Wired for sound? Children show auditory dominance when determining the emotions of others

Introduction

Recognising emotional cues is an essential facet of the development of social cognition in children, necessary for forming and maintaining relationships. Successfully labelling emotions has been linked to social competence (Izard et al, 2001; Leppanen and Hietanen, 2001), peer popularity (Edwards, Manstead, and McDonald, 1984; Leppanen and Hietanen, 2001), reduced behavioural problems (Izard et al, 2001), self-regulation (Neves et al, 2021), pro-social behaviour (Neves et al, 2021), and even academic achievement (Izard et al, 2001). As such, the ability to perceptually discriminate emotions, and in turn ascribe emotional states to others, is imperative for children's social functioning.

Emotion recognition is a function of multi-modal sensory input. Emotions are predominantly conveyed in the visual (facial expression, body language) and auditory (tone, pitch, volume) modalities, which are integrated to produce a unified percept. Congruency of information between these channels facilitates emotion recognition (Collignon et al, 2008; de Gelder and Vroomen, 2000; Kreifelts et al, 2007; Lambrecht, Kreifelts, and Wildgruber, 2013) with incongruent stimuli reducing both the accuracy of emotion classification (Collignon et al, 2008; Meeren, van Heijnsbergen, and de Gelder, 2005 ; Paulmann and Pell, 2011) and reaction time (Collignon et al, 2008; de Gelder and Vroomen, 2000; Meeren, van Heijnsbergen, and de Gelder, 2005). This is mirrored by neuroimaging studies (Kreifelts et al, 2007; Pourtois et al, 2005), showing enhanced activity in the bilateral posterior superior temporal gyrus (pSTG), middle temporal gyrus (MTG), and right thalamus during bimodal presentation of congruent emotional stimuli. Incongruence between visual and auditory information is not uncommon, however; for example, in the case of sarcasm, irony, insincerity, or the suppression of an emotional response, visual cues such as body language and facial expression will often contradict prosodic information, and judgement must be made on which cue to attend to. Moreover, these cues are transient, oftentimes ambiguous (Aviezer et al, 2008), and heavily dependent on context (Ngo and Isaacowitz, 2015).

So which cue do we pay attention to? The answer may be dictated by sensory dominance: a phenomenon where inputs from one sensory modality are preferentially processed over others

that are co-occurring, often in the context of incongruent information. Early work around sensory dominance by Colavita (1974) found visual stimuli to be ‘prepotent’ as compared to audio stimuli in a basic stimulus discrimination task – that is, visual stimuli seemed to be better attended. Participants were presented either a visual (light) or auditory (tone) stimulus, ostensibly unimodally, and instructed to report which stimuli they had perceived. In fact, a small proportion of the trials were bimodal, featuring synchronous audiovisual stimuli. Colavita engaged in deception to explain the presence of these unexpected bimodal presentations, demonstrating how they could occur from a supposedly unintentional error on part of the experimenter. Responses on the critical bimodal trials were of interest; participants often overlooked auditory stimuli, only responding to the presence of visual stimuli. Furthermore, on 16 of the 50 total bimodal trials, participants reported not registering the auditory stimulus at all. Even when the deception was dropped and participants were asked to press the tone key for bimodal stimuli, a strong bias towards the visual modality was noted. Not only did participants continue to press the visual key more often in conflict trials despite instructions to the contrary, Colavita also noted slowed reaction times, as participants would automatically reach for the light key before inhibiting this response to adhere to the task demands.

Subsequent research into the Colavita effect cemented its credibility across a range of conditions. Whilst it could be argued the original study’s results were ambiguous due to task demands - the large decrease in reaction times may have been due to interference with response selection, as the participants were forced to supress their responses to the visual component of the bimodal stimuli - later studies such as Koppen and Spence (2007a) corrected for this by allowing participants to report bimodal stimuli with a third key. A reduced, but still significant, visual dominance effect was found. The Colavita effect has been observed when haptic stimuli is used in place of audio (Hartcher-O'Brien, Levitan, and Spence, 2010; Hecht and Reiner, 2009), when the position of the stimuli is manipulated (Koppen and Spence, 2007b), when attentional bias to one modality is created (Koppen and Spence, 2007a; Sinnet, Spence, and Soto-Faraco, 2007), when the frequency of bimodal trials are altered (Koppen and Spence, 2007a; Sinnet et al, 2007) and when the target is changed from stimulus to repetition detection (Ngo et al, 2010). Thus, it remains robust across various experimental manipulations. Indeed, eliminating or reversing the Colavita effect in adults proves difficult, only achieved by significant alterations to the methodology to favour auditory stimuli (e.g. by using a temporal task, as in Sinnet et al, 2007), or by decreasing the

reliability of the visual stimuli (Alais and burr, 2004; Collignon et al, 2008; Heron, Whitaker, and McGraw, 2004; Moro and Steeves, 2012). One such study by Collignon et al (2008) found participants were initially biased towards categorising emotion using the visual component of a combined audio-visual affect expression, which could only be reversed by adding visual noise to decrease the reliability of the visual stimulus. Without such manipulations, it is clear that adults rely more heavily on visual information when perceiving bimodal stimuli.

Whilst adults preferentially attend visual stimuli, sensory dominance appears to be dynamic, changing over the course of typical human development. Research using children shows an early predisposition towards auditory stimuli – a reverse Colavita effect. Sloutsky and Napolitano (2003) found that after training participants to respond to compound audiovisual stimuli, 18-year-olds would rely on the visual component to make a same/difference judgement, whereas 4-year-olds exhibited reliance on auditory information. This was true both with a forced choice paradigm and a more open same/difference discrimination task. Critically, children were able to accurately discriminate visual stimuli when presented unimodally, showing preferential processing of auditory information, rather than difficulty with visual processing itself. A subsequent study by Robinson and Sloutsky (2004) showed this effect to be robust, replicating a strong auditory dominance in 4-year-olds whilst expanding these findings to infants, who also demonstrated the reverse Colavita effect. Furthermore, children were found to only encode information in their preferred modality when shown multimodal stimuli. After being trained to make predictions based on combined audiovisual cues, children could not accurately predict the location of a target based on their less preferred stimuli alone, whereas adults were able to. It may be, then, that processing and/or encoding of visual information is attenuated by the presence of auditory information.

Whilst a clear auditory dominance has been documented from infanthood to early childhood (Barnhart, Rivera, and Robinson, 2018; Havy, 2024; Lewkowicz, 1988; Nava and Pavani, 2013; Robinson and Sloutsky, 2004; Sloutsky and Napolitano, 2003; Wille and Ebersbach, 2016;), the point at which visual dominance develops is difficult to pinpoint. Nava and Pavani (2013) found, by using the original paradigm employed by Colavita (adjusted slightly for children), visual dominance begins to emerge around 9 years old, whilst auditory dominance persists in 6- to 7-year-olds. A similar timeline emerged when introducing semantically meaningful stimuli. Wille and Ebersbach (2016) tested 6-year-olds, 9-year-olds, and adults respectively, using semantically congruent (e.g. a picture of a dog paired with a

barking sound) or incongruent (e.g. a picture of a cow paired with a phone ringing) stimuli. Participants were instructed to press a key denoting whether they'd been presented with an auditory, visual, or audiovisual stimuli. 6 years old made significantly more auditory only judgements in bimodal trials, showing auditory dominance, whereas this was reversed in 9-year-olds and adults.

Studies using the other paradigms, such as the sound-induced flash illusion (Innes-Brown et al, 2011; Parker and Robinson, 2018; Sun et al, 2022), the oddball paradigm (Robinson, Ahmar, and Sloutsky, 2010; Robinson, Chandra, and Sinnet, 2016) and discrimination tasks (Dunifon, Rivera, and Robinson, 2016) have reported auditory dominance in older children and the young adult population. Innes-Brown et al (2011), using the sound induced flash illusion (whereby the number of beeps alters the perceived number of concurrently present flashes) reported an increased susceptibility to both fission (two beeps causes one flash to be perceived as two) and fusion (one beep causes two flashes to be perceived as one) illusions in children aged 8-17, as compared to adults. Conversely, Barnhart et al (2018) found mixed results in young adults, with most being deemed to use both auditory and visual cues on a same/difference task. It is important to note, however, that of the young adult participants who showed modality dominance, 60% were auditory responders.

Sensory dominance and emotional stimuli

The interplay between sensory dominance and affective cues in emotion recognition is an emerging area of interest. Emotion recognition seems to centre around facial and bodily expressions in adults. Collignon et al (2008) found adults participants preferentially attended visual stimuli to determine emotion when viewing dynamic clips of audiovisual expressions of emotion. Another study by Lambrecht, Kreifelts, and Wildgruber (2014) confirmed these findings; adult participants more accurately perceived emotion from facial expressions, versus vocal prosody. Furthermore, Van Den Stock, Righart, and De Gelder (2007) found whole body expressions to influence the perception of emotional voices in adults, even though participants were asked to exclusively judge the emotional quality of the voice.

Corresponding with the apparent timeline of sensory dominance using simpler stimuli, it appears children exhibit reliance on auditory information when recognising emotions up until school age. Brady, Ogren, and Johnson (2024) presented 2-year-old toddlers with two faces, alongside audio clips of sentences conveying conflicting affective and semantic content. The toddlers showed no preferential looking behaviour towards either face (one of which matched

the semantic content of the sentence, the other matching the emotional qualities of the sentence). This could be taken as auditory dominance; the infants attended audio, but not visual, stimuli.

Using a broader age range, Ross et al (2021) paired pictures of whole-body expressions with audio clips of emotionally meaningful stimuli such as laughter. In a series of tests, participants were instructed to determine the emotion of the person whilst ignoring either the body or the voice. Both younger (7 and under) and older (8 to 11-year old) children were unable to ignore what they heard, showing a preposition to select the emotion conveyed auditorily even when instructed to respond to visual stimulus. Using a similar paradigm, with emotional music clips in place of affect expressions, Ross et al (2023) found auditory dominance in both older and younger children, but not in adults.

Whilst these studies are indicative of a strong reliance on auditory cues demonstrated by children when determining the emotions of others, they are limited by their methodology. Both combined static pictures with audio clips, which, by their nature, are temporal. Multi-sensory integration depends on spatiotemporal binding of the stimuli; it must be feasible that they come from the same source, within roughly the same time-period, to be perceived as one event rather than several (Conrey and Pisoni, 2004; see Wallace and Stevenson, 2014, for review) Beneficial effects of multi-sensory integration of congruent information on reaction times and accuracy does not seem to appear when static images are used in the place of dynamic. Conversely, detrimental effects of incongruent multi-sensory information also aren't present. Schweinberger, Robertson, and Kaufmann (2007) found recognition of familiar voices to be faster and more accurate when voices were presented alongside either static or dynamic familiar faces, however this effect was significantly stronger when using dynamic faces than static. The researchers posited this was due to dynamic stimuli undergoing early multi-sensory integration, whereas static stimuli reflected a post-perceptual integration of information. Additionally, static incongruent faces did not introduce any significant cost to the identification of familiar voices, whereas dynamic incongruent faces disrupted identification of familiar voices. This may be because a mismatch in temporal features can alter one's ability to integrate stimuli.

If we consider the above, differences in studies such as Ross et al (2021) may be due to temporal factors rather than true sensory dominance. When combined with task demands that require inhibitory control, a characteristic not yet fully developed in young children

(Williams et al, 1999), it is difficult to dissociate auditory dominance from the development of executive function.

Here, we aim to resolve this problem. We used dynamic clips to present bodily and vocal emotion in a partial recreation of Ross et al (2021), and asked participants to name the emotion the person was feeling. We used four emotions: happiness, fear, sadness, and anger.

Methods

Participants

98 younger children, split into two categories, took part in this pilot study; thirty aged 7 and under (M = 5.83 years; 3 boys), and sixty-eight older children aged 8 – 11 (M = 9.22 years; 3 boys). Children were recruited from a junior scientist event held at Durham university. All participants had normal or corrected to normal vision and hearing. Informed consent was obtained from parents/guardians of child subjects. The study was approved by the psychology department ethics advisory subcommittee at Durham University.

Stimuli

Body stimuli were taken from a larger set created and validated by Kret et al (2011), consisting of sixty 2-second-long clips of actors depicting whole-body expressions, spanning six different emotional states. Each actor was dressed in black and filmed with a digital camera against a green background, with his or her face masked with a Gaussian filter to prevent the viewer from perceiving information from the face. Of these, we randomly selected forty: 10 happy, 10 fearful, 10 sad, and 10 angry clips. An even split of female and male actors were used.

Voice stimuli were taken from the Montreal Affective Voices database (Belin, Fillion-Bilodeau, & Gosselin, 2008), consisting of sixty short audio clips containing non-speech affective vocalisations, spanning nine different emotional states. Recordings were taken using a UMT800 condenser microphone, then edited to between 1.45 and 2.23s in length. In line with the body stimuli, we randomly selected 10 of each for the four basic emotions we tested.

Design and procedure

Participants took part in four experimental blocks. In blocks 1 and 2, participants were presented with unimodal stimuli (body alone and voice alone, respectively). 12 clips of each

modality were shown (3 per emotion). After viewing each stimulus, participants were asked "how is this person feeling?", with a choice of all four emotions per clip.

In block 3, participants were presented with combined audiovisual stimuli, with each modality depicting either anger or sadness. This gave the combinations of 5 Angry Body/Sad Voice, 5 Angry Body/Angry Voice, 5 Sad Body/Angry Voice, and 5 Sad Body/Sad Voice stimuli. Preceding each block, an instruction page was shown on screen, until participants pressed space to indicate they were ready. After viewing each stimulus, participants were asked "How is this person feeling?", with a forced choice of either angry (A Key) or sad (S Key).

In block 4, participants were presented with audiovisual stimuli depicting happiness or fearfulness. This gave the combinations of 5 Happy Body/Fearful Voice, 5 Happy Body/Happy Voice, 5 Fearful Body/Happy Voice, and 5 Fearful Body/Fearful Voice. Again, participants were asked "How is this person feeling?", with a forced choice of happy (H Key) or fearful (F Key).

A total of 64 trials took place across the two blocks, with testing taking approximately 10 minutes. Block order was randomised across participants, as were the order of clips displayed within each block.

Results

Body and Voice Alone

For block 1 and 2 (body alone and voice alone respectively), percentage of correct emotion identification was calculated and averaged across participants to give an overall accuracy score for each modality. Mean percentage accuracies were entered into a 4 (Emotion) × 2 (Modality) × 2 (Age Group) repeated analysis of variance (ANOVA)

We found a significant main effect of age group, $F(1, 96) = 6.516$, $p = .012$, $\eta p2 = .065$, driven by better performance in older children ($M = 77.79$, $SD = 9.31$) as compared to younger ($M = 71.83$, $SD = 11.18$) - ($M_{diff} = 5.959$, $SE = 2.334$). Age differences in the recognition of sad voices, fearful voices, and happy bodies were significant (sad voices: $p = .049$; fearful voices: $p = .023$; happy bodies: $p = .047$. A main effect of modality was also found, $F(1, 96) = 46.458$, $p < .001$, $\eta p2 = .326$. Emotion was better recognised in voices (M =

80.40, SD = 13.18) than in bodies (M = 66.92, SD = 15.19). There was no age group * modality interaction, F(1,96) = 0.036, p = .851, however, indicating voices were recognised better across groups. In fact, age did not interact with any other factor.

We found a significant effect of emotion, F(3, 288) = 3.368, p = .019. Post hoc Bonferroni-corrected t tests indicated this stemmed from better recognition rates of happiness as compared to fear. An emotion * modality interaction was also observed, F(3,288) = 65.505, p < .001, ηp2 = .406. Notably, participants were better at recognising happy and fearful voices than the bodies associated with their incongruent pairings (p < .001, p = .003). This did not hold true for anger and sadness.

Emotion – Modality combinations

For block 3 and 4, percentage of auditory based responding was calculated for each participant and averaged across all participants, then the incongruent trials entered into a 4 (Emotion–Modality Combination) × 2 (Age Group) ANOVA. We found a main effect of age groups, F(1,96) = 6.938, p = .010, ηp2 = .067, with younger children (M = 79.88, SD = 13.98) responding to the auditory component more often than older children (M = 70, SD = 22.74) An independent t-test conducted on incongruent trials revealed a significant difference between older and younger children for Happy Body / Fearful voice, T(96) = -2.48, p = .015 and Fearful Body / Happy Voice, T(96) = -3.44, p < .001, with younger participants showing higher levels of auditory dominance. In fact, older children failed to score significantly differently from chance on the Happy Body/Fearful Voice condition, T(13) = .62, p = 0.54.

We found a main effect of Emotion-Modality combo, F(2.43,233.699) = 28.332, p < .001, ηp2 = .228. This was driven by increased auditory dominance rates in the Angry Body / Sad Voice and Fearful Body / Happy voice, which were not significantly different from each other, but were significantly different from Sad Body / Angry voice and Happy body / Fearful voice. There was also an interaction of Emotion-Modality combo * age, F(2.43,233.699) = 3.064, p = .039, ηp2 = .031.

Discussion

Our research had two main goals: (a) to see if the reported auditory dominance in children would extend into dynamic emotional stimuli, and (b) to explore any differences in sensory dominance between younger and older children.

We found a distinct tendency for both younger and older children to rely on auditory information when making a judgement on how someone is feeling. Both groups were significantly more likely than chance to respond to the auditory component of an emotionally informative stimulus over the visual. From this, we can infer that auditory dominance in children does indeed extend to dynamic emotional stimuli. It also persists in older children, albeit to a lesser extent. Our findings are in line with Ross et al (2021), however, who, using the same age groups, found children responded to auditory components of bimodal emotional stimuli, even when instructed to ignore the voice. Similarly, Ross et al (2023) showed auditory dominance in older children.

Whilst this somewhat contradicts Nava and Pavani's (2013) timeline, we did find support for auditory dominance diminishing as children age. These differences could also be explained by the nature of the tasks used; Nava and Pavani simply tested whether children perceived the stimulus, whereas we added an additional component; children had to make judgement based on the stimulus. This could have introduced an additional element of response competition, one of the potential mechanisms of auditory dominance; children may have perceived both auditory and visual information but have difficulty inhibiting auditory responses (Robinson and Sloutsky, 2019). If this is the mechanism behind auditory dominance, it makes sense that our study yielded increased auditory dominance rates even in older children than Nava and Pavani (2013); we did not give children the option to report both emotions, whereas Nava and Pavani (2013) included the option to report bimodal stimuli.

Another area where our findings diverge from these studies is the difference in accuracy between modalities. Is it possible that here, given the higher recognition rates of voices, children could be relying on their most accurate modality? This seems unlikely. Performance in the unimodal conditions did not predict levels of auditory dominance in the bimodal trials. Older children are significantly better at recognising sad voices than sad bodies in isolation, but the recognition rate of angry bodies did not differ between age groups. If auditory dominance is caused by accuracy, then older children should show significantly higher auditory dominance in the Angry body / Sad voice condition than younger children, however only the Happy body / Fearful voice and Fearful body / Happy voice conditions yielded a significant difference between age groups. Additionally, children showed high auditory dominance in the Angry Body / Sad Voice condition, but angry bodies and sad voices had similar recognition rates in isolation.

Furthermore, the literature shows children tend to be less accurate when determining emotion from voice than body. Using a similar paradigm, Ross et al (2021) and Ross et al (2023) found no differences in modality across age groups. Most, Bachar, and Dromi (2012) found children to be less accurate in categorising emotion from prosody than body. Poignantly, the emotions they used to test this were the four included in this study. Another study by Nelson and Rusell (2011) tested preschoolers' recognition of emotions from the voice, face, and posture. Sadness from the voice was found to be identified with high accuracy rates (mirroring our results), however identification of the other three emotions through voice was very poor as compared to visual cues. From this, we must conclude that accuracy has no bearing on auditory dominance.

A similar argument by Hirst, Cragg, and Allen (2018) proposed a developmental explanation of auditory dominance. They argued auditory dominance in children may be the result of differing trajectories of maturation in the auditory and visual domains, with the auditory system deemed more reliable at younger ages. Upon consideration, we do not believe this to be the driving factor behind our results. Devous et al (2006) found the Heschl Gyrus (HG), which houses the Primary Auditory Cortex (PAC) and is involved in processing auditory information, to have fully matured by the age of 7. Temporal regions associated with higher auditory abilities, on the other hand, continue to develop throughout school age (Bishop et al, 2011; Devous et al, 2006), with some, such as distinguishing speech in noise, continuing to improve into the late teenage years (Talarico et al, 2006). When inferring emotions from voice alone, children reach adult-like rates at 14-15 years old (Grosbras, Ross, and Belin, 2018; Morningstar, Feldman, and Dirks, 2018).

In contrast, the visual cortex develops rapidly, with most anatomical features becoming adult like within the first two years of life (Siu and Murphy, 2018). Correspondingly, visual features such as binocular fusion and stereopsis are adult-like even in infancy (Birch and Petrig, 1996). Visual development extends throughout childhood, with adult like visual acuity, contract sensitivity, and visual search shown at age 8-9 (Almoqbel, Irving, and Leat, 2017; Dye and Bavelier, 2010; Ellemburg et al, 1999). At around this age, children are also adult like in recognising bodily emotions in full-light displays (Ross, Polton, and Grosbras, 2012). Given the protracted development of auditory abilities as compared to visual, we do not believe that younger children showed increased auditory dominance due to a relative advantage of the auditory system over the visual; in fact, it was older children who were significantly more accurate at detecting emotion from voice than body, whereas in younger

children this difference was not significant. As our data shows no difference in emotion recognition between age groups when perceiving bodies, with both showing high accuracy, we also do not believe developmental differences between our two age groups affected the ability to reliably detect emotion from the body.

The question then becomes why might children use potentially less accurate channels of information? Multiple theories seek to explain the processes behind sensory dominance. The modality appropriate hypothesis explains sensory dominance in terms of each modality's suitability for processing different types of information, with the visual modality being superior for spatial information, and the auditory modality more suitable for temporal information (Welch and Warren, 1980). If we interpret our results through this lens, the dynamic nature of the stimuli introduced an aspect of temporality, making audition the superior modality to process this event. Welch and Warren (1980) expand upon this model by introducing the concept of unity: the supposition that information from more than one modality is linked together, resulting in the perception of this information as a single event. The strength of the perceived unity of an event is thought to influence the magnitude of sensory dominance. Certainly, there is evidence for this; Koppen and Spence (2007b,c) found both spatial and temporal discrepancies between modalities modulated the Colavita effect, although they did not succeed in reversing the effect. Welch and Warren (1980) only supposed that unity predicted the magnitude, not the mode, of sensory dominance, however there is evidence that a lack of integration does result in auditory dominance in children. We know that audio-visual integration is immature in younger children, only reaching adult like levels at around 15 (Brandwein et al, 2011). Moreover, Schorr et al (2005) found that amongst normal hearing children who failed to integrate audio-visual information in the McGurk effect, 80% of them reported the auditory stimulus. Perhaps our results reflect the failure of younger children to properly integrate multi-sensory signals.

Limitations and implications

As mentioned, our experiment used forced choice when asking children to make an emotional judgement, which prevents us from measuring integration levels. It is possible that children recognised the presence of both the auditory and bodily emotions, but the binary nature of our study prevents us from testing this. Kestenbaum and Gelman (1995) show children as young as four can recognise blended emotions. Likewise, Van den stock, Righart, and de gelder (2007) show emotion expressions of the face, body and voice to bias each other. It may be

that our choice to exclude an option to acknowledge bimodal stimuli artificially inflated the rates of auditory dominance. Future studies could include an option for both emotions to delineate auditory dominance caused by the auditory modality interfering with processing of the visual modality from auditory dominance caused by response competition. Alternatively, participants could be allowed to include more than one emotion in their response, potentially with an accompanying rating of intensity, to measure the extent to which the signals mix.

Our study has significant implications for clinical populations. Establishing a timeline of sensory dominance gives us a baseline from which we can measure differences. Li et al (2024) showed children with ADHD exhibit a visual dominance in response selection, whereas neurotypical children do not. Conversely the ASD population seems to exhibit auditory dominance for longer, which may even persist into adulthood (Moro, Ghemraoui, and Steeves, 2012). If typically developing children begin to transition from auditory to visual dominance later in childhood, as our results suggest, discrepancies in the trajectory of sensory dominance can be used as early screening tools.

Additionally, our results have implications for improving educational approaches. If auditory dominance is a stable feature of early childhood, adopting auditory based teaching styles would possibly be more effective. Conversely, if children automatically attend auditory information, then background noise may act as a distractor, impairing children's ability to process visual information. Background noise should therefore be kept to a minimum.

Conclusion

Both younger and older children show auditory dominance when determining the emotions of others. When asked to make a judgement on a bimodal stimulus, children overwhelmingly responded based on the auditory component of the stimulus. Auditory dominance diminishes with age, however it is still present in children up to 11 years, and may persist beyond this. These findings have important implications for education and early screening of neurodivergences.

References

- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. Current Biology, 14(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029
- Almoqbel, F. M., Irving, E. L., & Leat, S. J. (2017). Visual Acuity and Contrast Sensitivity Development in Children: Sweep Visually Evoked Potential and Psychophysics. Optometry and Vision Science, 94(8), 830. https://doi.org/10.1097/OPX.0000000000001101
- Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & Bentin, S. (2008). Angry, Disgusted, or Afraid?: Studies on the Malleability of Emotion Perception. Psychological Science, 19(7), 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x
- Barnhart, W. R., Rivera, S., & Robinson, C. W. (2018). Different patterns of modality dominance across development. Acta Psychologica, 182, 154–165. https://doi.org/10.1016/j.actpsy.2017.11.017
- Birch, E., & Petrig, B. (1996). FPL and VEP measures of fusion, stereopsis and stereoacuity in normal infants. Vision Research, 36(9), 1321–1327. https://doi.org/10.1016/0042-6989(95)00183-2
- Bishop, D. V. M., Anderson, M., Reid, C., & Fox, A. M. (2011). Auditory Development between 7 and 11 Years: An Event-Related Potential (ERP) Study. PLOS ONE, 6(5), e18993. https://doi.org/10.1371/journal.pone.0018993
- Brady, S. M., Ogren, M., & Johnson, S. P. (2024). Effects of conflicting emotional cues on toddlers' emotion perception. British Journal of Developmental Psychology, 42(3), 376–391. https://doi.org/10.1111/bjdp.12501
- Brandwein, A. B., Foxe, J. J., Russo, N. N., Altschuler, T. S., Gomes, H., & Molholm, S. (2011). The Development of Audiovisual Multisensory Integration Across Childhood and Early Adolescence: A High-Density Electrical Mapping Study. Cerebral Cortex, 21(5), 1042–1055. https://doi.org/10.1093/cercor/bhq170
- Colavita, F. B. (1974). Human sensory dominance. Perception & Psychophysics, 16(2), 409–412. https://doi.org/10.3758/BF03203962

- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. Brain Research, 1242, 126–135. https://doi.org/10.1016/j.brainres.2008.04.023
- Conrey, B. L., & Pisoni, D. B. (2004). Detection of Auditory-Visual Asynchrony in Speech and Nonspeech Signals. Research on Spoken Language Processing, 26, 71–94.
- de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. Cognition and Emotion, 14(3), 289–311. https://doi.org/10.1080/026999300378824
- Devous Michael D., Altuna Dianne, Furl Nicholas, Cooper William, Gabbert Gretchen, Ngai Wei Tat, Chiu Stephanie, Scott Jack M., Harris Thomas S., Payne J. Kelly, & Tobey Emily A. (2006). Maturation of Speech and Language Functional Neuroanatomy in Pediatric Normal Controls. Journal of Speech, Language, and Hearing Research, 49(4), 856–866. https://doi.org/10.1044/1092-4388(2006/061)
- Dunifon, C. M., Rivera, S., & Robinson, C. W. (2016). Auditory stimuli automatically grab attention: Evidence from eye tracking and attentional manipulations. Journal of Experimental Psychology: Human Perception and Performance, 42(12), 1947–1958. https://doi.org/10.1037/xhp0000276
- Dye, M. W. G., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. Vision Research, 50(4), 452–459. https://doi.org/10.1016/j.visres.2009.10.010
- Edwards, R., Manstead, A. S. R., & Macdonald, C. J. (1984). The relationship between children's sociometric status and ability to recognize facial expressions of emotion. European Journal of Social Psychology, 14(2), 235–238. https://doi.org/10.1002/ejsp.2420140212
- Ellemberg, D., Lewis, T. L., Hong Liu, C., & Maurer, D. (1999). Development of spatial and temporal vision during childhood. Vision Research, 39(14), 2325–2333. https://doi.org/10.1016/S0042-6989(98)00280-6
- Grosbras, M.-H., Ross, P. D., & Belin, P. (2018). Categorical emotion recognition from voice improves during childhood and adolescence. Scientific Reports, 8(1), 14791. https://doi.org/10.1038/s41598-018-32868-3
- Hartcher-O'Brien, J., Levitan, C., & Spence, C. (2010). Extending visual dominance over touch for input off the body. Brain Research, 1362, 48–55. https://doi.org/10.1016/j.brainres.2010.09.036

- Havy, M. (2024). In the ear of the viewer: Toward an auditory dominance in early word learning. First Language, 44(2), 127–152. https://doi.org/10.1177/01427237231217249
- Hecht, D., & Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. Experimental Brain Research, 193(2), 307–314. https://doi.org/10.1007/s00221-008-1626-z
- Heron, J., Whitaker, D., & McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction. Vision Research, 44(25), 2875–2884. https://doi.org/10.1016/j.visres.2004.07.001
- Hirst, R. J., Cragg, L., & Allen, H. A. (2018). Vision dominates audition in adults but not children: A meta-analysis of the Colavita effect. Neuroscience & Biobehavioral Reviews, 94, 286–301. https://doi.org/10.1016/j.neubiorev.2018.07.012
- Innes-Brown, H., Barutchu, A., Shivdasani, M. N., Crewther, D. P., Grayden, D. B., & Paolini, A. G. (2011). Susceptibility to the flash-beep illusion is increased in children compared to adults. Developmental Science, 14(5), 1089–1099. https://doi.org/10.1111/j.1467-7687.2011.01059.x
- Izard, C., Fine, S., Schultz, D., Mostow, A., Ackerman, B., & Youngstrom, E. (2001). Emotion Knowledge as a Predictor of Social Behavior and Academic Competence in Children at Risk. Psychological Science. https://journals.sagepub.com/doi/epdf/10.1111/1467-9280.00304
- Kestenbaum, R., & Gelman, S. A. (1995). Preschool children's identification and understanding of mixed emotions. Cognitive Development, 10(3), 443–458. https://doi.org/10.1016/0885-2014(95)90006-3
- Koppen, C., & Spence, C. (2007a). Seeing the light: Exploring the Colavita visual dominance effect. Experimental Brain Research, 180(4), 737–754. https://doi.org/10.1007/s00221-007-0894-3
- Koppen, C., & Spence, C. (2007b). Audiovisual asynchrony modulates the Colavita visual dominance effect. Brain Research, 1186, 224–232. https://doi.org/10.1016/j.brainres.2007.09.076
- Koppen, C., & Spence, C. (2007c). Spatial coincidence modulates the Colavita visual dominance effect—ScienceDirect. https://www-sciencedirect-com.ezphost.dur.ac.uk/science/article/pii/S0304394006011566

- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. NeuroImage, 37(4), 1445–1456. https://doi.org/10.1016/j.neuroimage.2007.06.020
- Kret, M. E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. NeuroImage, 54(2), 1755–1762. https://doi.org/10.1016/j.neuroimage.2010.08.012
- Lambrecht, L., Kreifelts, B., & Wildgruber, D. (2014). Gender differences in emotion recognition: Impact of sensory modality and emotional category. Cognition and Emotion, 28(3), 452–469. https://doi.org/10.1080/02699931.2013.837378
- Leppänen, J. M., & Hietanen, J. K. (2001). Emotion recognition and social adjustment in school–aged girls and boys. Scandinavian Journal of Psychology, 42(5), 429–435. https://doi.org/10.1111/1467-9450.00255
- Lewkowicz, D. J. (1988). Sensory dominance in infants: I. Six-month-old infants' response to auditory-visual compounds. Developmental Psychology, 24(2), 155–171. https://doi.org/10.1037/0012-1649.24.2.155
- Li, X., Cai, S., Chen, Y., Tian, X., & Wang, A. (2024). Enhancement of visual dominance effects at the response level in children with attention-deficit/hyperactivity disorder—ScienceDirect. https://www-sciencedirect-com.ezphost.dur.ac.uk/science/article/pii/S0022096524000377
- Meeren, H. K. M., van Heijnsbergen, C. C. R. J., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. Proceedings of the National Academy of Sciences, 102(45), 16518–16523. https://doi.org/10.1073/pnas.0507650102
- Morningstar, M., Ly, V. Y., Feldman, L., & Dirks, M. A. (2018). Mid-Adolescents' and Adults' Recognition of Vocal Cues of Emotion and Social Intent: Differences by Expression and Speaker Age. Journal of Nonverbal Behavior, 42(2), 237–251. https://doi.org/10.1007/s10919-018-0274-7
- Moro, S. S., Ghemraoui, A. A., & Steeves, J. K. E. (2012). No Colavita effect: Lack of visual dominance in people with autism spectrum disorder. Journal of Vision, 12(9), 1038. https://doi.org/10.1167/12.9.1038
- Moro, S. S., & Steeves, J. K. E. (2012). No Colavita effect: Equal auditory and visual processing in people with one eye. Experimental Brain Research, 216(3), 367–373. https://doi.org/10.1007/s00221-011-2940-4

- Most, T., Bachar, D., & Dromi, E. (2012). Auditory, Visual, and Auditory-Visual Identification of Emotions by Nursery School Children. Journal of Speech-Language Pathology & Applied Behavior Analysis, 5.
- Nava, E., & Pavani, F. (2013). Changes in Sensory Dominance During Childhood: Converging Evidence From the Colavita Effect and the Sound-Induced Flash Illusion. Child Development, 84(2), 604–616. https://doi.org/10.1111/j.1467-8624.2012.01856.x
- Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. Journal of Experimental Child Psychology, 110(1), 52–61. https://doi.org/10.1016/j.jecp.2011.03.014
- Neves, L., Martins, M., Correia, A. I., Castro, S. L., & Lima, C. F. (2021). Associations between vocal emotion recognition and socio-emotional adjustment in children. Royal Society Open Science, 8(11), 211412. https://doi.org/10.1098/rsos.211412
- Ngo, M. K., Sinnett, S., Soto-Faraco, S., & Spence, C. (2010). Repetition blindness and the Colavita effect. Neuroscience Letters, 480(3), 186–190. https://doi.org/10.1016/j.neulet.2010.06.028
- Ngo, N., & Isaacowitz, D. M. (2015). Use of context in emotion perception: The role of top-down control, cue type, and perceiver's age. Emotion, 15(3), 292–302. https://doi.org/10.1037/emo0000062
- Parker, J. L., & Robinson, C. W. (2018). Changes in multisensory integration across the life span. Psychology and Aging, 33(3), 545–558. https://doi.org/10.1037/pag0000244
- Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? Motivation and Emotion, 35(2), 192–201. https://doi.org/10.1007/s11031-011-9206-0
- Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of Facial Expressions and Voices and of their Combination in the Human Brain. Cortex, 41(1), 49–59. https://doi.org/10.1016/S0010-9452(08)70177-1
- Robinson, C. W., Ahmar, N., & Sloutsky, V. M. (2010). Evidence for auditory dominance in a passive oddball task. In S. Ohlsson & R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society (pp. 2644--2649). Cognitive Science Society.

- Robinson, C. W., Chandra, M., & Sinnett, S. (2016). Existence of competing modality dominances. Attention, Perception, & Psychophysics, 78(4), 1104–1114. https://doi.org/10.3758/s13414-016-1061-3
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory Dominance and Its Change in the Course of Development. Child Development, 75(5), 1387–1401. https://doi.org/10.1111/j.1467-8624.2004.00747.x
- Robinson, C. W., & Sloutsky, V. M. (2019). Two mechanisms underlying auditory dominance: Overshadowing and response competition. Journal of Experimental Child Psychology, 178, 317–340. https://doi.org/10.1016/j.jecp.2018.10.001
- Ross, P., Atkins, B., Allison, L., Simpson, H., Duffell, C., Williams, M., & Ermolina, O. (2021). Children cannot ignore what they hear: Incongruent emotional information leads to an auditory dominance in children. Journal of Experimental Child Psychology, 204, 105068. https://doi.org/10.1016/j.jecp.2020.105068
- Ross, P., Williams, E., Herbert, G., Manning, L., & Lee, B. (2023). Turn that music down! Affective musical bursts cause an auditory dominance in children recognizing bodily emotions. Journal of Experimental Child Psychology, 230, 105632. https://doi.org/10.1016/j.jecp.2023.105632
- Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. Proceedings of the National Academy of Sciences, 102(51), 18748–18750. https://doi.org/10.1073/pnas.0508862102
- Schweinberger, S. R., Robertson, D., & Kaufmann, J. M. (2007). Hearing Facial Identities. Quarterly Journal of Experimental Psychology, 60(10), 1446–1456. https://doi.org/10.1080/17470210601063589
- Sinnett, S., Spence, C., & Soto-Faraco, S. (2007). Visual dominance and attention: The Colavita effect revisited. Perception & Psychophysics, 69(5), 673–686. https://doi.org/10.3758/BF03193770
- Siu, C. R., & Murphy, K. M. (2018). The development of human visual cortex and clinical implications. Eye and Brain, 10, 25. https://doi.org/10.2147/EB.S130893
- Sloutsky, V. M., & Napolitano, A. C. (2003). Is a Picture Worth a Thousand Words? Preference for Auditory Modality in Young Children. Child Development, 74(3), 822–833. https://doi.org/10.1111/1467-8624.00570

- Sun, Y., Zhou, H., Liu, C., Wang, A., Yue, C., & Zhang, M. (2022). Sound-induced flash illusion is modulated by the depth of auditory stimuli: Evidence from younger and older adults. Attention, Perception and Psychophysics, 84(6), 2040–2050. https://doi.org/10.3758/s13414-022-02537-9
- Talarico, M., Abdilla, G., Aliferis, M., Balazic, I., Giaprakis, I., Stefanakis, T., Foenander, K., Grayden, D. B., & Paolini, A. G. (2006). Effect of Age and Cognition on Childhood Speech in Noise Perception Abilities. Audiology and Neurotology, 12(1), 13–19. https://doi.org/10.1159/000096153
- Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. Emotion, 7(3), 487–494. https://doi.org/10.1037/1528-3542.7.3.487
- Wallace, M. T., & Stevenson, R. A. (2014). The construct of the multisensory temporal binding window and its dysregulation in developmental disabilities. Neuropsychologia, 64, 105–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.005
- Wille, C., & Ebersbach, M. (2016). Semantic congruency and the (reversed) Colavita effect in children and adults. Journal of Experimental Child Psychology, 141, 23–33. https://doi.org/10.1016/j.jecp.2015.07.015
- Williams, B. R., Ponesse, J. S., Schachar, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. Developmental Psychology, 35(1), 205–213. https://doi.org/10.1037/0012-1649.35.1.205