



OPEN Linguistic diversity shapes flexible speech perception in school age children

Ethan Kutlu^{1,2}✉, Keith Baxelbaum², Eldon Sorensen³, Jacob Oleson³ & Bob McMurray^{1,2}

Every day, listeners encounter a wide range of acoustic signals. Successfully solving this variability problem allows them to interpret these signals accurately. While this mechanism tends to be less effortful for adults, children need to learn stable categories in the face of such variability. It is unknown to what extent general maturation or diversity of the input plays a role in shaping different speech categorization profiles that children can employ. Here, we tested school-aged children's speech categorization with a continuous speech categorization task called the Visual Analogue Scaling (VAS) task. We measured the linguistic diversity in each child's social environment through a social network analysis. We found that increased linguistic diversity led to more flexible and gradient speech categorization. On the other hand, less diverse linguistic input led to more categorical speech categorization. We argue that these findings have implications for speech perception as well as linguistic diversity research.

In the course of everyday speech perception, listeners encounter highly varied acoustic signal that changes based on the talker, speaking rate, and neighboring sounds, among other things^{1–5}. Consider, for instance, the words *time* and *dime* in American English. To successfully discriminate between them, listeners rely on continuous acoustic cues such as Voice Onset Time (VOT), the time difference between the release of the articulators, and the onset of voicing. In American English, sounds like /d/ will have a VOT near 5–10 msec, and /t/ will have higher VOTs near 50–60 msec, though both categories span a considerable range. However, these prototypes can shift. For instance, a slower speaking rate can shift VOTs to be longer³ and individual talkers may have their own characteristic VOTs (even when controlled for rate⁶). Thus, the same VOT of 20 msec might be consistent with either /d/ or /t/ depending on factors like the speaking rate, differences among talkers, and the influence of neighboring sounds.

It is often assumed that listeners categorize speech input into meaningful linguistic units treating different ranges of the input as indicative of one category or the other (see⁷ for perspectives without units, but which still assume some form of functional categorization). However, because of the contextual factors described above, it is not enough to simply know the correct range of VOTs for each category (e.g., a /d/ has VOTs ranging from –20 to 15 msec); listeners must flexibly account for different sources of variance.

This problem is even more challenging when we consider development. Speech categories differ across different languages; consequently, children must learn the categories of their language(s) while, at the same time, navigating the complex problem of variability to understand speech. Research suggests that speech categorization may begin developing during early infancy. This is supported by evidence for infants' speech discrimination becoming attuned to the ambient language between 6 and 12 months of age^{8–12}. This was thought to be based on some form of statistical learning^{13–16}, based on statistical properties of speech cues (like VOT). However, in parallel, there is mounting evidence that children continue to develop speech categories through late childhood and even into adolescence¹⁷, a time when there are large changes in general cognition, as well as a host of language- and speech-specific developments with growth in vocabulary, improvements in speech production and the onset of reading. This complicates our understanding of category formation. Critically, it is currently unclear what aspects of speech perception develop during these later years, and how large a role input and maturational development play during this process.

This study focuses on school-aged children's development of speech categorization. This age group was targeted due to the significant developmental changes children experience at these ages (e.g., learning how to read, learning new vocabulary, expanding their social networks through schooling). We asked what specific aspects of speech perception are still developing and whether the linguistic input plays a role over and above

¹Department of Linguistics, University of Iowa, Iowa, USA. ²Department of Psychological and Brain Sciences, University of Iowa, Iowa, USA. ³Department of Biostatistics, University of Iowa, Iowa, USA. ✉email: ethan-kutlu@uiowa.edu

general maturation using novel empirical and statistical tools. In the next section, we start by briefly discussing what is known about the development of speech categorization. We then focus on new approaches in speech categorization and how these new approaches can expand our understanding of development, particularly in linguistically diverse contexts.

Development of speech categorization

It is well known that infants as young as six months begin to become sensitive to common speech contrasts^{8,9,18,19}. However, as they develop, their ability to discriminate speech sounds becomes more refined and tailored to the language that they are exposed to regularly, by about 10–12 months. During this time, the canonical view is that they lose access to many contrasts that are not attested in their language ecology¹⁰, and they increase their ability to discriminate contrasts that are attested⁹. This is a robust developmental progression that has been shown across multiple speech contrasts^{11,20,21}. At these ages, this rapid tuning to the language(s) is thought to derive mainly from infants' sensitivity to statistical properties of speech input, which enables them to align their internal categories to the surrounding language's distributional statistics¹². This statistical alignment leads them to narrow their perception to the demands of their language(s).

During this time, there is also evidence that infants are shaping their speech perception more broadly to the properties of their ambient language. For example, infants attune to the specific dialects they encounter by around 19 months – struggling to recognize words in unfamiliar dialects^{22,23}, though even older toddlers at 28 months old do show above-chance recognition of unfamiliar dialects²⁴. Moreover, infants exposed to two dialects at home seem to attune more slowly to either than those exposed to one²⁵. As a whole, this is consistent with a statistical attunement account where infants are narrowing in on the statistical structures appropriate for their input. However, this work on dialects in toddlers has only been examined in the context of whole word recognition – it is not clear how phoneme categories may adapt. More importantly, it is not clear if children exposed to variable dialects may have access to adaptations that could help support their varied listening needs. Such adaptations may run counter to the predominant trend in the infant literature that stresses the developmental goals of narrowing and focusing on the ambient language.

More recent research has extended work in infancy to suggest that speech categorization may develop well after infancy or even childhood. These studies typically use forced-choice categorization tasks in which listeners hear tokens from a speech continuum (e.g., ranging in small steps between *beach* and *peach*), and identify the word or phoneme in a Forced-Choice task (FC) (e.g., Fig. 1 for representative data). This provides a significantly richer picture of the nature of speech categorization than is possible with infant methods.

Collectively, these studies have shown several trends over the course of development, spanning from age 3 to as late as 18 and beyond. By far, the most prominent finding is that children improve on the precision of these categories, as indicated by the slope of the categorization function^{17,26,27} (e.g., Fig. 1). In these studies, the steepening of the speech categorization function over development (i.e., a steeper slope) is seen as consistent with perceptual narrowing. The assumption was that this steeper slope or sharper (i.e., more accurate and efficient) categorization indicates less sensitivity to fine-grained differences within a category. These findings parallel other studies showing that a shallower slope is also observed in language impairments^{28–31}, illiteracy³², and bi/multilingual exposure^{33–43}.

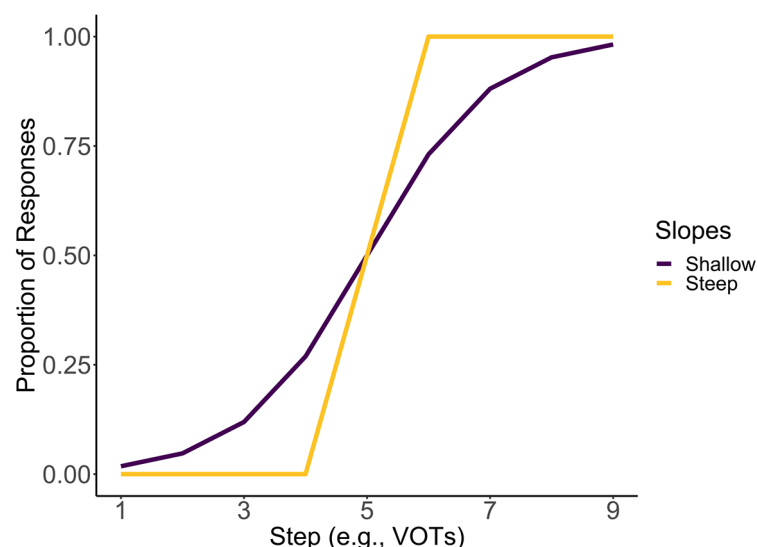


Fig. 1. Schematic results from a typical forced-choice / continuum task used with school age children and adults. The X axis represents a continuum ranging from (for example), /b/ to /p/; the Y axis is the proportion of responses to one of the endpoint (in this case, /p/). Older children typically show a steep slope, categorization patterns and younger children a shallow one.

However, this leaves open three critical questions. First, work with adults has come to question standard interpretations of forced-choice categorization tasks (see, for instance⁴⁴), it is not clear how to interpret this steepening slope, and consequently, it is not clear what aspects of speech categorization are developing. Second, while empirical and theoretical work in adults has documented clear plasticity in speech categorization in the lab^{45–47}, there is no empirical work linking the form of speech categorization in childhood to the statistics of the input in their environment. As a result, it is not clear if this development is experience-dependent (e.g., continued or distinct forms of statistical learning) or if it derives from more general growth in skills or even maturation. Finally, while theoretical models of learning suggest that development can in principle result in diverse outcomes^{14,45,48}, a typical assumption is that most children will arrive at a similar solution to the problem. However, these kinds of assumptions are rarely tested and as a result, it is not clear whether all children reach the same state or whether there may be diverse ways to solve the categorization problem that are tuned to the environmental demands faced by the child.

What is developing? Gradient vs. categorical speech perception

While children appear to show increasingly sharp or categorical speech categorization, mounting evidence from adult listeners suggests that gradient categorization—not a steep slope—is the norm. These studies have employed different methods such as priming⁴⁹, rating scales^{50,51}, eye-tracking^{52–55}, and electroencephalography^{55–57}. Moreover, it is now clear that this gradiency is beneficial for listeners (see⁵⁸ for a review), and can help them retain plasticity and flexibility⁵⁵. For instance, listeners who are more gradient categorizers are more sensitive to secondary cues^{59,60}, gain sensitivity to fine-grained, within-category differences, which can help them more flexibly recover from errors⁵⁴, and become increasingly efficient in accessing these categories in real-time¹⁷. Moreover, recent studies with adults also show that gradient listeners learn contrasts in other languages more accurately^{61,62}.

The increasingly categorical response functions seen over development would appear to run counter to the benefits of gradiency. This leaves us with a puzzling question: if gradient speech categorization is more functional for adaptation, plasticity, and recovery in language processing, what does it mean that the slope steepens with development?

This raises the possibility that the slope of the forced-choice categorization function may not accurately characterize the nature of the changing speech categorization system. That is, there may be multiple dimensions that change with development, and we may need to move to new empirical and conceptual tools to understand the precise nature of this development.

In considering the nature of this development, it helps to start from the standard interpretation of this data (which is still employed by many studies). Forced-choice tasks were originally grounded in Categorical Perception⁴⁴ which argues for “quasi-discrete” categories⁶³ with sharp boundaries. Consequently, previous studies considered a steep slope (a.k.a. crisp categorization) to be the ideal outcome for speech categorization, with a shallow or more graded response function linked to various language impairments^{28–31}, to illiteracy³², and to bi/multilingual exposure^{33–43}. However, these interpretations do not align with the current gradient views of speech perception^{15,44,47,55–58,61,62,64–66} or with the work that has emphasized the importance of variation in development across the lifespan^{67–73} (We would like to emphasize that work from speech sciences, phonetics, and certain areas of psycholinguistics on adults takes a nuanced view of speech categorization as a gradient and plastic process. However, these updated views have not been effectively communicated to other fields such as speech and hearing sciences, bilingualism and in some cases development, nor have they appeared basic psycholinguistic or cognitive psychology textbooks.). These findings also do not align with the more established computational frameworks of speech perception^{45,64} which embrace this gradiency as central.

Critically, we note that the evidence favoring a gradient view does not come from forced-choice tasks – instead, it comes from research using implicit and continuous measures like goodness ratings^{50,51}, priming⁴⁹, eye-tracking in the visual world paradigm⁷⁴, and ERPs⁵⁷. Such tasks can often show a gradient underlying representation even as forced choice categorization shows nearly categorical results. A good example of this disconnect between is a study¹⁷ testing school-aged children in a version of the Visual World Paradigm designed to isolate within-category (gradient) sensitivity to fine-grained acoustic differences⁵². With respect to the overt mouse click responses (which are analogous to a forced-choice task), this study found clear development over the 7–18 year period – with a gradually steepening slope of the identification function, just as had been observed in prior work^{26,27}. However, crucially, an analysis of the children's eye movements suggested that older children were becoming more sensitive to small within-category changes in VOT, not less. That is, developmentally, they were trying to achieve a more gradient representation of speech, even as their forced-choice responses looked more categorical. This challenges the notion that the steepening slope in forced-choice tasks is indicative of narrowing – here, it seems to accompany greater sensitivity to differences.

Part of the issue is that traditional forced-choice categorization tasks have an underlying mathematical ambiguity that makes it impossible to disentangle what exactly a shallow categorization slope means. In a forced-choice task, even if the underlying representation were gradient, children could still look categorical (if they always chose the most likely response at each step along the continuum). Alternatively, if a child showed a gradient response, they could be underlyingly gradient and attempting to match their underlying function, or they could be categorical, but the noise in the system is disrupting this^{44,75}.

A Visual Analog Scaling task (VAS) can tease these factors apart. Like the 2AFC task, in a VAS task, a sound from a speech continuum is heard. However, instead of making a discrete choice (i.e., 0 or 1), listeners respond on a continuous (analog) scale to indicate where the sound falls between endpoints. For instance, if a listener responds to tokens from a *time/dime* continuum, the scale has an image of *time* on the left and a *dime* on the right. Listeners then choose wherever on the continuum between the two items that they think the sound falls.

Critically, the VAS task overcomes the interpretative ambiguity of the 2AFC task as it can also measure trial-by-trial variability in addition to the average curves. Consider a situation where listeners have a categorical (discrete) category system but display noise in their encoding from trial to trial. In this case, the average VAS function should look shallow (i.e., gradient), but in fact, the listener is generally choosing the endpoints from trial to trial, showing high trial-by-trial variability (high categorization inconsistency). In contrast, if a shallower mean slope derives from a true gradient underlying representation, listeners will also show a gradient mean function. However, they should display individual responses that are tightly clustered around the mean (i.e., low categorization inconsistency).

The picture of speech categorization offered by the VAS task suggests that there may be multiple aspects of speech categorization that could be developing: the overall shape of the function (e.g., whether the slope reflects increasingly gradient or categorical response) and the nature of the trial-to-trial variability. The introduction of the VAS task raises a new dimension to consider (i.e., consistency) and may change our view of old ones (e.g., slope) by decoupling them from consistency.

The question then moves away from whether the slope of the mean function changes by getting more or less steep to the primary developmental change in the variability around the mean. Such changes could pattern together. For example, if children are becoming more categorical with age, this should appear as a steepening slope with increased consistency around the mean. On the other hand, it is also possible that children are tuning into the fine-grained differences in the signal¹⁷ – in this case, they would show reductions in slope alongside increased consistency (and it is the increased consistency that makes the 2AFC task *appear* categorical in older children). Thus, one goal of the present study was to use the VAS task to isolate what dimensions are changing as children continue to develop.

Is development experience-dependent and equifinal?

A second critical question is what is driving these changes in school-age children. One possibility is that they largely result from simply better overall cognitive ability – children gain executive function, working memory, and decision skills during this time, which could make them better at this task. However, a second possibility is that they result from ongoing experience-dependent plasticity.

In fact, a wealth of studies over the last two decades suggests that speech perception in adults is highly plastic. Studies on adult perceptual learning and plasticity suggest that in laboratory settings, listeners can adapt to new accents or ways of speaking rapidly and effortlessly^{46,66,76}. It seems highly likely that children can do the same. However, there is much less evidence – in either adults or children – for this kind of learning and adaptation outside of the lab, in the real world. Here, evidence would come from correlational studies linking systematic differences in a child's linguistic environment to differences in speech categorization.

The school experience affords a sort of natural experiment that can be conducted with developing children. When children transition from home or preschool to formal schooling, they may experience a stark change in the input statistics. Many children will simply encounter more talkers (with their own idiosyncratic ways of producing phonemes) as they transition from small-group preschools or homes to a more extensive and more diverse classroom. But for some children changes may be more dramatic. Some may be exposed to a high degree of dialect/accents variability, if their school features many children from non-English speaking homes (in the context of USA). This variability would not have been present to the same degree in their prior environment. Alternatively, they might have variability at home through exposure to different language(s), dialect(s), and accents. Consequently, the experience of schooling could potentially alter their speech categorization patterns.

It is important to note that children's experiences in school may differ even within the same school as a function of whom they interact with. Thus, to capture this, we must precisely characterize the nature of each child's social network. By relating speech perception to the nature of the child's social network (with respect to dialect, accent, and other language use), we can document the effect of the learning environment on speech category development in older children's speech perception. There are many recent studies that document differences in input through network questionnaires in infants and adults^{75,77–87}. The second goal of this study then is to harness this natural variation to ask if continued development is driven by the input. That is, we document children's input variability in their daily lives and ask if this impacts some aspects of speech categorization.

As we described, perceptual learning studies with adults offer substantial evidence for plasticity in general, but it is not clear how to map these learning paradigms onto the likely changes that most children experience during the school-age years. Most of these perceptual learning studies have examined situations where listeners shift existing boundaries to cope with a new type of input (e.g., exposure to a novel accent or a speaker with non-canonical productions). In contrast, by the time children have reached school age, they likely know the boundaries of their first language(s).

Work on individual differences in gradient speech categorization in adults suggests that being more gradient (e.g., a shallower slope and more consistent responses in the VAS task) can help cope with uncertainty. This is supported by two adult laboratory learning studies: when listeners encounter more variability in the speech cues (e.g., the “clusters” of cues that correspond to a category are wider), they adapt by becoming more gradient^{46,66}. However, such learning has not been demonstrated outside of the laboratory, nor in children.

Here, we apply the same logic to the real-world experience of developing children, and with the more sensitive VAS task. This predicts that children exposed to more diverse language backgrounds may adapt by adopting a shallower slope but with more consistency. That is, they are tuning into the fine-grained details of the input because they may need it to sort through the enhanced variability they face.

Critically, much of the work on dialect variability in older ages, seems to support a version of narrowing: as children become more proficient and adapted to the specific dialects they are exposed to, they may be less able to process other types of speech. In contrast, our hypothesis complements this work by suggesting that when

children are exposed to a lot of linguistic (e.g., dialects, accents) variation, they may adapt by simply becoming *more sensitive* to fine-grained differences.

These results highlight the possibility that children's speech categorization may be influenced by individual differences in listening experiences and environmental variability, leading to diverse speech processing outcomes. But this in turn leads to a potentially larger difference. If speech categorization is plastic and sensitive to differences in listening demands, it is possible that all children will not arrive at the same end state (e.g., a sufficiently gradient or categorical representation)— children may develop distinct strategies for processing speech, shaped by their linguistic environment and individual experiences. This is perhaps a given for children who are being raised multilingually, but here we suggest that the well-established plasticity in the speech categorization system means that even monolingually developing children may reach a different state.

The standard assumption in most language acquisition work is for equifinality: children reach (or attempt to reach) the same outcome (the skills they need to process their language(s)). Here, however, it is also possible that we will see the converse: children adapting distinct solutions to their own language needs.

The present study

The goal of this study is to address these three fundamental questions. First, we use the VAS task to more precisely capture what aspects of speech categorization are developing. This task has not yet been used with children and may offer new insight into the nature of this development. In this regard we also present the use of a novel statistical tool (latent profile analysis) that allows us to more precisely make sense of this multi-dimensional space. Second, we ask if the diversity of linguistic input in the child's environment (measured with a social network analysis) impacts the fundamental nature of speech categorization. This implies that the later school-age development is at least partially experience-dependent during this window. Critically, if gradiency functionally supports more flexible speech processing, this predicts that children with diverse inputs may show more gradient responses. Note that in this case, the more gradient pattern of behavior lies in fundamental opposition to work in bilingualism that has argued (on the basis of the 2AFC task) that a shallower slope indicates disrupted (less crisp) speech categorization^{39,42}. Here we predict the opposite: more gradient responding indicates richer sensitivity to fine-grained detail. Third and more broadly, if we find that the diversity of the input does impact speech categorization, it raises the possibility that development is not equifinal but may reach different results depending on the child's listening needs.

We thus tested a group of 64 children between the ages of 6 and 12. We tested children who only use English but have various degrees of exposure to other languages, dialects, and accents. To achieve this, we purposely recruited students from schools that were known to have either a higher or lower-than-average population of English Language Learners (in our local school district, ELL students are not distributed equally but clustered into schools where more services can be provided) with similar SES backgrounds. We then used a social network measure to calculate a linguistic diversity index for each child. A social network approach has been shown to be a more effective tool for capturing linguistic backgrounds^{82,88} and allowed us to precisely characterize the linguistic variability in children's home and school environments. Importantly, previous studies have shown that factors such as network size and relationship types are strong predictors for language development and processing^{78,79,84,86,87}. While some children did have caregivers who were speakers of other languages, caregiver-to-child interactions were solely in English (by parent report), with varying degrees of accent and dialect differences. Critically, from the social network analysis, we computed two key indices for each child: social network size (the number of individuals each child contacted on a weekly basis for the last 6 months), which has been widely used in previous language processing studies^{84–86}, and social network diversity (out of the total network size, how many individuals speak with a different language, accent, and/or dialect).

Each child then performed a VAS task examining categorization in eight different continua spanning several consonant and vowel contrasts to achieve a comprehensive picture of speech categorization across a range of contrasts. That is, our goal was not to observe how a particular speech cue (e.g., VOT) was interpreted in the face of a particular dialect (e.g., Spanish accented English), but rather to sample across multiple continua to characterize how speech perception as a whole may change in response to diversity more generally. On an exploratory basis, we also collected standardized measures of phonological processing, written word recognition (decoding), and oral language ability.

Results

Figure 2 A shows the mean VAS results as a function of the continuum step (averaged over all eight continua) and age group. It shows the expected developmental effects as the mean function had a steeper slope in older children. Similarly, Fig. 2B C show the differences between network diversity and network size measures. As we described, these differences could reflect true differences in the degree of gradiency of the function, or differences in the trial-by-trial response variability.

Supporting this, Fig. 3 shows three representative participants, showing both their mean function (the purple curve) and individual trial responses (the dots). The first panel has a steep slope and can be compared to the other two panels with similar (shallow) categorization slopes (i.e., shallow). Critically, however, the participant in the middle panel shows individual-trial responses that are clustered around the mean (they are canonically gradient), while the participant in the right-most panel shows how the same slope function might arise from higher categorization inconsistency.

To quantify these dimensions, we fit a Bayesian hierarchical non-linear model to the data, which simultaneously estimated the slope, boundary, and amplitude of the underlying function, as well as the average trial-by-trial response variability around that function (estimates from that model are numerically shown in each panel of Fig. 3). This was fit in a framework that allowed each subject to have their own specific parameters

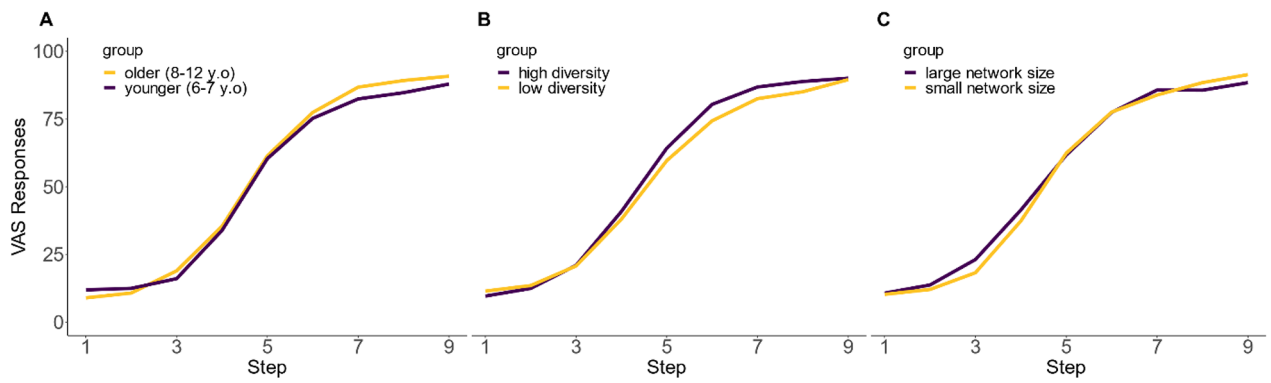


Fig. 2. (A) Two slopes showing older (8–12 years old) and younger (6–7 years old) children's VAS responses averaged over each continuum. (B) Two slopes showing median split social network diversity as high and low social network diversity averaged over each continuum. (C) The same structure compared to panels A and B, but this time the two slopes show network size (median split).

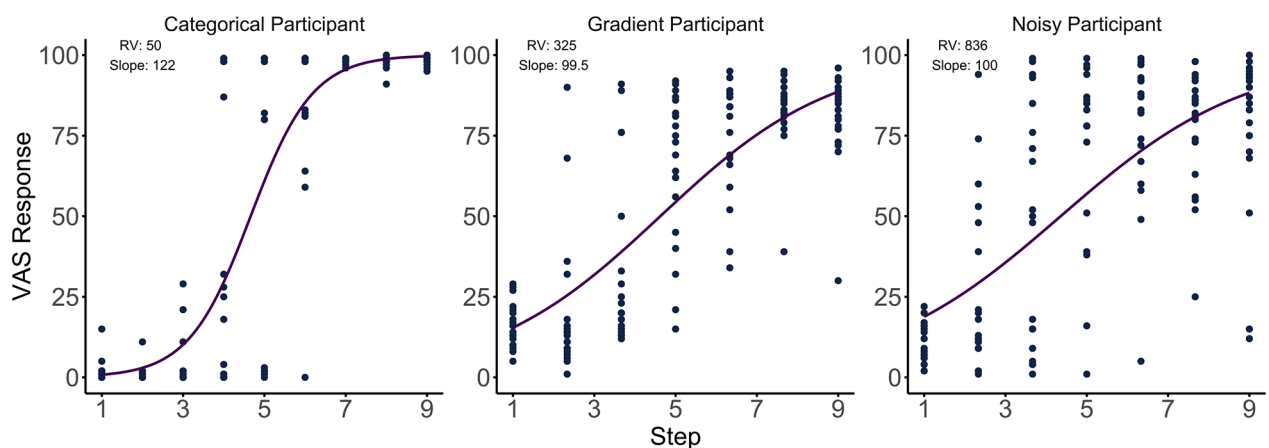


Fig. 3. Three participants' VAS responses (on the y-axis) and the 9-point step (on the x-axis) showing a similar pattern slope function emerging for two distinct types of Response Variability (i.e., Gradient and Noisy participants).

and trial-by-trial variance, while also capturing continuum-specific effects for each subject⁸⁹ to capture the fact that continua may differ in their slopes or boundaries. From this, we extracted estimates of the subject-specific curve parameters to use as the dependent variables in separate regression analyses that examined the effect of Network Diversity, Network Size, and age on each critical index of speech categorization. The Bayesian model is detailed more explicitly in the Methods section.

Effect of age and social network on speech categorization

The first model examined the effect of age and network indices on the slope of the categorization function, explaining approximately 24% of the total variance (see Fig. 4 for effect sizes of each predictor, Table 1 for complete results). The analysis revealed a significant relationship between Slope and Network Diversity ($\beta = -0.37$, $b = -19.49$, $SE = 6.14$, $t(60) = -3.17$, $p = .002$) and Network Size ($\beta = -0.35$, $b = -0.631$, $SE = 0.20$, $t(60) = -3.12$, $p = .002$). In both cases, increased linguistic diversity and increased network size led to a shallower (more gradient) slope.

Next, we repeated the same regression but with *Response Variability* as the dependent variable (Fig. 4 bottom panel, Table 1). This model explained approximately 25% of the variance. This model showed a starkly different pattern of results: only age had a significant effect on Response Variability ($\beta = -0.12$, $b = -0.193$, $SE = 0.05$, $t(60) = -3.383$, $p = .001$). As children age, their speech categorization becomes less noisy (see Table 2; Fig. 4 bottom panel).

It is important to highlight here that while the diversity of the input shapes children's speech categorization strategies, this diversity did not correlate with their language, reading, or phonological processing abilities (all p 's > 0.05). That is, we are not observing these effects because diversity slowed or altered overall language development; rather these effects reflect a specific adaptation in the speech perception system.

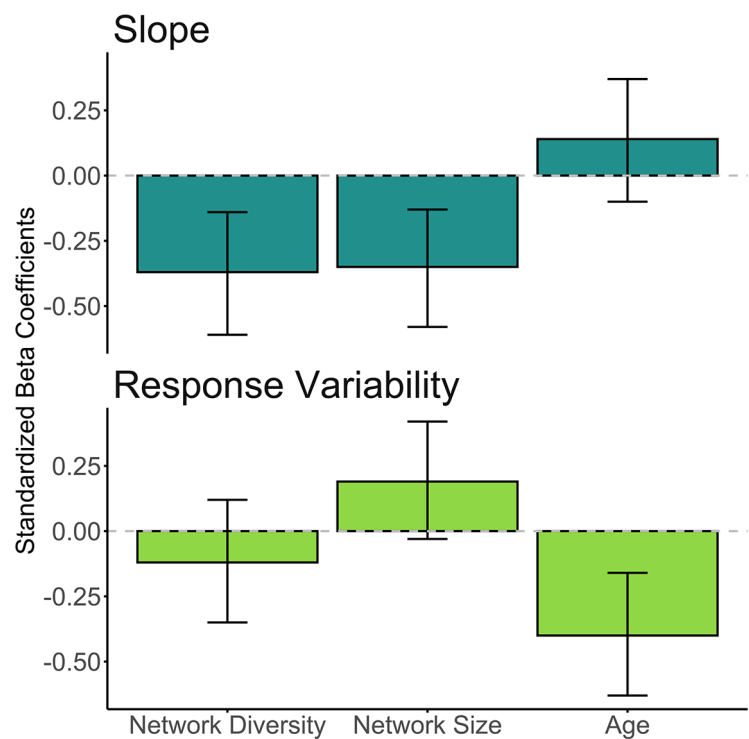


Fig. 4. Standardized regression coefficients (effect sizes) for the effect of network diversity, network size and age on Slope (top panel) and Response Variability (bottom panel). Error bars represent 95% CIs.

Predictor	Slope					Response Variability				
	<i>r</i>	<i>b</i>	<i>b</i> 95% CI [LL, UL]	<i>beta</i>	<i>beta</i> 95% CI [LL, UL]	<i>r</i>	<i>b</i>	<i>b</i> 95% CI [LL, UL]	<i>beta</i>	<i>beta</i> 95% CI [LL, UL]
(Intercept)		107.82**	[100.34, 115.29]				1.35*	[0.19, 2.51]		
Network Diversity	− 0.31*	−19.49**	[−31.79, −7.20]	−0.37	[−0.61, −0.14]	− 0.24	−0.95	[−2.86, 0.96]	−0.12	[−0.35, 0.12]
Network Size	− 0.33**	−0.63**	[−1.04, −0.23]	−0.35	[−0.58, −0.13]	0.21	0.05	[−0.01, 0.12]	0.19	[−0.03, 0.42]
Age	0.05	0.43	[−0.30, 1.17]	0.14	[−0.10, 0.37]	− 0.44**	−0.19**	[−0.31, −0.08]	−0.40	[−0.63, −0.16]
Fit	$R^2 = 0.240^{**}$ 95% CI [0.05, 0.38]					$R^2 = 0.243^{**}$ 95% CI [0.05, 0.38]				

Table 1. Results of regressions using slope and response variability as the criteria. Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. * indicates *p* < .05. ** indicates *p* < .0.

Classes	AIC	BIC	<i>p</i> (BLRT)
1	386.26	395.08	
2	377.63	393.06	0.02
3	349.65	371.70	0.01

Table 2. Goodness of fit measures of model fit for 1-class, 2-class, and 3-class LPA models. P values represent a likelihood ratio test between the current model and the simpler model immediately below it.

Together, these suggest a fairly clear picture: development during these ages is largely about achieving a more consistent encoding of speech cues from moment to moment, but the nature of the linguistic input is just as crucial for tuning the slope of the mean function (the underlying structure of the category).

A multi-dimensional characterization

Individually, both slope and response variability showed the effects of overall development and linguistic environment. However, these indices are not theoretically independent – a low response variability means something different when paired with a steep mean slope (categorical) than when paired with a shallow slope (gradient). A shallow slope could mean something very different depending on whether the response variability is high or low. Consequently, a two-dimensional space (Response Variability by Slope), may provide more insight.

Based on prior work on the VAS task (e.g.,^{55,65}) and on gradient speech perception using other tasks¹⁷, we proposed three potential underlying profiles of children (see preregistration for another project at <https://osf.io/q39yt> for more details). The first profile is children who are **Canonically-Categorical**, featuring low Response Variability and high Slope (Fig. 3A). Listeners who are in this profile have little sensitivity to within-category variation and a strong categorical response. The second profile is **Categorical-but-Noisy**. These listeners have high Response Variability and low Slope (Fig. 3C). These participants respond largely with extreme or even endpoint values (as in the categorical pattern), but they do so inconsistently from trial to trial, which leads the mean slope to be shallow or more gradient-looking. They are attempting a categorical profile but are too noisy to achieve it. The third profile—the **Canonically-Gradient** children (Fig. 3B)—also shows a shallow slope, but for a very different reason. Listeners in this profile have a low slope and low response variability, which suggests that they have a higher sensitivity to within-category variation (i.e., sub-phonemic differences), and they show this sensitivity with less noise.

To examine this possibility, Response Variability and Slope were analyzed with a Latent Profile Analysis (LPA) to identify any potential classes (tidyLPA⁹⁰). The best solution fitting LPA was a three-class solution, see Table 2). The model had a minimum class probability of 0.92, and a maximum class probability of 0.98. The Bootstrapped Likelihood Ratio Test (BLRT) indicated a significant improvement in fit for the three-class solution compared to a two-class solution ($p = .01$), but no significant gains for a 4-class model.

Figure 5 shows a breakdown of the sample, color-coded by the most likely profile for each child (though class membership is probabilistic in an LPA). It shows the expected three groups, with a large group of gradient children (in purple, at the bottom left), smaller groups of noisy children (green, top), and categorical children (yellow, far right).

We next extracted the probability that each participant was a member of each of the three classes (Fig. 5). These class probabilities (i.e., Canonically-Gradient, Canonically-Categorical, Categorical-but-Noisy) were then used as the dependent variables in a set of separate regression analyses that predicted class membership (which reflects children's speech categorization profile) from the linguistic environment and age.

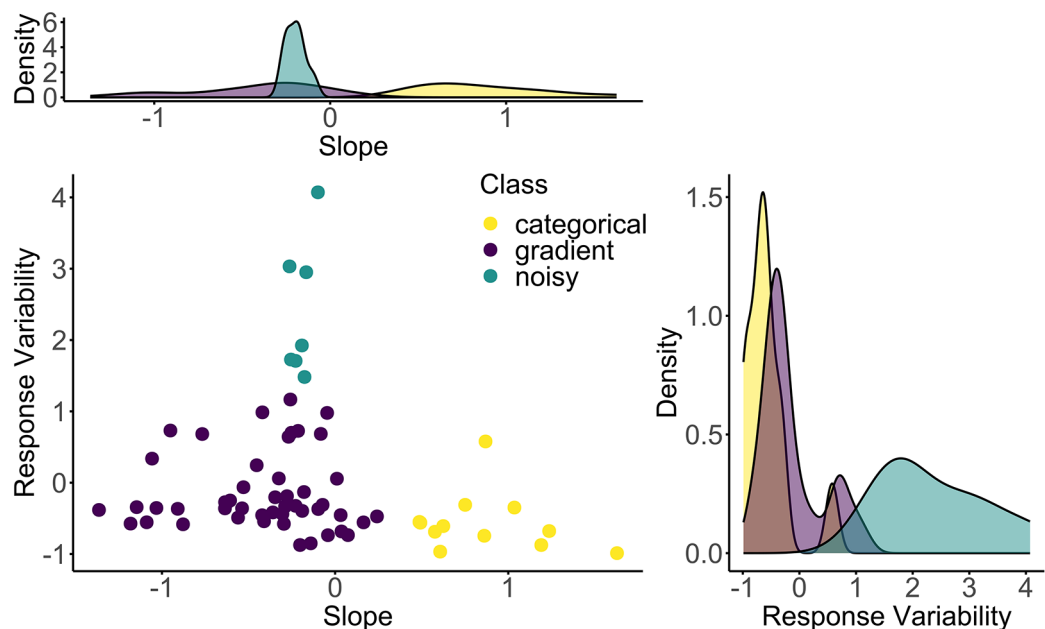


Fig. 5. The three profiles emerging in the two-dimensional space (Response Variability and Slope). The categorical profile emerges in the high slope (i.e., steep slope) and low Response Variability space. The gradient profile emerges in the low slope (i.e., shallow slope) and low Response Variability space. Finally, the noisy profile emerges in the low slope but high Response Variability space. These show the underlying differences between the shallow slopes that can emerge for two separate reasons.

The first model predicted the probability of being in the Canonically-Gradient profile. The model included Network Diversity, Network Size, and Age and explained approximately 18% of the variance (Fig. 6A; Table 3 for complete details; see Fig. 6A for a summary of the effects). The analysis revealed a significant relationship between membership in the Canonically-Gradient profile and Network Diversity ($\beta = 0.38$, $b = 1.270$, $SE = 0.415$, $t(60) = 3.055$, $p = .003$) such that children with more linguistic diversity are more likely to be in the Canonically-Gradient profile. Network Size had a marginally significant effect on gradiancy ($\beta = 0.23$, $b = 0.026$, $SE = 0.013$, $t(60) = 1.92$, $p = .059$) in the same direction. No significant relationships were found between Canonically-Gradient profile membership and Age ($\beta = -0.10$, $b = -0.02$, $SE = 0.024$, $t(60) = -0.80$, $p = .424$).

The second analysis used the probability of being in the Canonically-Categorical profile with the same predictors (Fig. 6B; Table 3). This model explained 25% of the variance. The analysis revealed a significant effect in Network Diversity ($\beta = -0.37$, $b = -1.04$, $SE = 0.203$, $t(60) = -3.125$, $p = .002$) such that increased linguistic diversity leads to a *reduced* likelihood of being in the Canonically-Categorical profile. There was also a significant effect of Network Size ($\beta = -0.03$, $b = -0.02$, $SE = 0.01$, $t(60) = -2.440$, $p = .017$). Similar to Network Diversity, increased Network Size led to a decrease likelihood of being in the Canonically-Categorical profile. Finally, as children get older, their chances of being in the Canonically-Categorical profile increased ($\beta = 0.35$, $b = 0.059$, $SE = 0.020$, $t(60) = 2.96$, $p = .004$).

The final model examined the probability of being in a Categorical-but-Noisy profile. This model explained 10% of the variance. The effect of age was the only significant factor: as children aged, their chances of being in the noisy profile reduced ($\beta = -0.27$, $b = -0.03$, $SE = 0.018$, $t(60) = -2.13$, $p = .036$) (see Fig. 6C, and Table 3).

Figure 7 shows a visualization of the effect of each factor (left panel: age, middle panel: network diversity, right panel: network size) on the likelihood of appearing in each group. The results are clear: increasing age is primarily associated with decreased noise. Higher network diversity and network size push children towards a more gradient response (lower slope and lower noise).

Discussion

The current study investigated the impact of linguistic diversity on speech categorization during a significant period of language development: the onset of formal schooling. This represents a sort of natural experiment on the development of speech perception abilities, as the new distribution of acoustic/phonetic cues is imposed upon the child in a way that is only partially related to the child's prior developmental experience.

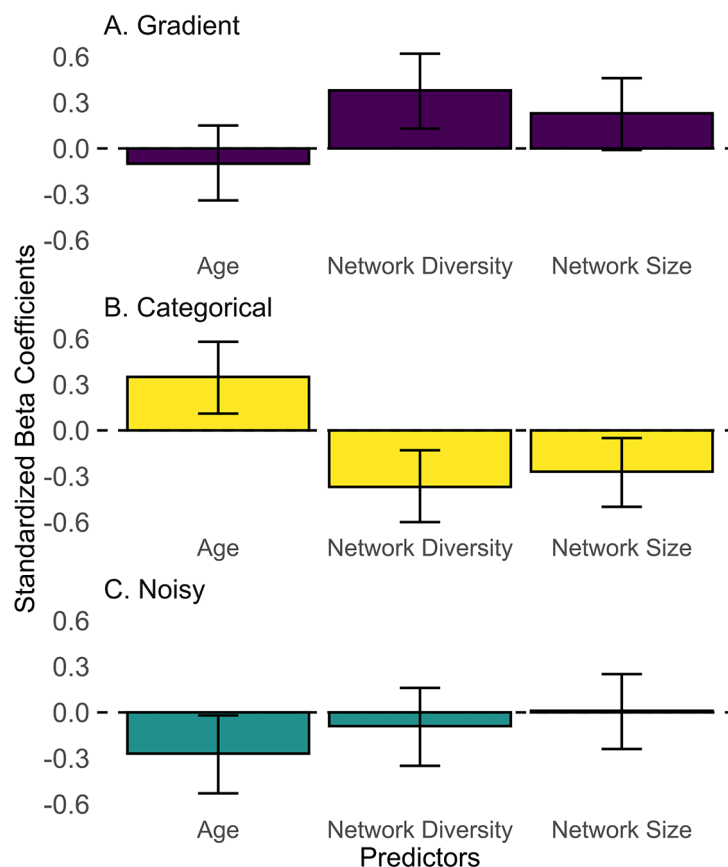


Fig. 6. Beta values with 95% CIs for each profile (Gradient, Categorical, and Noisy) with each of the network measures along with age.

Predictor	Canonically-Gradient				Canonically-Categorical				Categorical-but-Noisy			
	<i>r</i>	<i>b</i>	<i>b</i> 95% CI [LL, UL]	β	β 95% CI [LL, UL]	<i>r</i>	<i>b</i>	<i>b</i> 95% CI [LL, UL]	β	β 95% CI [LL, UL]	<i>b</i>	<i>b</i> 95% CI [LL, UL]
Intercept		0.39	[-0.11, 0.90]				0.13	[-0.28, 0.54]			0.47*	[0.10, 0.85]
Network Diversity	0.33**	1.27**	[0.44, 2.10]	0.38	[0.13, 0.62]	-0.25*	-1.05**	[-1.72, -0.38]	-0.37	[-0.60, -0.13]	-0.22	[-0.84, 0.39]
Network Size	0.21	0.03	[-0.00, 0.05]	0.23	[-0.01, 0.46]	-0.26*	-0.03*	[-0.05, -0.00]	-0.27	[-0.50, -0.05]	0.00	[-0.02, 0.02]
Age	-0.00	-0.02	[-0.07, 0.03]	-0.10	[-0.34, 0.15]	0.25*	0.06**	[0.02, 0.10]	0.35	[0.11, 0.58]	-0.04*	[-0.08, -0.00]
Fit	$R^2 = 0.172^{**}$, 95% CI[0.01, 0.31]				$R^2 = 0.250^{**}$, 95% CI[0.06, 0.39]				$R^2 = 0.098$, 95% CI[0.00, 0.22]			

Table 3. Results of regressions using the three profiles as the criteria. Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively. * indicates $p < .05$. ** indicates $p < .0$.

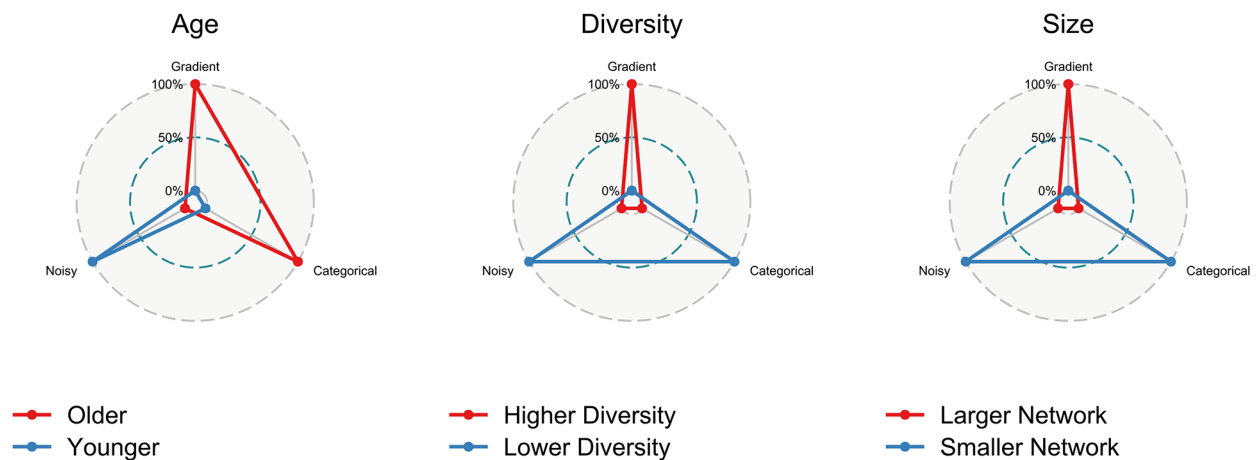


Fig. 7. The three profiles and their distribution with age (mean split as older and younger cohort), diversity (mean split higher and lower diversity), and network size (larger and smaller network).

The effect of social networks

We hypothesized that children with greater linguistic diversity in their social network would exhibit gradient speech categorization, distinct from noisy encoding, while children with less linguistic diversity would display more categorical speech categorization. Our findings support these hypotheses, demonstrating that linguistic diversity shapes children's speech categorization patterns.

First, we found that children with more diverse linguistic environments and who had larger network sizes had a shallower categorization slope. This is reminiscent of prior work on bilingualism^{33,34,42,91,92} which associates a shallow slope with more noise in the perceptual system. However, our use of the VAS allows a more precise characterization that points in a different direction. First, we found this same shallow slope in functionally American English monolingual children with no delays in language, reading, and phonological skills (see Methods). This belies the idea that linguistic diversity introduces a(n) “instability” or “less mature” categorization.

Traditionally, a shallow categorization slope has been interpreted as difficulty in categorization, and, at first glance, our findings align with the previous studies, which have shown that bilinguals or individuals exposed to diverse linguistic input exhibit shallower slopes than monolinguals^{91,92}. However, the VAS task allows us to disentangle whether this reflects disrupted categorization or a functional adaptation to diversity in the linguistic environment. If it's the former (disrupted categorization), we should have observed shallower slopes with higher response variability, but if it were the latter (functional adaptation), we should have seen a shallower slope with reduced variability.

When we considered both dimensions simultaneously – using a latent profile approach, results strongly favored the functional adaptation account. Critically, network diversity significantly increased the likelihood that children would fall into the gradient profile (shallow slope + reduced variability) and it reduced the chances of being in the noisy and the categorical profiles. This suggests the shallower slope associated with more diverse networks reflects true gradiency in speech categorization, not a disrupted categorization. This contrasts with prior approaches to infants and toddlers facing linguistically diverse environments^{33,34,39,41–43,92} which generally suggest a pattern of specialization and narrowing to the single dominant accent in the environment. In contrast, our work does not rule out such specialization, but suggests that when confronted with diverse talkers, children adapt by becoming more flexible.

Gradient representations of speech are widely considered to be more ecological as they may help listeners handle variations in the environment^{52,53,66,74,93–96}. Here, we show evidence that children whose social networks are more linguistically diverse are tuning their speech perception systems to harness this mechanism and are explicitly avoiding a categorical form of perception. This supports the hypothesized function of gradiency and shows that this may be intricately tied to the ambient language environment.

However, it is also important to note that in environments featuring less acoustic diversity, such gradiency may not be needed and could even slow processing. In these cases, children are still adapting – they are reducing noise in the system – but developing toward a more categorical system. Either solution is viable – and in fact both groups of children are developing language reading and phonological skills equally well – but which one is adopted appears to depend on the particular acoustic/phonetic environment they find themselves in.

We note that these findings show similar effects of both network size and network diversity (both lead to shallower slopes and a more gradient categorization profile). It is possible that a larger network size might be the underlying reason for increased linguistic diversity. It is difficult to answer this question with this data, as our network survey did not use an egocentric social network approach which uses a set network size (e.g., asking children to list a fixed number of contacts). Using a fixed number of nodes could, in principle, identify the relative proportion of diverse speakers (and is generally used in egocentric social networks); we avoided it

since this might be more challenging for young children and their families, who may vary widely in the size of their networks.

However, three factors suggest this is not a concern. First, we found no correlation between these two measures in our sample ($r = -.062$), suggesting they represent distinct aspects of the language learning environment. Second, we do have data from other studies with adult listeners using an egocentric social network approach, which shows similar findings for network diversity while controlling for network size (see preliminary results⁷⁵). Third, we point out that even within a dialect region, there is substantial variation in how talkers produce speech. A larger network size (even absent variation in dialect diversity) may also entail more significant acoustic phonetic variability. More broadly, whether or not network size and diversity play unique roles, both are critical factors documenting differences in the language learning environment (as can be seen in previous studies^{84,86}). This suggests that speech categorization is tied to the language learning environment, broadly construed.

We also note that fundamentally, this result is correlational, raising the possibility of reverse causation: do people who are more gradient seek out more diverse networks? This seems unlikely for two reasons. First, speech gradiency is generally not correlated with broader psychosocial or cognitive factors^{65,97}. It seems unlikely that subtle differences in how someone perceives speech would outweigh other factors in social affiliation. Second, this is a natural quasi-experiment – the diversity is imposed by the children's ecology. However, future longitudinal work may be better able to disentangle these issues.

Development

In addition to the effects of network, we also found a significant effect of age. Older children showed reduced response variability and were more likely to exhibit a categorical speech categorization profile.

It is possible that this reduced variability reflects relatively general processes such as the ability to stay engaged to complete the task. However, a companion study from our larger longitudinal work explicitly related VAS performance in 3rd grade students to cognitive and self-regulation as well as ADD symptoms (from parent report measures), and found no correlations⁹⁸. It also showed correlations between response variability in the VAS task and traditional measures of phonological processing (critical for later reading ability) offering some specificity to speech perception. Thus, it seems likely that this measure is capturing something speech specific, though clearly future work is needed to understand both speech-specific and domain-general contributors to performance.

If we take these results at face value, our results offer four critical insights concerning development. First, prior results demonstrate a steepening slope over-development (e.g.,²⁶). This could be consistent with changes to the underlying representation of the category or with changes in the moment-by-moment encoding of speech cues. Our results—showing developmental effects on response variability but not slope—suggest it is the latter. This raises important theoretical questions: current theoretical approaches to the development of speech perception largely focus on changes to stable long-term representations^{8,12,48}. These theoretical tools do not offer any clear way to mechanistically describe the moment-by-moment consistency of either the cue-encoding or the categorization process, nor do they offer mechanisms by which this consistency develops⁹⁸. This is an important goal of future work.

Second, the results of our LPA suggest that the primary effect of age is for children to move from the Categorical-but-Noisy to the Canonically-Categorical profile. This presents a conundrum. Gradiency is widely thought to be functionally beneficial, and at least at a group level, much of the evidence that suggests that speech categorization is gradient overall derives from adults. If this is the case, why is this profile not the “end goal” of development? We note several potential explanations. The assertion of a categorical or gradient profile from the VAS is relative, not absolute – even what we term the “categorical” children may be highly sensitive to fine-grained within-category detail. Indeed, work on the neural encoding of speech cues⁵⁵ suggests that even adults who show more categorical VAS profiles have gradient underlying processes. It is also possible that this categorical profile is an intermediate state in development. As children get older—even beyond the 6–12 range studied here, they are typically exposed to an even more diverse network of language users: their social lives are increasingly dominated by peers, and their network of non-familial adults increases (e.g., they have multiple teachers in high school), and their input is less bound to their home context. This increased diversity is what might be needed for children to reach a gradient adult state. Indeed, our own findings linking diversity to gradiency within this more limited age range may reveal the mechanisms of this longer-term development. Lastly, we also note that this study was conducted toward the end of the COVID-19 pandemic, which forced social isolation onto many children (and which may have had different effects at different ages).

Third, the findings here suggest that there may be more general developmental forces that are not fully described by the environmental factors captured by the network measures. That is, factors beyond the accumulation of the statistical distribution of the input may also play a role. We do not believe this developmental effect represents a general maturation. Rather, age here is likely a proxy for other developmental factors that may drive developmental changes, such as the growth of the vocabulary, gains in articulation skills, or changes in general perceptual precision. These changes are likely impacted by learning and the language learning environment, but they do not themselves represent the kind of input-driven plasticity that is usually posited in the development of speech perception as “distributional learning” (learning the precise distributions of speech cues).

In this regard, one critical factor may be reading and phonological awareness: as our children develop during this window, they learn to read an alphabetic language, which may put pressure on them to engage in more categorical encoding of speech. However, at a broader level, this underscores that development due to input-driven learning (e.g., adaptation to different linguistic environments) may have different impacts on speech perception than development due to these broader and more indirect factors. This highlights the diversity of influence on development at these older ages.

Finally, these findings challenge the notion of a critical period for the development of speech categorization⁹⁹ by suggesting that speech categorization undergoes ongoing development during the school-age years due to various causes—both the immediate learning environment and broader development of skills like language and reading. Here, we pin this development both to environmental input (diversity) and to other mechanisms (the effect of age). This dovetails with the explosion of work on adult plasticity as adults tune their speech categorization to the demands of their environment⁵⁸. However, as noted, most of these focus on the adaptation of specific boundaries to a new accent or style of speaking; here, we suggest that there may be broader adaptations designed to deal with uncertainty or variability more broadly^{2,46,61,66,76}. The present work suggests continuity between these mechanisms of adult adaptation and those of development: listeners shift between different profiles depending on the needs of their social network structure and diversity.

Conclusions and future directions

This study has several implications for future research. First, it highlights the need to consider the role of language diversity in development and that this diversity plays a central role in shaping cognitive mechanisms. It is therefore important to examine linguistically diverse communities or communities such as school-aged children where the social network structure changes. Importantly, our study demonstrates that linguistic diversity in children's social networks plays a crucial role in shaping their speech categorization patterns and that this diversity does not challenge their speech categorization. Children adapt to the needs of their linguistic environment by becoming more sensitive to sub-phonemic differences and variations. This finding challenges the views that a shallow slope in a forced-choice task indicates “less mature,” “less crisp” or unstable categorization, highlighting the importance of considering the linguistic environment when investigating the development of speech categorization.

Furthermore, our results emphasize the significance of employing tasks that can capture the gradient—and multi-dimensional—nature of speech categorization, allowing for a more nuanced understanding of speech processing in children with diverse linguistic experiences.

Most importantly, our study raises the possibility that there is no single form of speech perception (and possibly any other skills). For children facing less linguistic diversity, a more categorical approach may be quite functional, while for children facing more diversity, it may help to adopt a more gradient approach in order to enable the kind of flexibility needed for that environment. Neither is privileged (and it is quite likely that a categorical child confronted with more diversity could adapt), and each works for that specific child. This underscores that there may be no one-size-fits-all or standard approach; rather, the language system tunes itself to the particular demands it is presented with. This clear departure from an equifinal approach raises the possibility that multifinality may be visible elsewhere in language development, as a response to the diversity of language input and language needs.

For many decades, diverse linguistic input was scrutinized for its effects on language development, and diverse language users were excluded as deviant or simply too variant to incorporate into standard models of universal language skills (although see multiple scholars showing the limitations of such approaches^{70,100–104}). However, when captured ecologically and effectively, using measures sensitive to the diversity of people's environments and with a perspective that embraces variation among people in core problems of language, we see that diversity can fundamentally highlight the role of adaptive systems that are employed by the human cognitive system.

Method

All recruitment and experimental procedures were approved by the University of Iowa.

Institutional Review Board (IRB#202007363). All research was performed in accordance with the institutional review board guidelines. All participants and their legal guardians provided consent. Participants received \$22.5 for their participation in this experiment.

Data Availability

All data and scripts can be found on our OSF repository (https://osf.io/bwcz7/?view_only=2377429c7c6847feae8d6d0998644180).

Participants

Seventy participants took part in this experiment. Six participants were excluded from the experiment because of (1) for not paying attention during the experiment ($n=3$); (2) for not completing more than half of the experiment ($n=3$). This yielded 64 participants whose data were analyzed for this study (36 female, 28 male, Age Mean = 8.4, SD = 2, Range = 6–12). All participants were raised as dominant English speakers as measured by the language background questionnaire embedded within the social network survey.

Participants were recruited in a way that was intended to maximize the amount of social and linguistic diversity in the sample. We targeted the recruitment to several elementary schools in Iowa City, IA that were known to have high or low proportions of multi-racial and multi-ethnic children and general recruitment to the University community. Beyond that (and the restriction that children must speak American English as their dominant language), we sought to avoid any a priori sampling restrictions by the research team. Thus, we recruited children without any prescreening and employed the network approach to better understand the linguistic diversity in their environment. We argue that such approaches are better suited and more ecological for investigating the diversity of the linguistic environment⁸⁰.

General procedures

After providing informed consent (informed consent was obtained from all subjects and/or their legal guardian(s)), parents filled out the social network questionnaire while their child began the experiments.

Testing started with the VAS task which was conducted on a Microsoft Surface (touch screen) tablet. Once they completed this task, participants went through three language assessments (Elision, Oral Comprehension, Word attack). The whole experiment took approximately 75 min.

Assessing Speech categorization: VAS Task

Stimuli. Auditory stimuli used in VAS experiments consisted of monosyllabic minimal pairs in English. Eight continua were used included two voicing contrasts (beach-peach [b-p] dime-time [d-t]), five vowel contrasts (beet-boot [i-u], bet-bat [ɛ-æ], pen-pan [ɛ-æ], hat-hot [æ-ɑ], net-nut [ɛ-ʌ]), and one fricative contrast (sip-ship [s-ʃ]).

These continua were not selected on the basis of any acoustic-phonetic theory or any goal of testing a hypothesis about any specific auditory process. Rather, we sought to diversify our items to ensure that our task was tapping speech processing fairly broadly (as opposed to testing a specific contrast). This was important because participants would be exposed to a wide variety of input (a random effect), and we did not want to individually target speech contrasts relative to a particular language, dialect, or accent. Rather, we wanted a representative set of continua.

With this broad goal in mind, we selected these specific continua to balance several competing considerations. First, we were limited by what was technologically feasible for us to construct from natural speech (as we were concerned that children may struggle with synthetic speech). Second, we wanted to use continua that had been previously used in various prior VAS studies^{65,74,75,105,106}. Finally, we needed continua for which both endpoints were words that would be well known to children and picturable (since we were concerned about labeling the VAS endpoints with text as children around this age can be on a wide spectrum of reading abilities).

We did not expect all continua to behave identically (e.g., they were expected to have different boundaries and slopes), and our statistical model (described shortly) was built to capture this, while still providing generalizable estimates for each child. To support the generality of our approach we post-hoc computed correlations between the response variability estimates for each continuum (across subjects), and found generally robust correlations (see OSF page).

To construct the stimuli, we started by recording each endpoint word (e.g., *beach* and *peach*), spoken by an adult man with an American Mid-Western accent. The recordings were done in mono at a sampling rate of 44,100 Hz. These words were recorded in a carrier sentence to ensure uniform prosody and rate (i.e., “This is beach.”). We then selected one exemplar to serve as the endpoint for each continuum, seeking exemplars that were spoken with a uniform falling prosody and were free of any artifacts.

The *beach/peach* and *dime/time* continua manipulated VOT using a progressive cross splicing procedure using PRAAT⁷⁴. We selected segments of the aspiration from /p/ and /t/ whose duration corresponded to the intended VOT, and then replaced the corresponding quantity of the onset of /b/ and /d/ (respectively) with this aspiration.

For the *sip/ship* fricative continua, a spectral averaging procedure was used based on prior studies¹⁰⁵. In this procedure, the frication portion from *sip* and *ship* was first extracted from the full words. Second, the longer of the two was cut so that they would have equal length. Third, the overall spectral mean was calculated from the long-term average spectra of each fricative, and both spectra were aligned to the same average spectral mean. Fourth, a weighted average of the two spectra was constructed to create 9 spectra (one for each continuum step). Fifth, the spectral means of the spectra were shifted in frequency space to create nine steps. Finally, the resulting spectra were used to filter a segment of white noise, which had an envelope that was the average of the /s/ and /ʃ/ endpoints. This procedure was done in MATLAB.

To create the vowel continua, we used TANDEM STRAIGHT¹⁰⁶ (a MATLAB tool) which first extracts periodic information for each endpoint. Next, temporal anchors are manually placed at the beginning, middle, and end of the target sounds. Finally, continua were morphed across two endpoints in nine steps.

After constructing each continuum, we then conducted phonetic analyses measuring critical acoustic/phonetic cues (e.g., formants for the vowels, VOT for voicing contrasts, spectral peak for fricatives). These acoustic measures can be found on the OSF page associated with this project, along with the final stimuli. They generally confirm that these procedures resulted in the intended transformation of phonetic cues along the continua.

Visual stimuli were created using a picture norming process adapted from previous studies¹⁰⁷. For each item, a set of 3–5 candidates was downloaded from a commercial clipart dataset. These were then discussed by focus groups of laboratory members to select the prototypical image and recommend any changes (for a more consistent style, orientation, etc.).

Procedures. VAS testing started with three practice trials that were identical to the experimental trials and were intended to orient the child to the general task. After practice, an experimenter ensured that the child understood the task, and the experimental trials began. On each trial, participants were presented with one auditory token and two images on the screen along with the VAS line. They were asked to tap on the scale to indicate where the sound falls (participants did not report any issues with tapping as all participants were familiar with touch screen systems). To avoid biasing their responses, the line did not contain a slider or any marker until the participant made a response. Participants could change their responses by tapping on a different location on the scale. Once they were satisfied with their response, they hit the space bar to save the response and initiate the next trial.

Trials were blocked by continua to minimize the cognitive effort it would take to remap the scale to new words and/or continua. To control for order effects and image-position bias (i.e., seeing a *beach* image on the right), participants completed two blocks of each continuum, which counterbalanced the location of the endpoints along the response line. For example, if the participant saw a picture of a *beet* on the left and a *boot* on the right in the first block, they saw a *boot* on the left and a *beet* on the right on a later *beet/boot* block. Within a block of trials

(for each continuum), the order of the steps was fully randomized. Each block consisted of 2 repetitions of 9 steps for each continuum or 18 trials/block. With two blocks for each of the 8 continua, this led to 288 total trials.

VAS Analysis. A Bayesian hierarchical growth curve model was fit to the data using the *rstan* package in R (see⁸⁹ for details). Overall, the model estimated a 4-parameter logistic curve for each individual continuum pair. The four parameters of the curve correspond to the lower and upper asymptotes, the crossover or boundary, and the slope at the boundary. The log-variance about the curve was modeled as a quadratic function over the VAS steps to account for the trial-by-trial variance being dependent on the step in the continua (variance would necessarily be higher in the middle of the continua where VAS responses were neither endpoint). This model was fit using a Bayesian model built in a mixed effects framework in which the parameters of the curve (e.g., slope, boundary and asymptotes) and log-variance estimation was all done simultaneously, and included separate factors to account for subject and item level differences. Consequently, each of the 4 parameters in the logistic curve was modeled as a sum of four effects: the population-average parameter value (the overall slope), the continuum-level deviation from that (does a particular continuum have a higher or lower slope than average), individual-level deviation from the average (was a given subject higher or lower), and subject × continuum-level deviation. In addition, for the log-variance, subject, continuum, and subject × continuum effects were implemented as additive effects (i.e., essentially random intercepts) to the general log-variance function over VOT. Note that the use of the continuum-level deviation terms was explicitly included to model the fact that different continua were expected to have different slopes and boundaries, even as we sought to characterize a participants general slope / response variability across multiple continua.

Because the subsequent regression analyses required individual-level estimates while accounting for continuum, we next estimated individual subject’s parameter values. For subject-specific curve parameters, this was done by adding together the population average curve parameters and individual-level deviation effects and representing the average curve function for each individual across all the continua. To get a single subject-specific variance estimate, the log-variance for each individual was calculated at each of the VAS steps and then averaged. The final result, then, is an individual specific slope and an individual specific variance for each subject. To confirm this approach, we examined the correlation of by-subject/by-continuum estimates across continua. Response variability for all continua was highly correlated (a correlation matrix can be found on the OSF page).

Social network questionnaire

A social network questionnaire was created to examine children’s socio-linguistic environment. The questionnaire was administered through RedCap¹⁰⁸. Parents were asked to fill this survey out with their children. The questionnaire asked for basic demographic information as well as parents’ and child’s language background, the amount of child’s vocalization time in hours (M= 27.9 h, SD=17.5 h), the number of individuals the child interacts with on a weekly basis (known as an alter in network science), and the presence of linguistically diverse alters (family and friends). Network Size was quantified as the total number of alters for a given child (M=9.9, SD=3.6). Network Diversity was the number of linguistically diverse alters divided by the total number of alters (M=0.179, SD=0.124). Network estimates were highly independent of each other (see Table 4). The questionnaire can be found on the OSF page (similar network approaches have also been used in other developmental and language research see^{78,80,83–86,88}). In this study, we opted out of using a full egocentric network survey as these tend to be longer and more intensive, which made this type of survey a non-feasible option for this study.

Assessments

Three assessments were employed to characterize language, reading, and phonology abilities in the sample. **Oral language** was assessed with the Oral Comprehension subtest of the Woodcock-Johnson test of achievement (Version IV; ¹⁰⁹. In this test, children heard spoken sentences and were asked to complete them (e.g., “Water looks blue. Grass looks ...” in which the spoken response should be “green”). We also assessed word-level reading skills (decoding) using the Word Attack subtest of the Woodcock Johnson Tests of Reading Mastery Woodcock¹¹⁰. In this assessment children saw a phonetically regular nonsense word (e.g., *plurp*) and read them aloud. They were scored for the accuracy of the pronunciation. Finally, phonological awareness was assessed with the Elision subtest of the Comprehensive Test of Phonological Processing¹¹¹. In this task, children must remove phonological segments from spoken words to form other words (e.g., Say “toothbrush.” Now say “toothbrush” without saying “tooth”, in which the response should be “brush”). To account for differences in age and focus on relative ability, we report both the raw and standard scores for all three tests.

As Table 5 shows, the population average was close to 100. Only one participant had a score of 75. Table 6 shows that all three tests were correlated, which is predicted in this age: children with better language tend to have better reading and phonological processing (and vice versa). These children—regardless of their language environments—were acquiring language and reading skills at typical rates.

	Vocalization	Size	Diversity
Vocalization	1	0.04	-0.03
Size	0.04	1	-0.06
Diversity	-0.03	-0.06	1

Table 4. Correlation matrix for each of the network measures used.

Assessments	Raw Mean	Raw SD	Standardized Mean	Standardized SD
Phonological Processing (CTOPP-Elision)	23.23	7.01	100.89	11.48
Oral Language (WJTA - Oral Comprehension)	18.37	4.58	102.27	15.93
Reading/Decoding (WJRM-Word Attack)	24.40	10.05	105.38	13.61

Table 5. Summary statistics for the assessments both raw scores and standardized scores.

	Phonology	Reading	Language
Phonology	1	0.48	0.23
Reading	0.48	1	0.28
Language	0.23	0.28	1

Table 6. Correlation matrix for each of the assessments.

Data availability

All data and scripts can be found on our OSF repository (https://osf.io/bwcz7/?view_only=2377429c7c6847fea8d6d0998644180).

Received: 25 April 2024; Accepted: 19 November 2024

Published online: 21 November 2024

References

1. Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* **97**, 3099–3111 (1995).

2. Theodore, R. M., Miller, J. L. & DeSteno, D. Individual talker differences in voice-onset-time: contextual influences. *J. Acoust. Soc. Am.* **125**, 3974–3982 (2009).

3. Miller, J. L., Green, K. P. & Reeves, A. Speaking rate and segments: a look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* **43**, 106–115 (1986).

4. Viswanathan, N., Magnuson, J. S. & Fowler, C. A. Information for coarticulation: static signal properties or formant dynamics? *J. Exp. Psychol. Hum. Percept. Perform.* **40**, 1228 (2014).

5. Gay, T. Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am.* **63**, 223–230 (1978).

6. Allen, J. S. & Miller, J. L. Contextual influences on the internal structure of phonetic categories: a distinction between lexical status and speaking rate. *Perception Psychophysics* **63**, 798–810 (2001).

7. Ramscar, M. & Port, R. F. How spoken languages work in the absence of an inventory of discrete units. *Lang. Sci.* **53**, 58–74 (2016).

8. Kuhl, P. K. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* **5**, 831–843 (2004).

9. Kuhl, P. K. et al. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* **9**, F13–F21 (2006).

10. Werker, J. F. & Tees, R. C. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* **7**, 49–63 (1984).

11. Werker, J. F. Perceptual beginnings to language acquisition. *Appl. Psycholinguist.* **39**, 703–728 (2018).

12. Werker, J. F. & Curtin, S. PRIMIR: a developmental framework of infant speech processing. *Lang. Learn. Dev.* **1**, 197–234 (2005).

13. Maye, J., Weiss, D. J. & Aslin, R. N. Statistical phonetic learning in infants: facilitation and feature generalization. *Dev. Sci.* **11**, 122–134 (2008).

14. Pierrehumbert, J. B. Phonetic diversity, statistical learning, and acquisition of phonology. *Lang. Speech.* **46**, 115–154 (2003).

15. Feldman, N. H., Griffiths, T. L. & Morgan, J. L. The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* **116**, 752 (2009).

16. Feldman, N. H., Griffiths, T. L., Goldwater, S. & Morgan, J. L. A role for the developing lexicon in phonetic category acquisition. *Psychol. Rev.* **120**, 751 (2013).

17. McMurray, B., Danelz, A., Rigler, H. & Seedorff, M. Speech categorization develops slowly through adolescence. *Dev. Psychol.* **54**, 1472–1491 (2018).

18. Narayan, C. R., Werker, J. F. & Beddor, P. S. The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Dev. Sci.* **13**, 407–420 (2010).

19. Werker, J. F. & Tees, R. C. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev. Psychobiology: J. Int. Soc. Dev. Psychobiol.* **46**, 233–251 (2005).

20. Bergmann, C. et al. Promoting replicability in developmental research through meta-analyses: insights from language acquisition research. *Child Dev.* **89**, 1996–2009 (2018).

21. Galle, M. E. & McMurray, B. The development of voicing categories: a quantitative review of over 40 years of infant speech perception research. *Psychonomic Bulletin Review.* **21**, 884–906 (2014).

22. Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B. & Quann, C. A. Development of phonological constancy: toddlers’ perception of native- and jamaican-accented words. *Psychol. Sci.* **20**, 539–542 (2009).

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

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

25. Buckler, H., Oczak-Arsic, S., Siddiqui, N. & Johnson, E. K. Input matters: speed of word recognition in 2-year-olds exposed to multiple accents. *J. Exp. Child Psychol.* **164**, 87–100 (2017).

26. Hazan, V. & Barrett, S. The development of phonemic categorization in children aged 6–12. *J. Phonetics.* **28**, 377–396 (2000).

27. Slawinski, E. B. & Fitzgerald, L. K. Perceptual development of the categorization of the /r-w/ contrast in normal children. *J. Phonetics.* **26**, 27–43 (1998).

28. Joanisse, M. F., Manis, F. R., Keating, P. & Seidenberg, M. S. Language deficits in dyslexic children: Speech Perception, Phonology, and morphology. *J. Exp. Child Psychol.* **77**, 30–60 (2000).
29. Werker, J. F. & Tees, R. C. Speech perception in severely disabled and average reading children. *Can. J. Psychology/Revue Canadienne de Psychologie*. **41**, 48–61 (1987).
30. Serniclaes, W., Sprenger-Charolles, L., Carré, R. & Démonet, J. F. Perceptual discrimination of speech sounds in developmental dyslexia. (2001).
31. Manis, F. R. et al. Are speech perception deficits associated with developmental dyslexia? *J. Exp. Child Psychol.* **66**, 211–235 (1997).
32. Serniclaes, W., Ventura, P., Morais, J. & Kolinsky, R. Categorical perception of speech sounds in illiterate adults. *Cognition* **98**, B35–B44 (2005).
33. Pan, L., Ke, H. & Styles, S. J. Early linguistic experience shapes bilingual adults' hearing for phonemes in both languages. *Sci. Rep.* **12**, 4703 (2022).
34. Goh, H. L. & Styles, S. J. Perception of a phoneme contrast in Singaporean English-Mandarin bilingual adults: a preregistered study of individual differences. *44th Annual Meeting of the Cognitive Science Society (CogSci 2022)* (2022).
35. Mack, M. Consonant and vowel perception and production: early english-french bilinguals and English monolinguals. *Perception Psychophysics*. **46**, 187–200 (1989).
36. Flege, J. E., MacKay, I. R. & Meador, D. Native Italian speakers' perception and production of English vowels. *J. Acoust. Soc. Am.* **106**, 2973–2987 (1999).
37. Flege, J. E. Production and perception of a novel, second-language phonetic contrast. *J. Acoust. Soc. Am.* **93**, 1589–1608 (1993).
38. Williams, L. The perception of stop consonant voicing by spanish-english bilinguals. *Perception Psychophysics*. **21**, 289–297 (1977).
39. Llompart, M. Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Biling. Lang. Cogn.* **24**, 481–496 (2021).
40. Stölten, K., Abrahamsson, N. & Hyltén, K. Effects of age of learning on voice onset time: categorical perception of Swedish stops by near-native L2 speakers. *Lang. Speech*. **57**, 425–450 (2014).
41. Casillas, J. Production and perception of the /i/-/I/ vowel contrast: the case of L2-dominant early learners of English. *Phonetica* **72**, 182–205 (2015).
42. Casillas, J. V. Phonetic category formation is perceptually driven during the early stages of adult L2 development. *Lang. Speech*. **63**, 550–581 (2020).
43. Montanari, S., Steffman, J. & Mayr, R. Stop voicing perception in the societal and heritage language of spanish-english bilingual preschoolers: the role of age, input quantity and input diversity. *J. Phonetics*. **101**, 101276 (2023).
44. Apfelbaum, K. S., Kutlu, E., McMurray, B. & Kapnoula, E. C. Don't force it! Gradient speech categorization calls for continuous categorization tasks. *J. Acoust. Soc. Am.* **152**, 3728–3745 (2022).
45. Kleinschmidt, D. F. & Jaeger, T. F. Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol. Rev.* **122**, 148 (2015).
46. Theodore, R. M. & Monto, N. R. Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin Review*. **26**, 985–992 (2019).
47. Xie, X., Weatherholtz, K. & Bainton, L. Rapid adaptation to foreign- accented speech and its transfer to an unfamiliar talker. *J. Acoust. Soc. Am.* **143**, 2013–2031 (2018).
48. McMurray, B., Aslin, R. N. & Toscano, J. C. Statistical learning of phonetic categories: insights from a computational approach. *Dev. Sci.* **12**, 369–378 (2009).
49. Andruski, J. E., Blumstein, S. E. & Burton, M. W. The effect of subphonetic differences on lexical access. *Cognition* **52**, 163–187 (1994).
50. Massaro, D. W. & Cohen, M. M. Phonological context in speech perception. *Perception Psychophysics*. **34**, 338–348 (1983).
51. Miller, J. L. & Volaitis, L. E. Effect of speaking rate on the perceptual structure of a phonetic category. *Perception Psychophysics*. **46**, 505–512 (1989).
52. McMurray, B., Tanenhaus, M. K. & Aslin, R. N. Gradient effects of within-category phonetic variation on lexical access. *Cognition* **86**, B33–B42 (2002).
53. McMurray, B., Tanenhaus, M. K. & Aslin, R. N. Within-category VOT affects recovery from lexical garden-paths: evidence against phoneme-level inhibition. *J. Mem. Lang.* **60**, 65–91 (2009).
54. Kapnoula, E. C. & McMurray, B. edwards, jan Gradient Activation of Speech Categories Facilitates Listeners' Recovery from Lexical Garden Paths, but Not Perception of Speech-in-Noise. <https://osf.io/hw24k> doi: (2020). <https://doi.org/10.31234/osf.io/hw24k>
55. Kapnoula, E. C. & McMurray, B. On the Locus of Individual Differences in Perceptual Flexibility: ERP Evidence for Perceptual Warping of Speech Sounds. <https://osf.io/q9stn> doi: (2021). <https://doi.org/10.31234/osf.io/q9stn>
56. Sarrett, M. E., McMurray, B. & Kapnoula, E. C. Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. *Brain Lang.* **211**, 104875 (2020).
57. Toscano, J. C. & McMurray, B. Cue integration with categories: weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cogn. Sci.* **34**, 434–464 (2010).
58. Kapnoula, E. C., Jevtović, M. & Magnuson, J. S. Spoken Word Recognition: a focus on plasticity. *Annual Rev. Linguistics*. **10**, 233–256 (2024).
59. Nittrouer, S. Learning to perceive speech: how fricative perception changes, and how it stays the same. *J. Acoust. Soc. Am.* **112**, 711–719 (2002).
60. Nittrouer, S. The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults. *J. Acoust. Soc. Am.* **115**, 1777–1790 (2004).
61. Honda, C. T., Clayards, M. & Baum, S. R. Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *J. Exp. Psychol. Hum. Percept. Perform.* **50**(4), 370–394 (2024).
62. Fuhrmeister, P., Phillips, M. C., McCoach, D. B. & Myers, E. B. Relationships between native and non-native speech perception. *J. Experimental Psychology: Learn. Memory Cognition*. **49**, 1161 (2023).
63. Goldstone, R. L. & Hendrickson, A. T. Categorical perception. *Wiley Interdisciplinary Reviews: Cogn. Sci.* **1**, 69–78 (2010).
64. Xie, X., Jaeger, T. F. & Kurumada, C. What we do (not) know about the mechanisms underlying adaptive speech perception: a computational framework and review. *Cortex* **166**, 377–424 (2023).
65. Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J. & McMurray, B. Evaluating the sources and functions of gradiency in phoneme categorization: an individual differences approach. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 1594–1611 (2017).
66. Clayards, M., Tanenhaus, M. K., Aslin, R. N. & Jacobs, R. A. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* **108**, 804–809 (2008).
67. Seidl, A., Onishi, K. H. & Cristia, A. Talker variation aids young infants' phonotactic learning. *Lang. Learn. Dev.* **10**, 297–307 (2014).
68. Singh, L. Influences of high and low variability on infant word recognition. *Cognition* **106**, 833–870 (2008).
69. van Heugten, M. & Johnson, E. K. Input matters: multi-accent language exposure affects word form recognition in infancy. *J. Acoust. Soc. Am.* **142**, EL196–EL200 (2017).
70. Johnson, E. K. & White, K. S. Developmental sociolinguistics: children's acquisition of language variation. *Wiley Interdisciplinary Reviews: Cogn. Sci.* **11**, e1515 (2020).

71. Heugten, M. & Johnson, E. K. Toddlers' word recognition in an unfamiliar regional accent: the role of local sentence context and prior accent exposure. *Lang. Speech*. **59**, 353–363 (2016).
72. Drager, K. Sociophonetic variation in speech perception. *Lang. Linguist Compass*. **4**, 473–480 (2010).
73. Kraljic, T., Brennan, S. E. & Samuel, A. G. Accommodating variation: dialects, idiolects, and speech processing. *Cognition* **107**, 54–81 (2008).
74. McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J. & Subik, D. Gradient sensitivity to within-category variation in words and syllables. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1609 (2008).
75. Kutlu, E., Chiu, S. & McMurray, B. Moving away from deficiency models: Gradiency in bilingual speech categorization. *Front. Psychol.* **13**, 7428 (2022).
76. Xie, X., Theodore, R. M. & Myers, E. B. More than a boundary shift: perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 206–217 (2017).
77. Levy, H., Konieczny, L. & Hanulíková, A. Processing of unfamiliar accents in monolingual and bilingual children: Effects of type and amount of accent experience. *J. Child Lang.* **46**(2), 368–392 (2019).
78. Okocha, A., Burke, N. & Lew-Williams, C. Infants and toddlers in the United States with more close relationships have larger vocabularies. *J. Exp. Psychol. Gen.* **153**(11), 2849–2858 (2024).
79. Burke, N., Brezack, N. & Woodward, A. *Children's Social Networks in Developmental Psychology: A Network Approach to Capture and Describe Early Social Environments*. (2022).
80. Tiv, M. et al. Bridging interpersonal and ecological dynamics of cognition through a systems framework of bilingualism. *J. Exp. Psychol. Gen.* **151**(9), 2128–2143 (2022).
81. Tiv, M., Gullifer, J. W., Feng, R. Y. & Titone, D. Using network science to map what Montréal bilinguals talk about across languages and communicative contexts. *J. Neurolinguistics*. **56**, 100913 (2020).
82. Kutlu, E., Tiv, M., Wulff, S. & Titone, D. Does race impact speech perception? An account of accented speech in two different multilingual locales. *Cogn. Research: Principles Implications*. **7**, 1–16 (2022).
83. Feng, R. Y. et al. A systems approach to multilingual language attitudes: a case study of Montréal, Québec, Canada. *Int. J. Biling.* **28**(3), 454–478 (2023).
84. Lev-Ari, S. The influence of social network size on speech perception. *Q. J. Experimental Psychol.* **71**, 2249–2260 (2018).
85. Lev-Ari, S. Social network size can influence linguistic malleability and the propagation of linguistic change. *Cognition* **176**, 31–39 (2018).
86. Lev-Ari, S. Talking to fewer people leads to having more malleable linguistic representations. *PLoS ONE*. **12**, e0183593 (2017).
87. Lev-Ari, S. How the size of our Social Network influences our semantic skills. *Cogn. Sci.* **40**, 2050–2064 (2016).
88. Kutlu, E., Tiv, M., Wulff, S. & Titone, D. The impact of race on speech perception and accentedness judgements in racially diverse and non-diverse groups. *Appl. Linguist.* **43**, 867–890 (2022).
89. Sorensen, E., Oleson, J. & Kutlu, E. & McMurray, B. A Bayesian hierarchical model for the analysis of visual analogue scaling tasks. *Statistical Methods in Medical Research*.
90. Akogul, S. & Erisoglu, M. An approach for determining the number of clusters in a model-based cluster analysis. *Entropy* **19**, 452 (2017).
91. Bosch, L., Sebastián-Gallés, N. Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy* **2**, 29–49 (2001).
92. Sundara, M., Polka, L. & Genesee, F. Language-experience facilitates discrimination of/d-/in monolingual and bilingual acquisition of English. *Cognition* **100**, 369–388 (2006).
93. Bent, T., Buchwald, A. & Pisoni, D. B. Perceptual adaptation and intelligibility of multiple talkers for two types of degraded speech. *J. Acoust. Soc. Am.* **126**, 2660–2669 (2009).
94. Brown-Schmidt, S. & Toscano, J. C. Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Lang. Cognition Neurosci.* **32**, 1211–1228 (2017).
95. Gwilliams, L., Linzen, T., Poeppel, D. & Marantz, A. In Spoken Word Recognition, the future predicts the past. *J. Neurosci.* **38**, 7585–7599 (2018).
96. McMurray, B. & Jongman, A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychol. Rev.* **118**, 219–246 (2011).
97. Xie, X. & Myers, E. The impact of musical training and tone language experience on talker identification. *J. Acoust. Soc. Am.* **137**, 419–432 (2015).
98. Kim, H. et al. Inconsistent Speech Categorization in School-Age Children with Language and Reading Disabilities. (under review).
99. McMurray, B. The myth of categorical perception. *J. Acoust. Soc. Am.* **152**, 3819–3842 (2022).
100. Baese-Berk, M. & Reed, P. E. Addressing diversity in speech science courses. *J. Acoust. Soc. Am.* **154**, 918–925 (2023).
101. Singh, L., Killen, M. & Smetana, J. G. Global Science requires Greater Equity, Diversity, and Cultural Precision. *APS Observer* **36**, (2023).
102. Tiv, M., Kutlu, E. & Titone, D. Bilingualism moves us beyond the ideal speaker narrative in cognitive psychology. in *Bilingualism across the lifespan* 29–46 Routledge, (2021).
103. Kutlu, E. & Hayes-Harb, R. Towards a just and equitable applied psycholinguistics. *Appl. Psycholinguist.* **44**, 293–300 (2023).
104. McMurray, B., Baxelbaum, K. S., Colby, S. & Tomblin, J. B. Understanding language processing in variable populations on their own terms: towards a functionalist psycholinguistics of individual differences, development, and disorders. *Appl. Psycholinguist.* **44**, 565–592 (2023).
105. Galle, M. E., Klein-Packard, J., Schreiber, K. & McMurray, B. What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cogn. Sci.* **43**, e12700 (2019).
106. Kawahara, H., Masuda-Katsuse, I. & De Cheveigne, A. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun.* **27**, 187–207 (1999).
107. McMurray, B., Samelson, V. M., Lee, S. H. & Tomblin, J. B. Individual differences in online spoken word recognition: implications for SLI. *Cogn. Psychol.* **60**, 1–39 (2010).
108. Obeid, J. S. et al. Procurement of shared data instruments for research electronic data capture (REDCap). *J. Biomed. Inform.* **46**, 259–265 (2013).
109. Schrank, F. A. & Wendling, B. J. The woodcock–Johnson IV. *Contemporary Intellect. Assessment: Theor. Tests Issues*. **4**, 383–451 (2018).
110. Woodcock, R. W. & others. *Woodcock Reading Mastery Tests-Revised* American Guidance Service Circle Pines, MN., (1987).
111. Wagner, R. K., Torgesen, J. K., Rashotte, C. A. & Pearson, N. A. *CTOPP: Comprehensive Test of Phonological Processing* (Pro-ed Austin, 1999).

Acknowledgements

We would like to acknowledge the members of the MACLAB at the University of Iowa for their help in conducting this research, the two anonymous reviewers, and the associate editor for their guidance and helpful feedback.

Author contributions

Conceptualization (E.K., K.B., B.M.), Data Curation (E.K., B.M.), Formal Analysis (E.K., E.S., J.O., B.M.), Funding Acquisition (E.K., B.M.), Investigation (E.K., B.M.), Methodology (E.K., K.B., B.M.), Software (E.S., J.O., E.K., B.M.), Supervision (J.O., B.M.), Validation (E.K., E.S., J.O., B.M.), Visualization (E.K., B.M.), Writing - Original Draft (E.K., B.M.), Writing - Review & Editing (E.K., E.S., J.O., B.M.)

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024