voices matched a specified characteristic, or whether none of them matched the characteristic (see Fig. 1).

Prior to the main task, participants completed practice trials to ensure they understood the task. Participants were given the characteristic to listen for (e.g. *which voice is a dog*?) and when ready, listened to three voice clips (dog bark, baby laugh, birdsong). As each clip played in order, a corresponding numbered button in an array of three buttons loomed on-screen. After the last clip finished, if participants believed a voice with the characteristic was present, they clicked the button for the voice they believed was correct. If they did not believe a voice with the characteristic was present, they instead selected a 'none of them' option. Participants were not allowed to replay any audio stimuli before selecting their response. Two practice trials were target-present (TP) trials, in which the specified characteristic belonged to one of the voices in the array, and one practice trial was a target-absent (TA) trial, in which none of the voices matched the characteristic. To ensure participants understood how to respond in the context of both TP and TA trials, after each practice trial participants were given feedback as to whether their response was correct or incorrect; if incorrect, the same trial would repeat until a correct response was made.

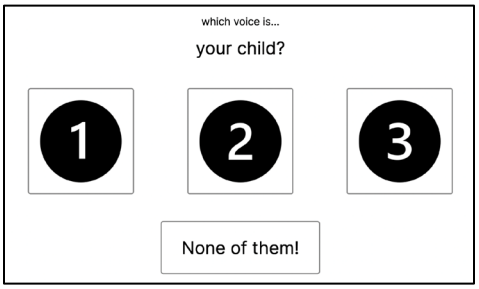

Figure 1: Sample response screen from main task.

In the main task, participants were asked to identify which voice belonged to their child, or to indicate if they believed their child's voice was not present. Trials were presented in the same way as in practice, but with no feedback or repetition. During each trial, participants heard 3 voice options out of the matched set of 4 (consisting of their child's voice and three other voices) and were asked to identify which voice was their child, or whether their child's voice was not present.

On each of 8 trials, each voice option was heard producing two isolated words selected at random from the recorded set, with 500 milliseconds of silence between each word. To facilitate direct comparison of voices, each voice option was heard producing the same words within each trial. No word item was repeated across trials.

Voices were played an equal number of times across trials, yielding 75% of trials (6) in which the target child's voice was present (TP trials), and 25% of trials (2) in which the target child's voice was not present (TA trials). Participants were not given any indication of how often to expect the target voice to be present or absent, and were simply told that sometimes their child's voice may not be present. Because voices were presented equally, there were no available cues to voice identity in their presentation frequency. That is, if participants were able to distinguish voices from each other but not identify which was their child's, they would not be able to surmise its identity by assuming it to be the most frequently-presented voice.

Participants later completed an additional set of trials presenting sentence recordings rather than words; however, data from sentence trials are not reported here. Only responses from the first portion of the task involving word stimuli are considered in the present study.

Analysis

Responses on each trial were coded according to which type of correct response or error they represented. On TP trials, one correct and two incorrect response types are possible. Participants can correctly select the target voice from the array (*correct identification* or CID), or they can incorrectly select a foil voice from the array (*false identification* or FID), or incorrectly respond that they believe none of the voice options are their child's voice (*miss* or MS). On TA trials, there are only two possible response types. Participants can correctly respond that their child's voice is not present (*correct rejection* or CR), or incorrectly choose a foil voice from the TA array (*false alarm* or FA). Note that while the CID response type is analogous to a *hit* in a signal detection theory framework, and CR and FA are shared, FID is a response type unique to the identification component of this

Table 1: Response type frequencies.

Trial type	Response	*n*	%
Target Present	CID	471	59.5
	MS	239	30.2
	FID	82	10.4
Target Absent	CR	199	75.4
	FA	65	24.6

compound decision task. Trial response type frequencies are reported in Table 1.

Although prior work involving compound decision tasks has largely utilized response type frequencies to derive and model estimates of *d'* and response bias as in a signal detection task (Duncan, 2006; Lee & Penrod, 2019), a mixed-effects modeling approach was preferred in this case to better account for anticipated individual differences in participants' voice identification abilities (see Shilowich & Biederman, 2016) as well as potential voice-level influences on identification performance (i.e. voice distinctiveness). This approach has gained prevalence in some other types of combined detection-decision tasks (e.g. Gokool et al., 2022).

To better account for voice factors in the model, observed response types were represented at the level of individual voices. Participants ultimately provided a single response to each trial, yet their selections can also be conceptualized as a set of judgments for each of the three voices heard. Within the constraints of the task, where participants were aware their child's voice could be 0 or 1 of the options in the array, choosing a voice from the array indicates that the participant believed that voice to be their child's, and believed the other two voices in the array to *not* be their child's. Therefore, a CID represents correct responses to all 3 voices (selection of target, rejection of two foils), MS represents an incorrect response to the target but a correct response (rejection) to both foils, FID represents an incorrect response to the target and the incorrectly-chosen foil but a correct response to the unselected foil, FA represents an incorrect response to the chosen foil but correct responses to the other two unchosen foils, and CR represents correct responses to all three unchosen foils.

Accuracy of voice judgments was predicted using a mixed-effects logistic regression model, constructed using the lme4 package in R (Bates et al., 2015). Fixed effects were children's Age (centered at the mean and scaled) and Trial Type (contrast-coded as TP = 1, TA = -1); an interaction term was also included. Random intercepts of Participant and Child Identity were entered one by one and selected for inclusion via model comparison. Additional random intercepts of Trial Number (to account for potential learning effects across trials) and Array Position (to account for primacy/recency effects of whether the voice being judged was presented first, second, or third in the response array) were considered, but dropped during model selection for failure to improve model fit (Trial Number: $\chi^2(1) = 0.41$, $p = 0.52$; Array Position: $\chi^2(1) = 0.24$, $p = 0.63$). The final model is summarized in Table 2.

Table 2: Log odds of making correct judgment of voice identity as own child or not own child.

Predictor	Estimate	*SE*	*z*	*p*
Intercept	2.73	0.18	15.02	< 0.0001
Age	0.40	0.18	2.26	0.02
TrialType	-0.78	0.15	-5.32	< 0.0001
Age*TrialType	-0.28	0.14	-2.02	0.04

```
glmer(Correct ~ Age*TrialType +
(1|Participant) + (1|Child), family =
"binomial", data = data)
```

Participants' judgments were generally accurate, with the model intercept corresponding to 93.9% correct. Of note, there is no true chance level in compound decision tasks, and task constraints, where the target child's voice could belong to only 0 or 1 voices in a trial, result in trial-wise error types being represented as errors on only one voice (miss and FA) or two voices (FID) of three. As a result, floor performance for the task (FID on all of 6 TP trials, FA on both 2 TA trials) is 41.7%.

The model revealed a main effect of Age, where older children's voices were more likely to be judged correctly ($\beta = 0.40$, $SE = 0.18$, $z = 2.26$, $p = 0.02$). This aligns with work in unfamiliar voice identification and provides further support for voices of younger children being more difficult to identify than more adult-like voices (Cooper et al., 2020; Creel & Jimenez, 2012).

Additionally, Trial Type was found to have a significant effect on voice judgment accuracy, yet in the opposite direction of the expected pattern ($\beta = -0.78$, $SE = 0.15$, $z = -5.32$, $p < 0.0001$); judgments on TP trials were less accurate than TA trials. This difference, and what it suggests about how listeners approached the task, is discussed further in the following section.

The model also revealed a significant interaction between Age and Trial Type ($\beta = -0.28$, $SE = 0.14$, $z = -2.02$, $p = 0.04$). This interaction is visualized in Fig. 2. Within TP trials, incorrect judgments were spread relatively evenly across the full range of child ages, yet in TA trials, voices of younger children were judged incorrectly more often than the voices of older children.

Discussion

Listeners find it incredibly difficult to distinguish and identify recently-learned children's voices (Cooper et al., 2020; Creel & Jimenez, 2012). Yet the most critical and most common scenarios for adults to accurately identify child voices do not center around unfamiliar voices, but the highly-familiar voices of the children for whom they are caregivers. In the present study, we explored whether difficulty identifying children's voices extends to familiar voices by examining caregivers' ability to identify their own children by voice alone.

Children's age was found to have an effect on participants' ability to correctly judge whether voices did or did not belong to their child, where younger children's voices were more

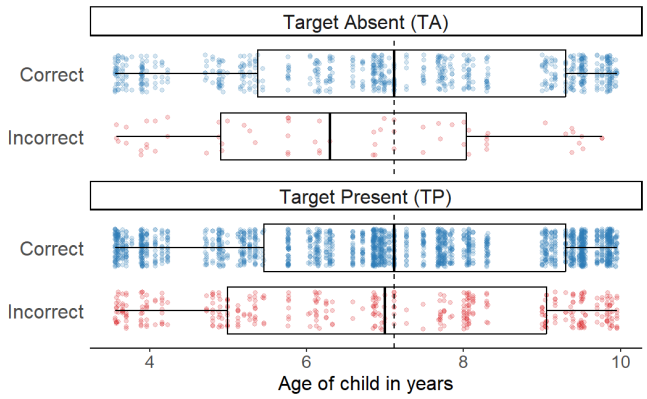

Figure 2: Voice judgments by trial type. Points represent individual voice judgments; dashed line represents the median age of all voices presented.

likely to be judged incorrectly than older children's. One explanation, as has been suggested previously in unfamiliar voice identification, is that greater variability in young children's speech, from one production to the next, may result in less reliable cues for voice identification (Cooper et al., 2020). Here, the demonstration of greater difficulty identifying young children's voices suggests that variability may also be an issue for familiar child voices, but that this poses less of an issue for voices of older children as they become more consistent and adult-like in their productions.

In the context of familiar voice identification, a second possible explanation for the effect of age is the relative instability of children's voice characteristics over time. Acoustic properties of children's speech change over development in concert with physical growth in the vocal tract and improvements in motor control. Developmental trajectories of many acoustic measures are notably nonlinear, with periods of rapid change in early childhood (see Vorperian & Kent, 2007 for review). In a few months' span of time, a nine-year-old's voice may change very little, yet a three-year-old's may differ much more substantially from their younger self's. If familiar voice identification typically becomes easier with greater degrees of exposure and voice representations become more highly specified, young children's voices are a moving target, rendering extended high-level exposure less useful or even potentially misleading, if voice representations do not shift in step with speech development. These two explanations are not mutually exclusive; moment-to-moment variability and change over time are deeply intertwined in the context of children's motor control and vocal development. Yet because the latter is uniquely an issue in the context of listening to familiar voices, future work directly comparing the processing of familiar and unfamiliar child voices has the potential to elucidate the relative challenges they pose to listeners.

However, both of these potential explanations are at odds with findings in familiar voice identification. Taken together, work in this area suggests gestalt-like representations for familiar voices, with less reliance on specific acoustic features, that are acquired rapidly through processes similar to one-trial learning (see Stevenage, 2018 for review). This work has focused almost entirely on adult voices, therefore further examination of familiar child voices is necessary to reconcile these differences and to enhance our understanding of the mechanisms of familiar voice identification in general. It is worth noting that despite high levels of accuracy, many of our participants reported that they found this task very difficult. Findings in other types of speech processing tasks, such as listening-in-noise tasks, suggest that increased difficulty in the speech signal can prompt listeners (counter-productively and somewhat paradoxically) to shift listening strategies, increasing their reliance on low-level acoustic information and bottom-up processing and decreasing their use of top-down processing (Mattys, Brooks & Cooke, 2009). If listeners respond similarly when encountering the greater difficulty inherent to processing young children's voices, with their greater variability and lower intelligibility, this could help account for differences in the patterns seen here and the predictions of existing models of adult familiar voice processing.

Participants' apparent reliance on low-level acoustic cues in identifying their child's voice may also account for another unexpected result in the context of the familiar voice identification literature: greater accuracy on target-absent trials compared to target-present trials. Possible sources of greater difficulty in TP trials are made clearer in looking at the particular types of errors observed in both target-present and target-absent trials, and their relative prevalence. Misses, in which the target child's voice was a response option but the participant believed their child's voice was not present, were the most common error type, accounting for most of participants' difficulty with target-present trials. Meanwhile, false identifications (choosing a foil child's voice over the target child's on a target-present trial) were rare and false alarms (choosing a foil child's voice when the target child's voice was not present on the trial) were in between. Both false identifications and false alarms involve selecting the wrong child's voice as one's own; their infrequency suggests that listeners do not have exceptional difficulty distinguishing between their child's voice and other unfamiliar voices, and appropriately rejecting the unfamiliar voices. Yet the comparatively higher rate of misses suggests that in this task, obtaining a strong enough match between a voice sample and a memory representation to *detect* a familiar voice is more problematic than any issue of mistaken identity. It may be the case that caregivers' voice representations are over-specified, and fail to accommodate all of the variability in children's productions; alternatively, this type of error may be especially sensitive to the type of audio stimuli used (utterance length, spontaneous vs. non-spontaneous speech recordings, audio quality) and participants' assumptions about the task. Simple detection tasks may be a useful method to narrow in on the dynamics surrounding these errors, over conventional identification-focused tasks such as voice lineups. In our opening playground scenario, caregivers

monitor a complex auditory scene without consistent access to visual cues as to whether their child is currently speaking. They must decide moment-to-moment if an incoming speech signal is personally relevant and whether to respond. In such contexts, our results suggest that it may be easier to momentarily miss the voice of one's child than to mistakenly respond to an unfamiliar child calling for someone else.

False alarms, however, are the focal point of an interaction between task trial types and age: caregivers were more likely to make false alarm errors with younger children's voices. This was an expected pattern, as target-absent trials were thought to represent the scenario in which participants would most need to rely on auditory pattern-matching to succeed, and yet with younger children, would also have the least consistent acoustic patterns available to track and recognize. Of greater interest is the main effect of age, which suggests that listeners rely on low-level acoustic cues outside this context as well.

A limitation of the present study is the necessity of conducting child voice recordings online, over Zoom. Due to bandwidth limitations, voice recordings were consistently lower-quality than could be obtained in a lab setting. Variation in equipment and home recording environments also resulted in some recordings having better audio quality than others; while voices within each participant's lineup were matched for audio quality, it is possible, though unlikely, that participants could utilize additional quality-related differences in the audio to group and differentiate voices rather than relying on vocal features alone. While bandwidth-limited (telephone) speech recordings have been shown to reduce identification accuracy of once-heard voices (McDougall et al., 2015) and enhance perceived voice similarity (Nolan et al., 2013), it is unlikely that highly-familiar voice recognition would be substantially affected by the audio quality typical of telephone or videocall transmission. Moreover, we have no reason to expect any systematic differences in home recording quality with respect to child age, meaning that our results could not be due to a confound in stimulus quality.

In real-world scenarios, voice identification rarely must be based on isolated, single-word utterances as in the present study. With their greater length, samples of connected speech provide more information upon which to base judgments of voice identity, but they also provide additional types of suprasegmental cues, such as sentence-level prosody. While outside the scope of the current study, future studies should examine the extent to which age-related difficulties in identification of children's voices can be overcome with the additional, and potentially more reliable, cues to voice identity that are available in connected speech.

As the first large-scale study to our knowledge that has examined familiar child voice identification, we demonstrate that even high degrees of familiarity do not prevent adults from experiencing some difficulty in recognizing voices of children, particularly the acoustically variable voices of young children. Child voices present unique difficulties to voice processing, and have unique potential to expand our understanding of the mechanisms underlying familiar voice recognition. Further work can shed light on the extent to which processes in unfamiliar and familiar voice identification are shared or distinct, and how voice representations form and shift to accommodate variability and change over time in speech development.

Acknowledgements

We would like to thank Melynna Duong, Maripaz Gonzalez, Danyel Ore, and Lisa Hotson, as well as the other members of the Child Language and Speech Studies Lab for their support. This work was supported by grants from the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Barsics, C., & Brédart, S. (2012). Recalling semantic information about newly learned faces and voices. *Memory*, *20*(5), 527–534. https://doi.org/10.1080/09658211.2012.683012

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *27*, 464–472. https://doi.org/10.1037/h0082498

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Cooper, A., Fecher, N., & Johnson, E. K. (2020). Identifying children's voices. *The Journal of the Acoustical Society of America*, *148*(1), 324–333. https://doi.org/10.1121/10.0001576

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*(4), 487–509. https://doi.org/10.1016/j.jecp.2012.07.007

Gerosa, M., Lee, S., Giuliani, D., & Narayanan, S. (2006). Analyzing Children's Speech: An Acoustic Study of Consonants and Consonant-Vowel Transition. *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, *1*, I–I. https://doi.org/10.1109/ICASSP.2006.1660040

Gokool, V. A., Crespo-Cajigas, J., Mallikarjun, A., Collins, A., Kane, S. A., Plymouth, V., Nguyen, E., Abella, B. S., Holness, H. K., Furton, K. G., Johnson, A. T. C., & Otto, C. M. (2022). The Use of Biological Sensors and Instrumental Analysis to Discriminate COVID-19 Odor Signatures. *Biosensors*, *12*(11), Article 11. https://doi.org/10.3390/bios12111003

Johnson, E. K., Bruggeman, L., & Cutler, A. (2018). Abstraction and the (Misnamed) Language Familiarity

Effect. *Cognitive Science*, *42*(2), 633–645. https://doi.org/10.1111/cogs.12520
Lee, J., & Penrod, S. D. (2019). New signal detection theory-based framework for eyewitness performance in lineups. *Law and Human Behavior*, *43*(5), 436–454. https://doi.org/10.1037/lhb0000343
Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686
Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. https://doi.org/10.1016/j.cogpsych.2009.04.001
McDougall, K., Nolan, F., & Hudson, T. (2015). Telephone Transmission and Earwitnesses: Performance on Voice Parades Controlled for Voice Similarity. *Phonetica*, *72*(4), 257–272. https://doi.org/10.1159/000439385
Nolan, F., McDougall, K., & Hudson, T. (2013). Effects of the telephone on perceived voice similarity: Implications for voice line-ups. *International Journal of Speech, Language and the Law*, *20*(2), 229–246. https://doi.org/10.1558/ijsll.v20i2.229
Shilowich, B. E., & Biederman, I. (2016). An estimate of the prevalence of developmental phonagnosia. *Brain and Language*, *159*, 84–91. https://doi.org/10.1016/j.bandl.2016.05.004
Stevenage, S. V. (2018). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia*, *116*, 162–178.
Stevenage, S. V., Neil, G. J., & Hamlin, I. (2014). When the face fits: Recognition of celebrities from matching and mismatching faces and voices. *Memory*, *22*(3), 284–294. https://doi.org/10.1080/09658211.2013.781654
Vorperian, H. K., & Kent, R. D. (2007). Vowel Acoustic Space Development in Children: A Synthesis of Acoustic and Anatomic Data. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1510–1545. https://doi.org/10.1044/1092-4388(2007/104)