Daylong Mobile Audio Recordings Reveal Multitimescale Dynamics in Infants' Vocal Productions and Auditory Experiences

Anne S. Warlaumont[1],

Kunmi Sobowale[2],

Caitlin M. Fausey[3]

[1]Department of Communication, University of California, Los Angeles

[2]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles

[3]Department of Psychology, University of Oregon

Abstract

The sounds of human infancy—baby babbling, adult talking, lullaby singing, and more—fluctuate over time. Infant-friendly wearable audio recorders can now capture very large quantities of these sounds throughout infants' everyday lives at home. Here, we review recent discoveries about how infants' soundscapes are organized over the course of a day based on analyses designed to detect patterns at multiple timescales. Analyses of infants' day-long audio have revealed that everyday vocalizations are clustered hierarchically in time, vocal explorations are consistent with foraging dynamics, and musical tunes are distributed such that some are much more available than others. This approach focusing on the multi-scale distributions of sounds heard and produced by infants provides new, fundamental insights on human communication development from a complex systems perspective.

Keywords

infancy; vocalization; music; timescales; day-long audio

Over the course of a day, a baby may babble playfully, coo socially, scream manipulatively, attempt to produce spoken words and phrases, laugh, cry, observe quietly, and sleep silently. And they may hear adult speech, sibling screaming, soothing lullabies, recorded voices, water running, dogs barking, clothes rustling, and many, many other sounds. These will depend on the infant's age, physical environment, culture, family structure, personality, and other factors, some of which may be relatively stable and others of which may change within or across days, weeks, and months.

All theories that attempt to explain human communication development (atypical or typical) make assumptions (implicit or explicit) about the statistics of the inputs infants receive.

Corresponding author: Anne S. Warlaumont, University of California, Los Angeles BOX 951538, 2225 Rolfe Hall, Los Angeles, CA 90095-1538 (310) 794-9243, warlaumont@ucla.edu.

And all must account for the statistics of the sounds children produce and how they change over time. It is therefore crucial that ecological data on children's input and productions be recorded in naturalistic settings and with long enough durations to capture the range of contexts and fluctuations infants actually experience and exhibit.

Thanks to innovation in infant-friendly wearable audio recorders and related tools for quantifying patterns in everyday soundscapes (Casillas & Cristia, 2019; Gilkerson et al., 2017; VanDam et al., 2016), we can now begin to characterize infants' sound experiences over the course of an entire day. Foundational discoveries about how these everyday soundscapes matter for young children used machine estimates of overall quantities of specific event types (e.g., number of adult words heard over the day). Human listeners' annotations of short sections of audio sampled from day-long recordings have led to further insights. Now, an additional suite of discoveries is emerging based on analyses that focus on *how sounds are distributed over the course of a day*.

One overarching finding emerging from these studies is that structure in infants' auditory and vocal experiences is nested across seconds, minutes, and hours. There is a general tendency for acoustic events to be distributed non-uniformly. Sounds occur in nested clusters with a combination of many short gaps between sounds along with relatively fewer large gaps. When it comes to sound types, a few types a lot and cumulate to long total durations of experience with that sound type, and many other sound types are experienced less often. We suggest that this non-uniform organization has important implications for understanding and studying human communication development.

Hierarchical clustering of infant and adult vocalizations in time

A complex system can be defined as a system comprised of many interacting components organized at multiple levels. The human brain-body-environment system is one of many naturally occurring complex systems. A substantial body of research has analyzed the behaviors of many natural and simulated complex systems in search of commonalities across domains. One result is an understanding that complex systems tend to generate behavior that fluctuates at a range of nested scales (Kello et al., 2010; Kello, 2013; Viswanathan et al., 2011). This leads to similarity in how a pattern looks when zooming in or out, or *fractality*. There are many reasons scientists have found fractality in behavior intriguing. For one, the *degree* to which there is such nesting in animal behavior often correlates with environmental features. In one example, fractality of human spatial search on a computer screen was higher when resources were clustered compared to when they were uniformly randomly distributed (Kerster et al., 2016). In another example, fractality of albatross foraging was greater when food resources were scarce compared to when they were plentiful (Viswanathan et al., 2011). It is possible that changes in fractality of search patterns are adaptive to one's environment.

Another reason for interest in nested fluctuations is that changes in fractality can be predictive of important state transitions. For example, Stephen et al. (2009) found that there was a predictable peak (an increase followed by a decrease) in the amount of nested structure in adult participants' eye movements immediately before they exhibited instances

of mathematical insight. Stephen et al. noted that this pattern—increase in nested structure during times of reorganization—is a common feature of complex systems. Often this reorganization is purely self-organized, meaning that it results from the internal evolution of the system's state as its components interact with each other. Reorganization can also be initiated or influenced by external inputs to the system.

Bringing a nested-structure focus to study human infant communication, Abney et al. (2016) assessed the degree of hierarchical clustering when infants and caregivers vocalized during day-long recordings. They used the LENA (Language Environment Analysis) system (Gilkerson et al., 2017). LENA enables recording up to 16 hours of infant-centered audio and provides automatic tagging of when infant and adult vocalizations occurred, categorizing infant vocalizations into pre-speech sounds (cooing, babbling, squealing, talking, etc.) versus reflexive or vegetative sounds (cries, laughs, coughs, etc.). Abney et al. found that the difference between the number of vocalizations observed within one time interval and the number in the next consecutive interval is positively correlated with the size of the time interval. In other words, there is more difference in vocalization quantity from one hour to the next hour than from one five-minute interval to the next five-minute interval, consistent with the nesting of vocalization clusters apparent in Figure 1.

In addition to demonstrating hierarchical clustering of both infant and adult vocalizations, Abney et al. (2016) found that the degree of nesting tended to match between infant vocalizations and adult vocalizations. This matching effect held even after controlling for matching in overall rates of vocalization and for temporal proximity between infant and adult vocalizations. Matching was also found to increase with infant age due to adults' scaling pattern becoming more similar to infants'.

Vocalization-to-vocalization changes: a foraging perspective

Analyses of foraging by humans and other animals have also yielded many examples of multi-timescale, non-uniform patterns in behavior. Foraging can be considered broadly to include a wide range of resource types and realms being searched (Todd & Hills, 2020). For example, when animals forage in space for prey and when human adults forage in cognitive semantic networks for items of a particular type, resources tend to be found in nested clusters over time and space (Kerster et al., 2016; Montez et al., 2015; Viswanathan et al., 2011). One way to characterize foraging behavior is to quantify the transitions between consecutive resource-gathering events. From one event to the next (this transition is sometimes called a "step"), we can measure the distance individuals "travel" within a physical or feature space. We can also measure the time between the two events.

Using LENA recordings and their associated automatically identified vocalization onsets and offsets, Ritwika et al. analyzed the acoustic differences and time elapsed from vocalization to vocalization (Figure 2, left panel). They asked whether infant and adult vocalizations could be construed as foraging through pitch and amplitude space. They also looked for evidence that vocal responses from others serve as "resources" for the foraging individual. Inspired by prior foraging research, Ritwika et al. fit mathematical distributions—normal, exponential, lognormal, or pareto (power law)—to the observed acoustic step sizes (Figure

2, right panel) and inter-vocalization intervals. These step sizes and time intervals spanned large ranges. Importantly, the longer step-sizes and inter-vocalization intervals require long recordings to be observed and measured.

Ritwika et al. found that for both infants and adults, the more time that elapsed between consecutive vocalizations, the bigger the change in pitch and amplitude. This aligns with foraging in other domains (Montez et al., 2015; see also Hills et al., 2012). They also found that less time elapsed from one vocalization to the next when infants and adults were interacting with each other. This fits the hypothesis that vocalization is a type of foraging for social responses. It also corresponds with prior research on infant-adult turn-taking. Ritwika et al. also observed that infants' vocalization-to-vocalization pitch movements increased with age (suggesting increasing pitch exploration) while amplitude movements shrank. Adult vocalization steps in both acoustic dimensions grew with infant age. These results connect existing research on infant-adult turn-taking with interdisciplinary work on foraging dynamics. They indicate that vocalization can be construed as an exploratory foraging process and that infant and adult vocal exploration patterns change with age.

Multiple timescales in the sound types infants experience

Focusing now on what day-long audio can reveal about the distributions of specific *types* of sounds infants encounter, we turn to recent discoveries about musical sounds available throughout the day. Second-by-second manual annotation identified which seconds of infants' days were musical as well as the specific voices and tunes within the day's music (Mendoza & Fausey, in press). Because full waking days were annotated, it was possible to observe relatively rare musical voices and tunes and to observe the proportional differences between more and less available musical identities. As shown in Figure 3, instead of each instance of a musical tune cumulating to the same proportion of daily music, certain tunes were *much* more available than others.

How does this distributional non-uniformity matter for infant learning? One possibility is that highly familiar tunes ground musical recognition, providing a base of deep expertise from which infants can learn to generalize to novel tunes. Experiences of numerous less available tunes may help infants establish this generalization capability (see also Smith et al., 2018, for related hypotheses about early learning in other domains). Indeed, it has long been known that word frequencies in natural language follow highly skewed (Zipfian) distributions and this non-uniformity can help adults learn words (Hendrickson & Pefors, 2019). Skewed distributions can also improve adults' category generalization (Carvalho et al., 2021).

What could account for the non-uniformity of musical tune distributions? As with many aspects of infant-caregiver interactions, possible factors might include infant and caregiver preferences, the availability of a particular option in a caregiver's memory, the appeal of novelty, and the comfort of familiarity. Given the many related factors involved, and knowing that complex systems composed of many interacting components often self-organize to generate multi-scale patterns of behavior, the answer is likely to be complicated.

Future work along these lines may enable researchers to compare how distributions of musical (or other audio) stimuli are affected by differing living situations, family structures, and early childhood education experiences. The initial work, especially given that the dataset is available for re-use by other researchers (Mendoza & Fausey, in press), may also enable machine learning researchers to test whether the non-uniformities of input experienced by human infants yield improved capacities for machine learning (see also Bambach et al., 2018; Ossmy et al., 2018). This may in turn enhance our understanding of how these distributional features affect human infants' perceptual learning.

Detecting events within day-long audio: Automatic vs. manual annotation

Automated algorithms available for annotating day-long child-centered audio include the LENA system's proprietary software as well as a handful of open-source alternatives (Le Franc et al., 2018; Räsänen et al., 2021; Schuller et al., 2017). LENA annotates recordings with a closed set of mutually exclusive sound source labels and estimates counts of adult words, child vocalizations, and back-and-forth conversational turns between the child and adults. One huge advantage of automated annotation is that annotation time does not scale prohibitively with recording length. Another advantage is that the exact same algorithm can be shared across projects, eliminating variation that can occur when human annotators with different life and professional experiences interpret sounds differently.

However, automatic annotation accuracy is often lower than that of human listeners (e.g., Ferjan Ramírez et al., 2021). Further, algorithms originally trained with specific datasets for specific purposes may not generalize well. For example, LENA was developed for the purpose of obtaining word, vocalization, and turn counts at the 5-minute, 1-hour, and day-long levels (Gilkerson et al., 2017). Inaccuracies in the annotation might make the algorithm unsuitable for research projects that demand higher accuracy or that use the labels for other purposes. Moreover, for many meaningful units within everyday recordings, no automatic algorithms are currently up to the task (e.g., Adolph, 2020). One issue with day-long child-centered audio recordings is that they are among the most difficult types of conversational speech data for automated systems to accurately tag (Casillas & Cristia, 2019).

An alternative is for human listeners to perform annotation. This can be an enormous undertaking—for example, 6400 person hours were required to manually annotate the features, voices, and tunes in 35 day-long audio recordings (Mendoza & Fausey, in press). Infrastructure supporting sharing data and protocols (e.g., Gilmore et al., 2018; VanDam et al., 2016) helps to maximize value of such investments. For example, sharing manual annotations provides training and evaluation for machine algorithms (e.g., Le Franc et al., 2018; Räsänen et al., 2021; Schuller et al., 2017), which in turn provides new tools for annotating day-long recordings.

We expect that as speech recognition and other automatic audio processing algorithms improve, coupled with increasing availability of datasets of human-annotated audio, it will become possible to automatically identify words, emotions, and more within child-centered day-long audio recordings. Such advances will permit analyses of nested clustering in

additional domains. They might also enable the detection of interactions across domains that partially contribute to the skewed distributions and nested clustering patterns within domains.

Capturing day-long real-world audio recordings also raises privacy concerns. Researchers must explain the issues and enable participants to make informed decisions about participation and use of their data. Some devices, like the TILES recorder (Feng et al., 2018), provide investigators the flexibility to extract and collect only specific features (e.g., speech onset and offset time, pitch estimates) from the audio input. Collecting features alone may better preserve privacy but may not be suitable for every research question and limits re-analysis when improved automatic audio processing tools become available.

Broader implications and future directions

It is clear that, over the course of a day, infant vocalizations and auditory experiences are organized in patterns that unfold at multiple timescales, from seconds to hours. The patterns likely extend to longer timescales (days, weeks, months) as well as to shorter timescales within utterances (Kello et al., 2017). Such multi-scale behavior is characteristic of complex systems involving many interacting components, such as networks of neurons and networks of locally interacting social agents (Kello et al., 2010). It fits with the view that infant development emerges within a complex system of richly interacting components within and external to the infant (Frankenhuis et al., 2019; Oakes & Rakison, 2020; Wozniak et al., 2016).

Future research should explore how patterns of productions and input at shorter and longer timescales are related to other features of the physical and social environment (e.g., material resources, culture, family structure). Such work could help identify some of the mechanisms contributing to the multi-timescale patterns described above. It would also build bridges with other disciplines, like anthropology (Cristia et al., 2017; Frankenhuis et al., 2019).

Future research should also explore the extent to which fractal analyses provide unique information from other methods used to analyze time series data that do not focus on degree of self-similarity across timescales (e.g., Jebb et al., 2015). An explicit focus on dynamics across timescales and ecological contexts enables these comparisons.

Multi-scale patterns in human infant auditory and vocal experiences may also relate to brain plasticity, mental and physical health, and cognitive development. Research with adult humans has documented individual differences in the balance of exploration and exploitation across a range of spatial and cognitive foraging tasks. These differences are often consistent across domains and associated with performance (Todd & Hills, 2020). Regarding early-life development, experimental research using rodent models suggests that differences in physical environment (e.g., a cage having or not having adequate nesting materials) can lead to differences in the predictability of maternal behavior, in turn leading to changes in offspring brain development and variations in cognition, memory, and anhedonia (Glynn & Baram, 2019). Predictable, repeated interactions between a caregiver and infant may signal safety and slow the maturation of corticolimbic circuitry, increasing plasticity

and improving future emotion regulation (Gee & Rhodes, 2021). However, unpredictable rare positive experiences, like listening to a New Year's Holiday song, also seem to prolong brain plasticity (Tooley et al., 2021). Most findings about how environmental experiences affect brain plasticity derive from animal models. Translating this research to humans will be facilitated by detailed data on the distributions of different events types at day-long timescales and in highly naturalistic contexts— such assays would enable the operationalization of predictability and environmental enrichment in human development.

Conclusion

Data on infant vocal productions and auditory experiences acquired from day-long real-world recordings reveal multi-timescale fluctuations and skewed distributions of event types across domains of infant experience. Such patterns often arise through self-organization of complex systems of many interacting components. The findings thus support a complex-systems orientation and underscore the richness and complexity of development as it unfolds in a diverse range of physical, social, and physiological contexts.

Acknowledgements

References

Abney DH, Warlaumont AS, Oller DK, Wallot S, & Kello CT (2016). Multiple coordination patterns in infant and adult vocalizations. Infancy, 22(4), 514–539. doi:10.1111/infa.12165 [PubMed: 29375276]

Adolph KE (2020). Oh, Behave! Infancy, 25, 374–392. doi:10.1111/infa.12336 [PubMed: 33100922]

Bambach S, Crandall DJ, Smith LB, & Yu C (2018). Toddler-inspired visual object learning. In NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems (pp. 1209–1218).

Carvalho PF, Chen C, & Yu C (2021). The distributional properties of exemplars affect category learning and generalization. Scientific Reports, 11, 11263. doi: 10.1038/s41598-021-90743-0 [PubMed: 34050226]

Casillas M, & Cristia A (2019). A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings. Collabra: Psychology, 5(1), 24. doi:10.1525/collabra.209

Cristia A, Dupoux E, Gurven M, & Stieglitz J (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. Child Development, 90(3), 759–773. doi:10.1111/cdev.12974 [PubMed: 29094348]

Feng T, Nadarajan A, Vaz C, Booth B, & Narayanan S (2018). TILES audio recorder: An unobtrusive wearable solution to track audio activity. In WearSys '18: Proceedings of the 4th ACM Workshop on Wearable Systems and Applications (pp. 33–38). doi:10.1145/3211960.3211975

Ferjan Ramírez N, Hippe DS, & Kuhl PK (2021). Comparing automatic and manual measures of parent-infant conversational turns: A word of caution. Child Development, 92(2), 672–681. doi: 10.1111/cdev.13495 [PubMed: 33421100]

Frankenhuis WE, Nettle D, & Dall SRX (2019). A case for environmental statistics of early-life effects. Philosophical Transactions of the Royal Society B, 374, 20180110. doi: 10.1098/rstb.2018.0110

Gee DG, & Cohodes EM (2021). Influences of caregiving on development: A sensitive period for biological embedding of predictability and safety cues. Current Directions in Psychological Science doi:10.1177/09637214211015673

Gilkerson J, Richards JA, Warren SF, Montgomery JK, Greenwood CR, Oller DK, Hansen JHL, & Paul TD (2017). Mapping the early language environment using all-day recordings and automated analysis. American Journal of Speech-Language Pathology, 26(2), 248–265. [PubMed: 28418456]

Gilmore RO, Kennedy JL, & Adolph KE (2018). Practical solutions for sharing data and materials from psychological research. Advances in Methods and Practices in Psychological Science, 1(1), 121–130. doi: 10.1177/2515245917746500 [PubMed: 31157320]

Hendrickson AT (2019). Cross-situational learning in a Zipfian environment. Cognition, 189, 11–22. doi: 10.1016/j.cognition.2019.03.005 [PubMed: 30903853]

Hills TT, Jones MN, & Todd PM (2012). Optimal foraging in semantic memory. Psychological Review, 119(2), 431–440. doi:10.1037/a0027373 [PubMed: 22329683]

Jebb AT, Tay L, Wang W, & Huang Q (2015). Time series analysis for psychological research: Examining and forecasting change. Frontiers in Psychology, 6, 727. doi: 10.3389/fpsyg.2015.00727 [PubMed: 26106341]

Kello CT (2013). Critical branching neural networks. Psychological Review, 120(1), 230–254. doi:10.1037/a0030970 [PubMed: 23356781]

Kello CT, Brown GDA, Ferrer-i-Cancho R, Holden JG, Linkenkaer-Hansen K, Rhodes T, & Van Orden GC (2010). Scaling laws in cognitive sciences. Trends in Cognitive Sciences, 14, 223–232. doi:10.1016/j.tics.2010.02.005 [PubMed: 20363176]

Kello CT, Dalla Bella S, Médé B, & Balasubramaniam R (2017). Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes. Journal of the Royal Society Interface, 14(135). doi: 10.1098/rsif.2017.0231

Kerster BE, Rhodes T, & Kello CT (2016). Spatial memory in foraging games. Cognition, 148, 85–96. doi:10.1016/j.cognition.2015.12.015 [PubMed: 26752603]

Le Franc A, Riebling E, Karadayi J, Wang Y, Scaff C, Metze F, & Cristia A (2018). The ACLEW DiViMe: An easy-to-use diarization tool. In Proc. Interspeech 2018 (pp. 1383–1387). doi:10.21437/Interspeech.2018-2324

Mendoza JK, & Fausey CM (in press). Everyday music in infancy. Developmental Science, e13122. doi:10.1111/desc.13122

Montez P, Thompson G, & Kello CT (2015). The role of semantic clustering in optimal memory foraging. Cognitive Science, 39(8), 1925–1939. doi:10.1111/cogs.12249 [PubMed: 25959228]

Oakes LM, & Rakison DH (2020). Developmental cascades Oxford University Press.

Ossmy O, Hoch JE, MacAlpine P, Hasan S, Stone P, & Adolph KE (2018). Variety wins: Soccer-playing robots and infant walking. Frontiers in Neurorobotics, 12, 19. doi: 10.3389/fnbot.2018.00019 [PubMed: 29867427]

Ritwika VPS, Pretzer GM, Mendoza SM, Shedd C, Kello CT, Gopinathan A, & Warlaumont AS (2020). Exploratory dynamics of vocal foraging during infant-caregiver communication. Scientific Reports, 10, 10469. doi:10.1038/s41598-020-66778-0 [PubMed: 32591549]

Räsänen O, Seshadri S, Lavechin M, Cristia A, & Casillas M (2021). ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. Behavior Research Methods, 53, 818–835. doi:10.3758/s13428-020-01460-x [PubMed: 32875399]

Schuller B, Steidl S, Batliner A, Bergelson E, Krajewski J, Janott C, Amatuni A, Casillas M, Seidl A, Soderstrom M, Warlaumont AS, Hidalgo G, Schnieder S, Heiser C, Hohenhorst W, Herzog M, Schmitt M, Qian K, Zhang Y, & Zafeiriou S (2017). THE INTERSPEECH 2017 computational paralinguistics challenge: Addressee, cold & snoring. In Proc. Interspeech 2017 (pp. 3442–3446). doi:10.21437/Interspeech.2017-43

Smith LB, Jayaraman S, Clerkin E, & Yu C (2018). The developing infant creates a curriculum for statistical learning. Trends in Cognitive Sciences, 22(4), 325–336. doi: 10.1016/j.tics.2018.02.004 [PubMed: 29519675]

Stephen DG, Boncoddo RA, Magnuson JS, &Dixon JA (2009). The dynamics of insight: Mathematical discovery as a phase transition. Memory & Cognition, 37, 1132–1149. doi:10.3758/MC.37.8.1132 [PubMed: 19933457]

Todd PM, & Hills TT (2020). Foraging in mind. Current Directions in Psychological Science, 29(3), 309–315. doi:10.1177/0963721420915861

Tooley UA, Bassett DS, & Mackey AP (2021). Environmental influences on the pace of brain development. Nature Reviews Neuroscience, 22, 372–384. doi: 10.1038/s41583-021-00457-5 [PubMed: 33911229]

VanDam M, Warlaumont AS, Bergelson E, Cristia A, Soderstrom M, De Palma P, & MacWhinney B (2016). HomeBank: An online repository of daylong child-centered audio recordings. Seminars in Speech and Language, 37(2), 128–142. doi:10.1055/s-0036-1580745 [PubMed: 27111272]

Viswanathan GM, Da Luz MGE, Raposo EP, & Stanley HE (2011). The physics of foraging Cambridge University Press.

Wozniak RH, Leezenbaum NB, Northrup JB, West KL, & Iverson JM (2016). The development of autism spectrum disorders: Variability and causal complexity. WIREs Cognitive Science doi: 10.1002/wcs.1426

Recommended Readings

Casillas M, & Cristia A (2019). See References. Provides a tutorial on how to start working with day-long audio recordings

Cychosz M, Romeo R, Soderstrom M, Scaff C, Ganek H, Cristia A, Casillas M, de Barbaro K, Bang JY, & Weisleder A (2020). Longform recordings of everyday life: Ethics for best practices. Behavior Research Methods, 52, 1951–1969. doi:10.3758/s13428-020-01365-9. Provides an extensive discussion of ethical considerations when working with day-long audio recordings of children. [PubMed: 32103465]

Kello et al. (2010). See References. Discusses multi-scale dynamics in cognitive science in relation to underlying system characteristics and connecting to other scientific domains

Rowe ML, & Snow CE (2019). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. Journal of Child Language, 47(1), 5–21. doi: 10.1017/S0305000919000655. An overview of research on language input and its role in language development. [PubMed: 31668157]

Warlaumont AS (2020). "Infant vocal learning and speech production." In Lockman JJ (Ed.) and Tamis-LeMonda CS (Eds.), The Cambridge handbook of infant development (pp. 602–631). Cambridge University Press. doi: 10.1017/9781108351959.022. An overview of research on infant vocal production and speech development.

Onsets of child vocalizations

Full day–long recording (~10 hours)

1 hour within the day

5 minutes within the hour
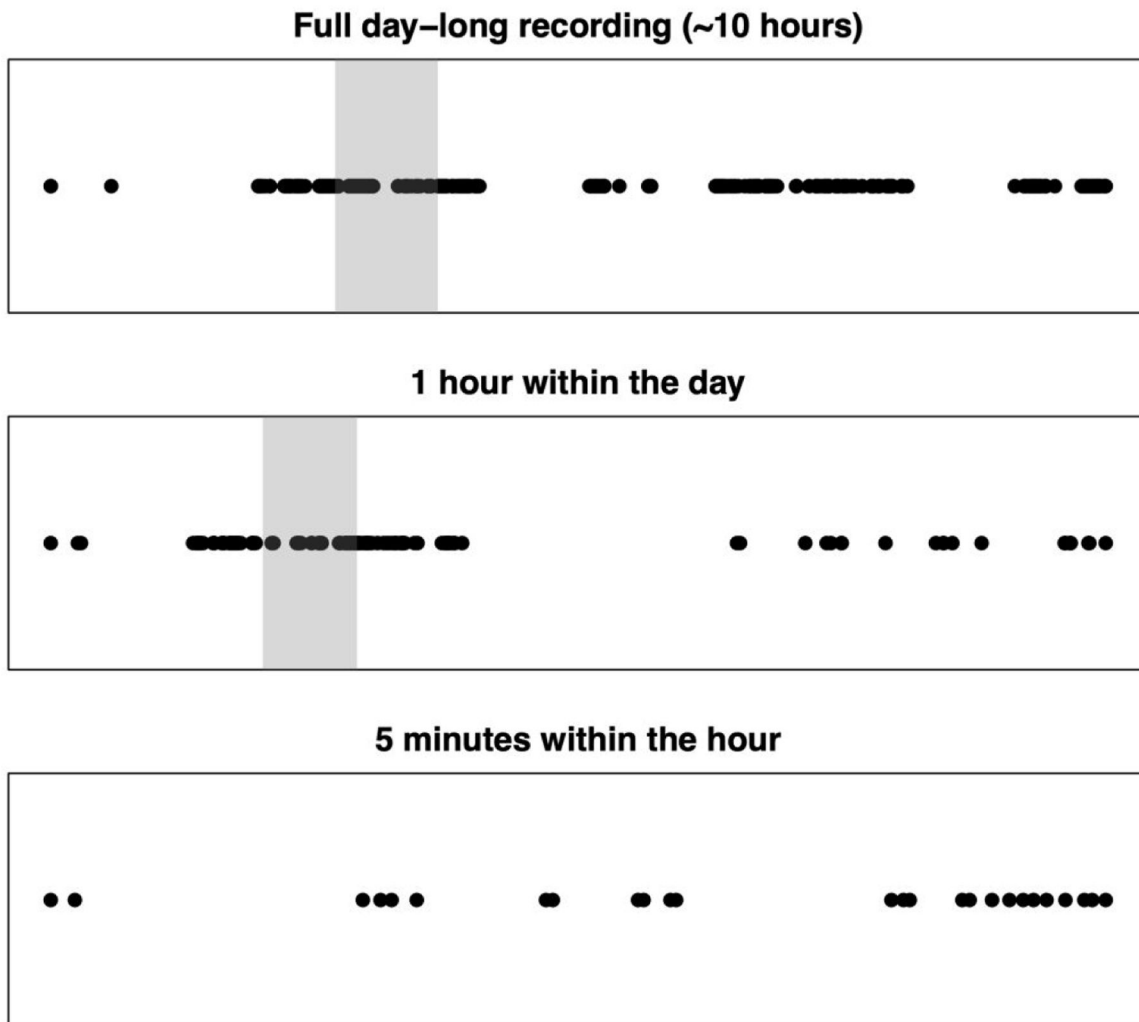
Figure 1.
The circles in the top panel show the onsets of automatically-identified vocalizations in a 9-month-old infant's day-long audio recording. The horizontal position of each circle represents time, with left being earlier in the day and right being later in the day. It is apparent that the infant vocalized in clusters over the day, with some of the clusters being denser and/or lasting longer than others. The area with gray background is an hour-long period and forms the basis of the middle panel. The middle panel thus presents a zoomed-in version of the top panel. It can be seen that within that hour, the infant vocalized in clusters, with the pattern of clustering appearing similar to the clustering at the day-level even though the timescale is much smaller. The area with gray background is a 5-minute-long period and forms the basis of the bottom panel. It is apparent that even within the 5-minute period, the infant vocalized in clusters. Again, the nature of the clustering shows similarity in its patterning to the clustering at hour-long and day-long scales. This figure thus highlights that
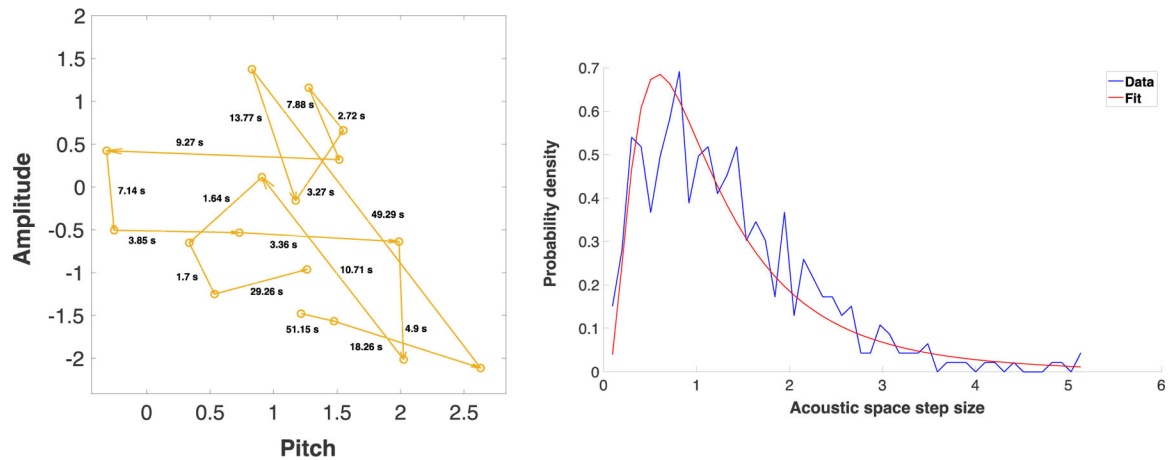
there is fractality, i.e., self-similarity when zooming in or zooming out, in the how infant vocalizations pattern into clusters over time.
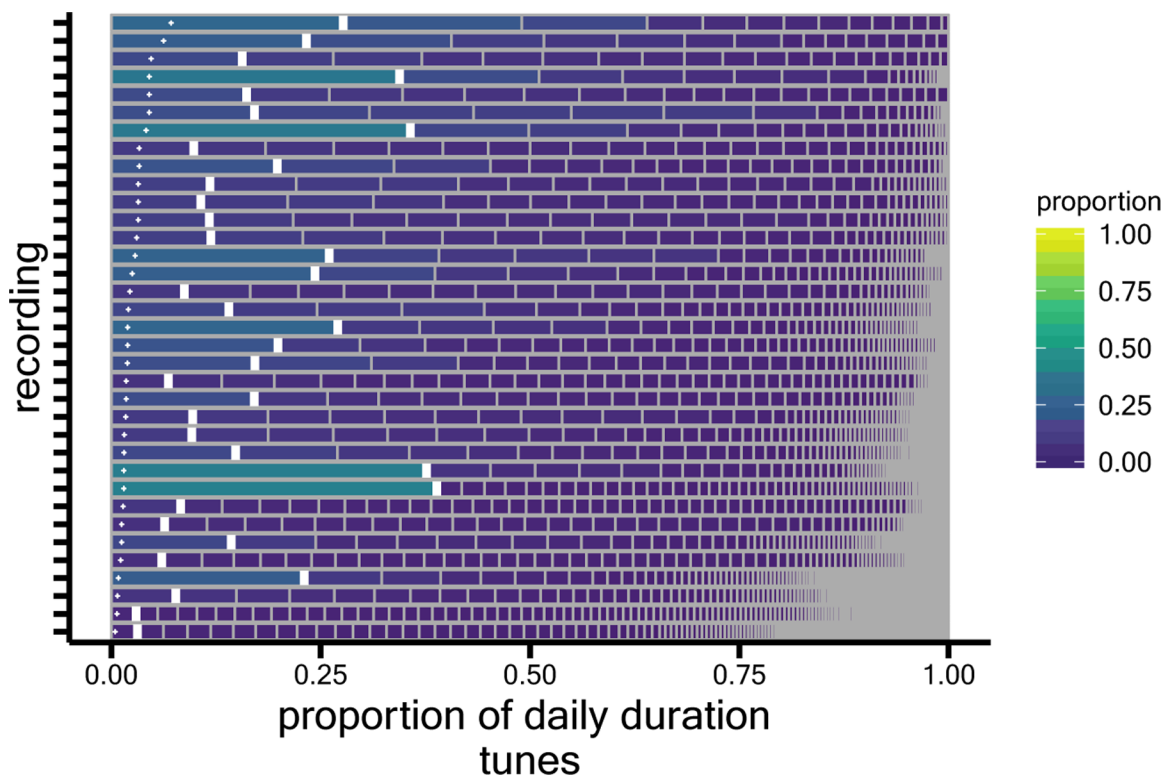
Figure 2.

Left: A sample of some of the vocalization "movements", a.k.a. "steps", of a 3-month-old infant. Each point represents a single vocalization. The horizontal axis represents the mean pitch of the vocalization (log-transformed and normalized with respect to the entire infant vocalization dataset of Ritwika et al., 2020). The vertical axis represents the mean intensity of the vocalization in dB (also normalized). Each arrow corresponds to one step between consecutive infant pre-speech sounds. The numbers next to each arrow represent the duration of time that elapsed between the two vocalizations (i.e., the inter-vocalization interval). Acoustic space step size was defined as the distance in the two plotted acoustic dimensions between the two vocalization points. Right: An example of acoustic step size distribution for a 2-month-old infant's recording, focusing on the infant's pre-speech sounds and specifically on steps where the first vocalization did not receive an adult response. The x-axis plots the acoustic space step size (i.e., the difference, taking into account both pitch and amplitude, between two consecutive infant vocalizations). The y-axis shows the likelihood of observing steps of a given size. It can be seen that smaller step sizes are generally more frequent, but that larger step sizes (spanning more than 2 and up to 5 standard deviations in the acoustic dimensions) do occur. The blue curve shows the histogram of step sizes from the raw data. In this case a lognormal distribution was the best type of function to fit that histogram. The lognormal fit is shown by the red curve. The specific parameters of lognormal fits can be compared across recordings and interactive contexts. This can provide information about how infant vocal dynamics change, such as with age or related to whether the infant is or is not engaged in vocal interaction with caregivers. *Right panel adapted from* Ritwika et al. (2020).

Figure 3.
How much each of a day's many tunes contributes to the full daily tune distribution for each of 35 infants. Each row corresponds to one day-long audio recording, segmented into unique tune identities (e.g., *Twinkle Twinkle Little Star*, *Itsy Bitsy Spider*, *Shake It Off*, *Everybody loves potatoes*, *Short whistle*, and so on). Within the row, each distinct tune's relative duration (i.e., the proportion of the day's musical time) is shown, organized from most available to the infant on the left to least available on the right. The observed proportion of each recording's most available tune is marked by each thick white vertical line. The small white +'s show the proportion that would be expected per tune if each tune were equally available to the infant. Recordings are sorted with those containing the fewest total number of distinct tunes on the top. *Figure adapted/reprinted from* Mendoza & Fausey (in press).