

Hearing water temperature: Characterizing the development of nuanced perception of auditory events

Tanushree Agrawal (t2agrawa@ucsd.edu), Michelle Lee (m.lee@ucsd.edu), Amanda T. Calcetas (acalceta@ucsd.edu), Danielle Clarke (djclarke@ucsd.edu), Naomi Lin (ntlin@ucsd.edu),

Adena Schachner (schachner@ucsd.edu)

University of California, San Diego, Department of Psychology
9500 Gilman Drive M/C 0109, San Diego, CA 92093-0109 USA

Abstract

Without conscious thought, listeners link events in the world to sounds they hear. We study one surprising example: Adults can judge the temperature of water simply from hearing it being poured. How do these nuanced perceptual skills develop? Is extensive auditory experience required, or are these skills present in early childhood? In Exp.1, adults were exceptionally good at judging whether water was hot vs. cold from pouring sounds ($M=93\%$ accuracy; $N=104$). In Exp.2, we tested this ability in $N=113$ children aged 3-12 years, and found evidence of developmental change: Age significantly predicted accuracy ($p<0.001$, logistic regression), such that 3-5 year old children performed at chance while 85% of children age 6+ answered correctly. Overall our data suggest that perception of nuanced differences between auditory events is not part of early-developing cross-modal cognition, and instead develops over the first six years of life.

Keywords: auditory development; cross-modal perception; auditory perception; nature vs nurture; individual differences

Introduction

Without conscious thought, we make inferences about events in the world from sounds we hear: We distinguish heavy from light rain, or whether a car is hurtling towards us at a reasonable versus dangerous speed. Sounds convey information about their sources (Gibson 1966, 1977; Gaver, 1993a, 1993b), and we use even subtle differences in everyday sounds to guide our decisions and actions: Using sound alone, we are able to detect when a water bottle or cup is close to full, and use this information to stop pouring and prevent overflowing (Cabe and Pittenger, 2000; Perfecto, Donnelly, and Critcher 2018). By reasoning about the sources of auditory input, we link perceptual input to a rich, cross-modal representation of what had to happen in the world to cause the sound (Cusimano, Hewitt, Tenenbaum & McDermott, 2019; Gerstenberg, Siegel & Tenenbaum, 2018).

How do humans develop the ability to perceive the sources of auditory events in a nuanced way? What is the role of experience, versus innate or early-developing factors? The literature provides two conflicting theories of how humans develop such skills. Work on adult perception of auditory events posits that such skills require experience with the sounds of relevant everyday events (e.g. Gaver, 1993a; 1993b). By this account, people accumulate cross-modal experience of various events (pouring, striking, splashing), involving various materials (liquids, metal, wood), and learn the types of acoustic features associated with materials and

types of events from this experience. This account makes the prediction that young children may be unable to categorize nuanced properties of auditory events, until they have specific experience with the relevant events.

Contrastingly, developmental literature on cross-modal perception suggests that some relevant aspects of auditory event perception develop early in infancy. By 1 month of age, infants realize that sounds and impacts go together, and expect to hear a sound when they see an impact between two objects, or between an object and a surface, such as a bouncing ball (Bahrick, 1983, 1988; Kopp, 2014; Lewkowicz, 1992, 1994; Spelke, 1979). By this age, infants also match speech sounds (e.g. vowel sounds) to a face moving in ways appropriate to produce that sound (Kuhl & Meltzoff, 1982). Infants also appear to systematically map certain kinds of speech sounds to visual shapes with specific properties, in the same manner as adults and toddlers (the ‘bouba/kiki’ effect; Ozturk, Krehm, & Vouloumanos, 2013). Together, developmental data suggest that many foundational components of auditory cross-modal perception are present from early in life, and that extensive auditory experience is not always needed to link sounds to the events that caused them.

A case study: Hearing water temperature

Here we explore the origins of the ability to make subtle and nuanced inferences about events from sounds, focusing on an informative case study: Our ability to hear water temperature. Adults can judge the temperature of water simply from hearing the sound of it being poured into a cup (Velasco, Jones, King, & Spence, 2013). In spite of expecting to fail at this task, adults are surprisingly accurate, judging correctly whether water they heard poured was hot or cold with ~80% accuracy (Velasco et al., 2013). This ability is driven by temperature-related changes in the sound: For all sound clips, the only variable manipulated was the temperature of the water (see Figure 1), and manipulating relevant acoustic features was sufficient to make adults categorize the sounds as hot or cold water (Velasco et al., 2013).

How do such nuanced perceptual abilities develop over childhood? To our knowledge, no empirical studies have addressed the origins of nuanced cross-modal perception about everyday sounds. Developmental work has instead focused on perception of broader, more abstract aspects of cross-modal events, such as the distinction between the sound

of impact and lack of impact (Spelke, 1979). If experience plays a crucial role in our ability to link sound with specific materials and events, we would expect that the ability to hear temperature from sound would develop later in childhood than these more foundational components.

Here we report two experiments on the ability to hear water temperature, as a case study in the origins of nuanced auditory event perception. To provide a baseline measure of adult accuracy, we first test adults using the same two-alternative forced choice method we use with children in Exp. 2. We also explore individual differences in adulthood, asking whether there are individual differences in accuracy, and whether any differences may be explained by differing levels of experience with relevant sounds (Exp. 1). We then characterize the development of this ability in childhood, testing children 3-11 years of age (Exp. 2). If this aspect of cross-modal perception is innate, or requires only a few months of experience (e.g. Spelke, 1979), then children of all ages in this range should succeed at this task. If instead more extensive experience is needed, young children may fail to distinguish the sound of hot versus cold water, succeeding only later in childhood.

Stimuli, data, methods, and supplemental results for both experiments can be found in an OSF repository at the following link: <https://osf.io/j4ksu/>.

Experiment 1

In a first experiment, we tested adult accuracy in judging whether water is hot or cold from the sound of it being poured, using a two-alternative forced choice method, to provide a baseline with which to compare child performance. Participants listened to pairs of pouring sounds – one hot, one cold; and otherwise matched – and selected which sound was hot and which was cold water. In addition, participants completed a short questionnaire to measure the extent of their experience with relevant sounds (hot or cold liquids being poured), to explore how any individual differences in performance related to levels of experience with relevant sounds.

Stimuli, methods, and analyses were pre-registered and are accessible here: <http://asppredicted.org/blind.php?x=pp29ys>.

Method

Participants N=104 undergraduate students (Mean age = 21.3 yrs; SD = 2.32 yrs; 75 female) at a large public university in Southern California participated in exchange for course credit. Five participants were additionally tested but excluded based on criteria established a priori: Answering before or without listening to the audio stimuli (n=3); technical difficulties in presentation (n=2).

Stimuli All acoustic stimuli were professional recordings taken directly from previously published work, provided by the author (Velasco et al., 2013). Each sound was a pouring event five seconds long, of either hot or cold water being poured into a cup. Hot water was 180-183°F, and cold water was 43-46°F. The recordings involved pouring the same

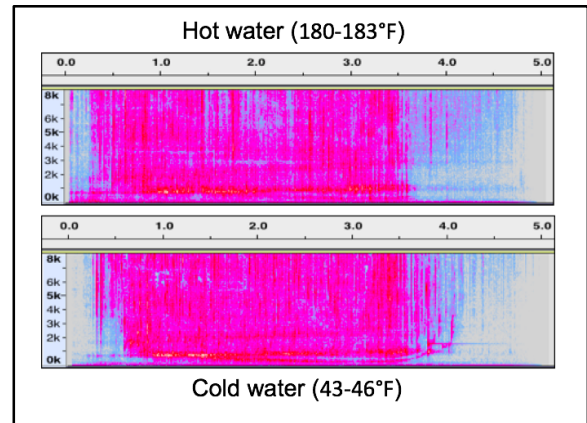


Figure 1: Spectrograms of Acoustic Stimuli. Spectrograms of hot versus cold water being poured from the same height at the same rate into the same cup. X-axis is time in seconds, y-axis is acoustic frequency in Hz.

amount of water (20 mL) from the same height (10 cm) at the same flow rate (40 mL/s).

One pair of sounds (hot, cold) was created for each of four different cups, each of a different material (paper, plastic, porcelain, glass), for a total of 8 stimulus items (4 pairs), presented across four unique trials corresponding to the four cup materials.

Design On each trial, participants heard two pouring sounds (one hot, one cold; matched for cup material), and were asked to judge which was hot and which was cold. Participants each completed 32 trials (8 sets of the 4 unique trials). The order of questions (whether “hot” or “cold” was queried first) was counterbalanced across participants; the order of trials and the order of sounds (whether hot or cold was played first) was randomized.

Procedure Participants completed the task in the lab in a web-browser on an iMac desktop computer, using high-quality headphones (Bose QuietComfort® 25 headphones) and via Qualtrics survey software.

Participants were instructed that they would be judging whether water being poured was hot or cold, from the sound of it being poured; and that every trial would include one hot and one cold sound. Before the task, participants judged whether they expected to be able to hear water temperature, on a 5-point confidence scale from -2 to +2 (‘Very confident that I cannot do this task’ to ‘Very confident that I can do this task’), and completed a brief sound check in which they identified two common sounds.

Each trial was presented on a separate page, and consisted of a prompt to listen to each of two sounds, followed by two questions: “Which of the sounds was hot water? (1, 2)” and “Which of the sounds was cold water? (1, 2)”. Participants could listen to the clips as many times as they liked before making their choices, and clicking to move on to the next trial.

After completing all 32 trials, participants then filled out a survey measuring their levels of experience with relevant sounds, based on previous established measures (Beverage Intake Questionnaire in Hedrick, Comber, Estabrooks, Savla, & Davy, 2010). Items were: Number of [hot/cold] drinks you hear being prepared/made per week; Number of [hot/cold] drinks you consume per week; Do you enjoy [hot/cold] beverages; Do you consider yourself to be someone who never/rarely drinks [hot/cold] drinks. Participants answered the above questions once for hot beverages and once for cold beverages. The study concluded with a brief demographics section, followed by a debriefing if requested. The study session lasted for 20 to 30 minutes on average.

Results

Adult accuracy at judging temperature from sound was extremely high: Participants were accurate on 93.0% of trials (29.8/32; $SD=2.78$), performing significantly above chance ($t(103)=107.4$, $p<0.001$, t -test vs. chance of 50%). Performance was heavily skewed, with 35% of participants accurate on 100% of trials, and 75% of participants correct on 90% of trials or more (see Figure 2). This high accuracy contrasted with participants' predictions before starting the task: 28% of participants were very or moderately confident that they would fail at the task, however these individuals (like the rest of the sample) had extremely high accuracy ($M=93.8\%$; $SD=1.41$).

To further characterize the data, we examined effects of cup material and trial order using logistic regression. The model predicted accuracy on each trial ('correct' vs. 'incorrect') with the predictors of cup material (paper, plastic, porcelain, glass), and both trial number (1 to 32) and subject as random effects. Nested model comparisons revealed there was a significant effect of cup material ($\chi^2(3)=112.5$, $p<0.001$), such that stimuli created with certain cup materials, particularly plastic cups, were more challenging than others (mean proportion correct: Plastic 85.6%; Paper 92.8%; Glass 96.5%; Porcelain 97.0%). There were individual differences in performance ($\chi^2(1)=135.5$, $p<0.001$), such that some participants were more accurate than others. There was no significant effect of trial number; that is, performance did not significantly change over the course of the session ($\chi^2(1)=3.08$, $p=0.079$).

Finally, we explored whether performance on this task differed based on how much experience participants had with relevant sounds. To do this, we binned participants into low- and typical-experience groups based on our binary measure: Whether participants identified as drinking hot or cold beverages rarely/never; or not. We found that the low-experience participants performed, numerically, slightly worse than the rest of the sample; however, this difference was not statistically significant; $t(98.71)=1.903$, $p=0.059$, two-tailed two sample t -test). Our power to detect such effects was low due to ceiling effects: Overall, both groups were successful at the task (Typical-experience group, $n=62$: Mean trials correct=30.3/32, $SD=2.42$ trials; Low-experience group, $n=42$: Mean trials correct= 29.3/32, $SD= 3.02$).

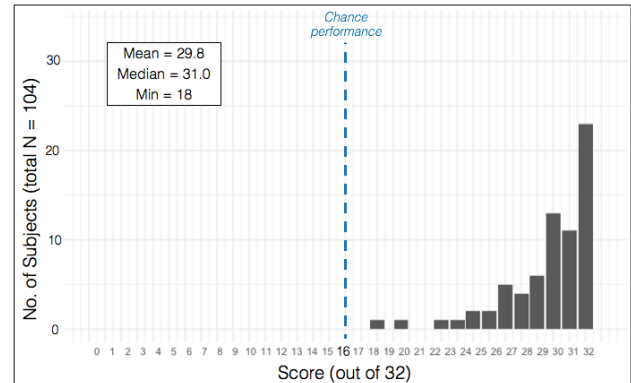


Figure 2: Results, adult participants. Adult participants were extremely accurate at judging the temperature of water from the sound of it being poured. Scores are out of 32 trials per participant; each trial was a two-alternative forced choice identifying which sound was hot vs. cold.

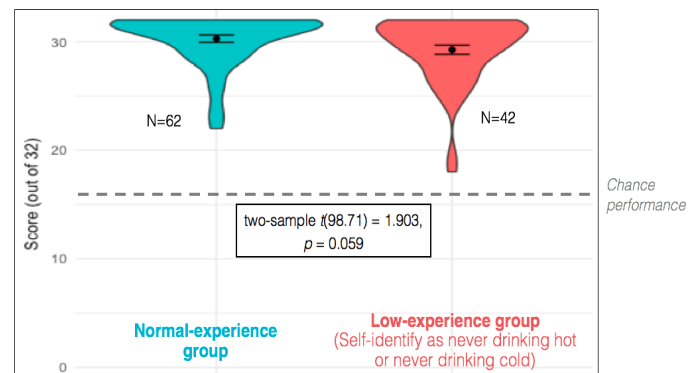


Figure 3: Comparing adults with normal vs. low levels of relevant auditory experience. Low-experience participants performed numerically though not significantly lower than the rest of the sample; ceiling effects limited our power to detect effects of experience.

Discussion

Exp. 1 confirms that adults can successfully judge the temperature of water from the sound of it being poured. This provides a baseline of adult performance to which we can compare children's performance, in Experiment 2. Performance in our task was substantially more accurate than the single prior study (Velasco et al., 2013), on which participants succeeded on an average of 72.5% of trials, vs. 93% here. This is as expected, due to methodological differences: Our task involved auditory discrimination (comparing two sounds) rather than outright identification (listening to one sound at a time). Our participants were also allowed to repeatedly listen to the stimuli before making their judgments, in contrast with prior work (Velasco et al., 2013). Both of these modifications were chosen intentionally, to

make the task easier for children and allow for future developmental comparison (in Exp. 2).

We also observed a significant main effect of cup material, which indicates that there were differences in people's ability to succeed at this task across our four stimuli. Plastic cups, in particular, made it more challenging to hear the difference between hot and cold water being poured. This difference in accuracy could result from differential everyday experience with various materials; for example, it is uncommon for people to pour hot water into a lightweight, disposable plastic cup (as used in this stimulus). However, this result may not generalize: Lower accuracy for these stimuli was not seen in previous work, which used the same stimuli with a slightly different method (Velasco et al., 2013). Thus, the effect of cup material on performance should be replicated and tested with new stimuli in future work.

Finally, we explored whether people's level of experience with relevant sounds predicted individual differences in accuracy. The answer to this question is less straightforward: Many people in our low experience group could nevertheless do the task very well, and overall the low experience group numerically performed lower but was not statistically different from the normal experience group. This could imply that experience with relevant sounds may not be required to perform these nuanced perceptual judgements. However, it is also plausible that all adults have substantial experience with relevant pouring sounds, even adults who do not themselves consume hot/cold beverages. In this case, individual differences in adults' experience may not be sufficient to find an effect of experience, should it truly exist. In Experiment 2, we therefore conducted a developmental study, as young children may have substantially less experience with relevant sounds.

Experiment 2

In a second experiment, we performed what to our knowledge is the first developmental study of nuanced auditory event perception. Taking a cross-sectional approach and testing children across a wide range of ages (3-11 years), we asked whether young children can hear water temperature, and whether there is a change in this ability over development. In doing so, we aimed to address the role of auditory experience from a different perspective – instead of asking adults how much experience they have with such sounds in everyday life (as in Exp. 1), here we consider age as a proxy for amount of relevant experience. If the ability appears in early childhood, this would suggest that extensive relevant auditory experience is not needed for this ability.

Children performed a two-alternative forced choice task in which they were asked to identify which of two sounds was hot vs. cold water. The task was designed to parallel the task given to adults in Exp. 1, with modifications to make the task developmentally appropriate (length of task, format of presentation). To ensure that all children in our final sample understood the task, we included pre-test questions to ensure that children understood the concepts of hot and cold (i.e.

could identify hot and cold scenes and objects), and could identify other sounds they heard (animal sounds).

Method

Participants N=113 children from the San Diego metro area participated, recruited from a database of local families interested in research, as well as from local preschools. Children ranged from age 3-11 years (3 years, 0 months - 11y, 7m; Mean age = 5y,10m; 46 female), and were tested by convenience within this range. The final sample included a large group of 4-5 year old children, and smaller samples of other ages (see Figure 5). An additional 23 participants were tested, but excluded according to a priori exclusion criteria: Answering one or more pre-test questions incorrectly (n=22); having participated previously (n=1).

Stimuli Acoustic stimuli (water pouring sounds) were the sounds of hot and cold water being poured into a glass, from Experiment 1. Stimuli for additional pre-test trials (to check task comprehension) consisted of two animal sounds (a cow mooing, a dog barking); cartoon images of hot and cold scenes (a sunny desert; a snowy scene); and cartoon images of hot and cold drinks (iced lemonade; a steaming mug).

Design Each participant completed one test trial, with the order of presentation of hot/cold sounds counterbalanced across participants. Participants also completed the three pre-test trials, during which the question order, order of sounds and position of images on pre-test trials was counterbalanced across participants. The study session took approximately five minutes.

Procedure Children were tested individually in the lab or at their preschool, in a quiet room seated at a child-size table across from an experimenter. To ensure that the experimenter remained blind to condition, sounds were presented via high quality headphones (MPOW 059 Bluetooth Headphones). The study was conducted on an iPad, with stimuli presented using Keynote slides, controlled by the experimenter via a remote control (with touchscreen disabled using Guided Access to prevent children from accidentally advancing the slides).

Pre-test trials: Three pre-test trials tested whether children understood the task and the concepts of hot and cold. Children were first asked to identify common animal sounds, using a similar procedure to the main task: The screen displayed two circles numbered 1 and 2; each corresponded to a sound by pulsing slightly as the sound played. Children listened to two animal sounds (cow mooing; dog barking) and were asked two questions: "Which one sounded like a [cow/dog], the first one or the second one?" Participants were then presented with two pictures of hot and cold scenes side by side, and asked: "Which of these pictures is somewhere [hot/cold]?" Lastly, participants were shown two pictures of hot and cold beverages, and asked "Which of these is a [cold/hot] drink?"



Figure 4: Developmental change in children’s ability to hear water temperature. Each dot represents one child’s response, which was either correct (on the top) or incorrect (on the bottom). The blue line indicates the probability of a correct response as a function of age (logistic regression; grey represents 95% confidence intervals).

Test trial: Children were instructed “Now in the next part, you’re going to hear two sounds. Each one is the sound of water being poured into a cup. But one of them is hot water, and one of them is cold water. I want you to listen and think about if it sounds like hot water, or cold water.” Children saw two circles, numbered 1 and 2, each corresponding to one sound (of hot or cold water being poured) by pulsing slightly as the sound played. After hearing both sounds, children were asked if they wanted to hear them again, and stimuli were replayed if requested. Children were then asked: “Which one sounded like [hot/cold] water, the first sound or the second sound?”

Results

To test for developmental change in children’s accuracy, we conducted a logistic regression, predicting accuracy based on age (Figure 4). Participants’ age significantly predicted their accuracy in identifying the sound of hot and cold water (Nested logistic model comparisons, Likelihood ratio test, $\chi^2(1) = 12.91, p < 0.001$). Notably, children aged 5 and under performed at chance, in spite of our testing a substantial sample of four and five year-olds (45.7% accuracy, 36/79 children, $p = 0.499$). In contrast, 85% of children age 6 and older answered correctly (29 of 34 children age 6+; $p < 0.001$; Figure 5).

Discussion

We find evidence of developmental change over childhood in the ability to hear water temperature: While children over 6 performed above chance in distinguishing the sound of hot versus cold water, children age 3-5 years performed at chance, in spite of our testing a substantial sample of young children. Crucially, this failure was not due to lack of understanding of the task: All children succeeded at using a

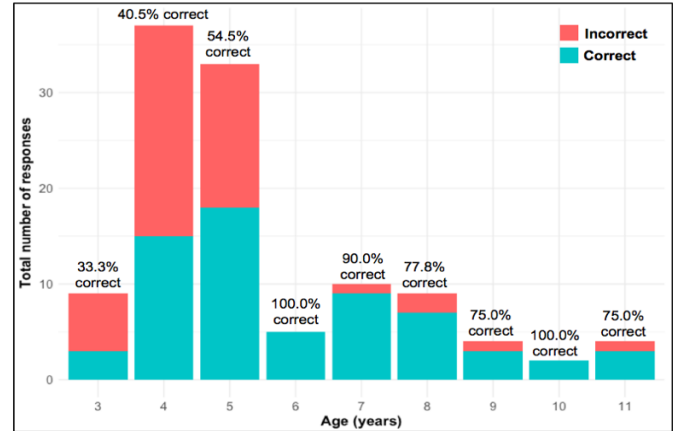


Figure 5: Accuracy and sample size by year of age. Children under the age of 6 performed at chance, while older children succeeded. The overall height of each bar shows the number of children tested in each age group; 4 and 5 year olds performed at chance in spite of a large sample of participants tested at these ages.

similar method to distinguish other sounds, and to distinguish and categorize other (visual) hot and cold stimuli.

General Discussion

Across two experiments, while adults and older children were able to judge water temperature from sound with high accuracy, preschool-aged children notably failed to do so. This failure was in spite of young children’s comprehension of the task, and the concepts of hot and cold. Overall these data suggest that perception of nuanced differences between auditory events is not part of early-developing auditory cognition, and instead develops over the first six years of life.

Developmental work has shown that even infants can make some broad links between sounds and the events that cause them (Spelke, 1979). The current work suggests, however, that sound-event mapping is not entirely present in early childhood, and that that more nuanced aspects of sound-event mapping develop in a protracted manner over childhood.

This developmental change may help explain differences in how people perceive sounds, including musical sounds. When perceiving musical sounds, people perceive meaning in part due to the relationship of musical sounds to the sounds of everyday events (Clarke, 2005). The current data suggest that young children and adults differ in our ability to link sounds to specific meaningful events. This represents a novel way in which auditory perception changes over developmental time, and an important potential source of differences in the music perception of adults and young children.

Why does the ability to hear water temperature appear only later in childhood, rather than earlier in life? Experience with relevant everyday sounds may play a crucial role in our ability to link sound with specific materials and events. By this account, young children may not have sufficient

exposure to relevant events to have mapped sounds to the somatosensory experience of hot and cold. Children may particularly lack acoustic and somatosensory experience with the pouring of hot water.

We believe that this experience-based account is the most parsimonious account of the developmental change we report. However, the developmental change may alternatively be explained as due to biological maturation. For example, maturation of the auditory system could potentially change sensitivity to the relevant aspects of the acoustic stimuli, making it easier for older children to hear the acoustic frequencies that contain the relevant information for distinguishing hot and cold pouring sounds.

A training study, which we plan in future work, may be able to tease apart these two accounts. If differences are due to experience and not maturation, then an intervention to provide preschool-age children with experience of relevant pouring events should be able to increase their accuracy, at young ages where they previously fail. The current data thus provide a foundation for answering broad questions about the role of innate factors and experience in the development of auditory perception.

Acknowledgments

We thank Viola Stormer and Tim Brady for helping to inspire this line of work.

References

- Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior and Development*, 6(4), 429–451.
- Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, 59, 197–209.
- Cabe, P. A., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 313.
- Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. New York, NY: Oxford University Press.
- Cusimano, M., Hewitt, L. B., Tenenbaum, J., McDermott, J. H. (2018). Auditory scene analysis as bayesian inference in sound source models. *Proceedings of the fortieth annual conference of the cognitive science society* (pp. 510-516). Austin, TX: Cognitive Science Society.
- Gaver, W. (1993a). How Do We Hear in the World? Explorations in Ecological Acoustics. *Ecological Psychology*, 5, 285–313.
- Gaver, W. W. (1993b). What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5(1), 1–29.
- Gerstenberg, T., Siegel, M. H., & Tenenbaum, J. (2018). What happened? Reconstructing the past through vision and sound. *Proceedings of the fortieth annual conference of the cognitive science society* (pp. 510-516). Austin, TX: Cognitive Science Society.
- Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.
- Gibson, J.J. (1977). The theory of affordances. In R. Shaw and J. Bransford (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. New York, NY: Routledge.
- Hedrick, V. E., Comber, D. L., Estabrooks, P. A., Savla, J., & Davy, B. M. (2010). The Beverage Intake Questionnaire: Determining Initial Validity and Reliability. *Journal of the American Dietetic Association*, 110(8), 1227–1232.
- Kopp, F. (2014). Audiovisual temporal fusion in 6-month-old infants. *Developmental Cognitive Neuroscience*, 9, 56–67.
- Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, 218(4577), 1138–1141.
- Lewkowicz, D. J. (1992). Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus. *Perception & Psychophysics*, 52(5), 519–528.
- Lewkowicz, D. J. (1994). Limitations on infants' response to rate-based auditory-visual relations. *Developmental Psychology*, 30(6), 880.
- Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, 282(5737), 403–404.
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, 114(2), 173–186.
- Perfecto, H., Donnelly, K., & Critcher, C. R. (2018). Volume Estimation Through Mental Simulation. *Psychological Science*, 30(1), 80–91.
- Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology*, 15(6), 626–636.
- Velasco, C., Jones, R., King, S., & Spence, C. (2013). The Sound of Temperature: What Information do Pouring Sounds Convey Concerning the Temperature of a Beverage. *Journal of Sensory Studies*, 28(5), 335–345.