The functional significance of cross-sensory correspondences in infant-directed speech

Anna Madeleine Barnett, B.Sc.

A thesis submitted to Lancaster University in

partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Lancaster University

December 2019

Declaration

I declare that this thesis is my own work and that it has not been submitted in substantially the same form for the award of a higher degree elsewhere.

Signed: ABarnett

Date: 18th December 2019

Abstract

Evidence suggesting that infants appreciate a range of cross-sensory correspondences is growing rapidly (see Dolscheid, Hunnius, Casasanto & Majid, 2014; Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004; Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson, 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018), and yet there is no known attempt to establish the functional significance of these correspondences in infancy. Research shows that speakers manipulate their prosody (i.e. melody of spoken language) to communicate the meaning of unfamiliar words and do so in ways that exploit the cross-sensory correspondences between, for example, pitch and size (Nygaard, Herold & Namy, 2009) and pitch and height (Shintel, Nusbaum & Okrent, 2006). But do infants attend to a speaker's prosody in this context to interpret the meaning of unfamiliar words? The aim of this thesis is to further establish how infant-directed speakers use prosody to communicate the cross-sensory meanings of words and, for the first time, identify whether infants capitalise on their sensitivity to cross-sensory correspondences to resolve linguistic uncertainty. In Experiment 1 – 4 we identify a list of novel pseudowords to use in all experiments being reported. These pseudowords were judged by participants as being *neutral* in terms of their sound-symbolic potential, allowing us to rule out the impact of sound-symbolism in our investigation. Experiment 5 provides support for earlier studies revealing cross-sensory correspondences in infant-directed speech. When presented with pseudowords spoken in a prosodically meaningful way, 13-month-old infants demonstrated a preference for objects that were contradictory to the cross-sensory acoustic properties of speech (e.g. lower-pitch voice with higher objects) (Experiment 6), and adults failed to match pseudowords with objects based on the prosodic information that was provided (Experiment 7). However, Experiment 8 provides evidence that 24-month-olds match pseudowords spoken in a higher-pitch voice, and at a faster rate, with objects that are visually higher in space. The implications of these findings are discussed, with suggestions as to how they can be usefully extended.

For Grandad. The first academic in my life.

I wish you could be here to share in this.

Sock it to ‘em!

Acknowledgements

I will begin by saying an enormous thank you to my supervisors, Gavin Bremner and Peter Walker. Gavin, you saw something in me when I was just a fresh-faced, enthusiastic undergraduate and have journeyed with me ever since. If it had not been for you and your “I wondered if you had ever thought of doing a PhD?” I would not have had the confidence to pursue this. Peter, the endless hours we have spent putting the world to rights has finally paid off. Thank you for sharing your wisdom and for teaching me how to be a proper researcher.

I extend a huge thank you to the Leverhulme Trust. Your support has allowed me to explore what I am truly capable of (and visit a lot of cool places along the way).

I am incredibly fortunate to be surrounded by the most loving and supportive family and friends. Mum, you are my greatest role model (“My mum’s a head teacher, you know.”) and the first best friend I ever had. Pops, you have been cheering me on for as long as I can remember. Thank you for reminding me to always be myself (I’ll never stop pulling faces in the mirror). Thank you to Sophie for forcing me to put on my running shoes and to Maisie for injecting some magic into my life!

Jonny, what a journey! And to think, I only went to university in the first place because you thought I had potential (I guess you were right, as usual). Thank you for being by my side. I can’t wait to share the next chapter with you, my number one!

List of Figures

List of Tables

Table of Contents

CHAPTER 1

Literature Review

1.1 General introduction and chapter overview

There is growing evidence for the existence of cross-sensory correspondences in early infancy (see Dolscheid, Hunnius, Casasanto & Majid, 2014; Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004; Wagner, Winner, Cicchetti & Gardner, 1981; Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson, 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018). That is, infants perceive that some features most directly related to one sensory modality (e.g. auditory pitch) can be mapped to the features of another (e.g. visuospatial height). In short, higher-pitch sounds are perceived as being pointier, smaller, thinner, brighter, lighter in weight, and higher in space than their lower-pitch counterparts. This early detection of cross-sensory correspondences, and in many cases before language develops, indicates that at least some correspondences are not linguistically mediated. Despite a prominent linguistic principle that the associations between words and their referents are arbitrary, the presence of these correspondences during infancy leads to the possibility that they support and are reflected in the form language takes. In other words, language has evolved to reveal cross-sensory correspondences in a number of ways. As just one example, auditory pitch is described in terms of its *height* by many languages, reflecting the known cross-sensory correspondence that exists between auditory pitch and visuospatial elevation.

When considering other ways in which cross-sensory correspondences are revealed through language, obvious examples include metaphors, such as sour tastes being described as *sharp*, melodies as sounding *smooth* or patterns as appearing *loud*.

Metaphors such as these indicate the extent to which correspondences govern everyday experiences and provide an outlet by which they can be communicated. In the context of paralinguistic communication (i.e. communication via the non-lexical elements of speech), how speakers capitalise on cross-sensory correspondences is largely unclear, but recent advances have been made to establish the role that prosody (i.e. the melody of spoken language) plays as a non-emotional cue to meaning (Herold, Nygaard, Chicos & Namy, 2011; Herold, Nygaard & Namy, 2011; Nygaard, Herold and Namy, 2009; Perlman, Clark & Falck, 2015; Shintel, Nusbaum & Okrent, 2006). For example, Shintel, Nusbaum and Okrent (2006) demonstrated how speakers will adjust their spoken pitch to reflect the directional movement of a dynamic object, adopting a higher-pitch tone of voice to refer to objects positioned higher (as opposed to lower) in space. Contributing to this literature, this thesis aims to demonstrate the functional significance of cross-sensory correspondences in infant-directed speech. Whilst there is some evidence to suggest that infant-directed speakers use prosody in meaningful ways (although not exclusively in regards to cross-sensory correspondences, see Nygaard, Herold and Namy, 2009), to date, there is no research identifying whether infants attend to these prosodic cues to resolve linguistic uncertainty (i.e. interpret the meaning of novel words).

In this chapter, a review of the relevant literature is presented. To begin, a summary of the research demonstrating the range of cross-sensory correspondences infants are sensitive to is provided, with emphasis on those correspondences that are related to auditory pitch (see section 1.2). Having only been outlined briefly thus far, the relationship between cross-sensory correspondences and language is then defined (see section 1.3). In this section, the assumption that language forms the basis for cross-sensory correspondences is also challenged in relation to the empirical evidence both within and across languages. Next, the importance of sound-symbolism is discussed in relation to this thesis (see section 1.4), with the rationale for the first four experiments outlined in some detail. Finally, an overview of the research demonstrating how speakers manipulate prosody in meaningful ways (particularly if in accordance with cross-sensory correspondences) is reviewed (see section 1.5). Attention is also given to the evidence suggesting that adult and child listeners infer the intended meaning of ambiguous utterances by attending to the prosody in which they are spoken.

1.2 Infants' sensitivity to cross-sensory correspondences

Higher-pitched sounds are judged by adults and infants alike to be pointier, smaller, brighter, thinner, lighter in weight and higher in visual space (see Chiou & Rich, 2012; Gallace & Spence, 2006; Keetels & Vrooman, 2010; Marks, 1987; Marks, 1989; Mondloch & Maurer, 2004; Parkinson, Kohler, Sievers & Wheatley, 2012; Spence, 2011; Walker, 2012). These associations are examples of *cross-sensory correspondences*: the tendency for people to associate progressively more extreme feature values in one sensory modality with more extreme feature values in another modality. So far, research has provided evidence for an early sensitivity to pitch-size (Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004), pitch-thinness (Dolscheid et al., 2014), pitch-brightness (Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004), pitch-pointiness (Walker et al., 2010) and pitch-visuospatial height (Dolscheid, Hunnius, Casasanto & Majid, 2014; Wagner, Winner, Cicchetti & Gardner, 1981; Walker et al., 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018) correspondences in preverbal infants.

In an early study, Wagner, Winner, Cicchetti and Gardner (1981) found that infants between the ages of 6- and 14-months look preferentially (measured by accumulated looking) towards the visual display of an arrow pointing in the same direction as an ascending or descending auditory tone. Researchers have since questioned the suitability of their visual stimuli in measuring the relationship between auditory pitch and visuospatial height, given that preverbal infants are unlikely to be familiar with arrows and their meaning (Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson, 2010). However, one alternative interpretation of these findings is that infants are not responding to an arrow's meaning (i.e. its direction) per se when making judgements about its corresponding tone, but instead are demonstrating a tendency to match stimuli that are visually top- or bottom-heavy with an ascending or descending tone, respectively. Given that the largest and most salient part of a top-heavy shape (i.e. an arrowhead) is positioned higher up than its bottom-heavy counterpart, these findings might still be interpreted as reflecting infants' sensitivity to pitch-visuospatial height correspondences. That is, when an auditory tone was increasing in auditory pitch, infants associated it with an object that was visually *higher* in terms of its centre of mass (and vice-versa).

Walker, Bremner, Mason, Spring, Mattock, Slater, and Johnson (2010) were amongst the first researchers to demonstrate the emergence of cross-sensory correspondences in infants as young as three months old and have since replicated some of these same findings with newborns (Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018). In their study (Walker et al., 2010), infants were presented with the visual display of either a ball travelling up or down in space or with a geometric shape morphing between a state of pointiness and roundedness, alongside the sound of a sliding whistle, sweeping in frequency between 300Hertz (Hz) and

1700Hz. They found that infants looked longer at congruent events (i.e. those that are in line with the core set of cross-sensory correspondences evident in the literature) than incongruent events. That is, they associated higher-pitch sounds with objects that were positioned higher in space or pointier.

For pitch-height correspondences only, Lewkowicz and Minar (2014) have failed to replicate the findings of Walker et al. (2010), despite Dolscheid, Hunnius, Casasanto and Majid (2014) doing so successfully with four-month-olds and Walker et al. (2018) doing so with newborns. In their study, Lewkowicz and Minar found that only older infants (aged six months old) could detect differences between congruent and incongruent pitch-height events, and only if the dynamic stimuli were presented at twice the rate as Walker et al.'s (2010) original experiment. They concluded that the findings of Walker et al. (2010) could be explained by infants detecting pitch-loudness interactions only. That is, as the auditory pitch and loudness (i.e. amplitude) of the tone rises or falls simultaneously, infants are attracted to the tone's internal congruence. In Walker et al.'s study, the loudness of the tone increased during the first half of the tone's rise or fall in frequency, peaked at mid-range and then decreased during the second half of its rise or fall in frequency, respectively. Because trials always began with the ball presented in its lowest visuospatial position, Lewkowicz and Minar argue that the congruency between auditory pitch and loudness at the start of a trial attracts infants' attention, which has been wrongly interpreted as a preference for congruent pitch-height events over incongruent events. In response to this, Walker, Bremner, Mason, Spring, Mattock, Slater and Johnson (2014) highlight that, whilst perhaps true for congruent pitch-height events, this argument fails to explain why infants also demonstrated a looking preference for congruent pitch-shape events. In these trials, the morphing shape always began in its pointiest form so, for congruent events, the relationship between auditory pitch and loudness was internally incongruent (i.e. pitch decreases whilst loudness increases). If infants were responding to pitch-loudness congruencies only, an association between higher-pitch sounds and more rounded shapes would be expected (contrary to Walker et al.'s observation).

Alternatively, if Lewkowicz and Minar's (2014) interpretation of looking behaviour in Walker et al.'s study is right (i.e. infants were attracted to the tone's internal congruence between auditory pitch and loudness), then their theory continues to support the claim that infants are sensitive to correspondences between auditory pitch and height, rather than disproves it. In fact, by making this suggestion, they even extend the findings to cover loudness-height correspondences also. That is, responding to the convergence of auditory pitch and loudness suggests that infants possess some non-linguistic notion of height, with higher-pitch and louder sounds being treated as *higher* than lower-pitch and quieter sounds, respectively. In other words, in order to detect that auditory pitch and loudness are internally congruent, infants must first interpret them as travelling in a particular direction.

Like Walker et al. (2010), Lewkowicz and Minar (2014) chose to increase the loudness of the tone during the first half of its rise in frequency. However, rather than decreasing its loudness during the second half of its rise in frequency, they held loudness constant. Walker et al. (2014) emphasise the importance of maintaining a tight association between stimuli that are presented auditorily and visually, hence their decision to reduce the tone's loudness as the moving ball approached the top and bottom of the display. They predict that Lewkowicz and Minar's null result might actually reflect their decision to hold loudness constant during the second half of its rise in frequency, causing infants to dissociate the auditory tone and moving ball and, as a result, fail to detect correspondences between the tone's pitch and the ball's visuospatial position.

As discussed briefly in the previous examples, not all research related to cross-sensory correspondences has achieved comparable findings across ages. This is also the case within studies, where differences in performance with increased age are reported (Fernández-Prieto, Navarra & Pons, 2015; Lewkowicz & Minar, 2014). When Fernández-Prieto, Navarra and Pons presented four- and six-month-old infants with the visual display of two balls independently increasing or decreasing in size alongside the auditory presentation of a tone increasing or decreasing in auditory pitch, they found that only older infants displayed a sensitivity to pitch-size correspondences. This was measured as significantly longer looks towards the ball whose size was congruent with the pitch of the tone (i.e. smaller is higher pitched). These findings support theories predicting that cross-sensory correspondences are a learned aspect of perception and challenges those that advocate their innateness. Despite this, six-month-old infants are preverbal and so the findings of Fernández-Prieto et al. also lend support to the idea that cross-sensory correspondences (at least for pitch-size correspondences) do not depend entirely on language.

Despite Fernández-Prieto et al. (2015) failing to establish pitch-size correspondences in younger infants, other research has had more success in recording correspondences of this kind with a similar age group (Dolscheid, Hunnius, Casasanto & Majid, 2014). In a preferential looking task, four-month-olds looked longer towards congruent, rather than incongruent, pitch-thinness events (Dolscheid et al., 2014). That is, infants preferred objects that were visually thinner when accompanied by a tone rising in auditory pitch. However, whilst the motivation for this research was to explore infants' sensitivity to pitch-thinness correspondences, thinness was conflated with size. In their study, Dolscheid et al. manipulated the thinness of an object by holding its height constant and increasing (thick) or decreasing (thin) its width. Increasing the width of objects like these, results in an increase of object size, such that objects with a larger width also have a larger footprint. Perhaps in this study, infants were responding to the size of objects rather than their thinness or perhaps it is the additive nature (i.e. thinness *and* size) of Dolscheid et al.'s stimuli that contributed to cross-sensory correspondences being recorded in four-month-olds. As just one example, this observation demonstrates the need for stimuli employed in cross-sensory research to be unambiguous in their interpretation.

Whilst the combination of some features (e.g. thinness and size) might contribute to the emergence of cross-sensory correspondences in younger participants, other featural combinations might eradicate cross-sensory effects entirely, perhaps in an instance whereby one feature is more salient than another. In a violation-of-expectation task, Haryu and Kajikawa (2012) found that ten-month-olds failed to associate auditory pitch with size (despite Fernández-Prieto, Navarra & Pons, 2015, demonstrating this ability with six-month-olds). In their experiment, infants were habituated to the display of two objects, each of varying sizes. One of the objects (e.g. the smaller of the two) travelled vertically and at a constant rate, ricocheting off the top and bottom of a screen, whilst the other object (i.e. larger) remained stationary throughout. Whenever the moving object was in contact with the bottom of the screen, a tone was presented. Test trials followed a similar procedure, except that the object that was moving during habituation trials (e.g. smaller) was now static and the object that was stationary during habituation trials (e.g. larger) was now dynamic. Whenever the moving object was in contact with the bottom of the screen, a tone with either a relatively higher- or lower-pitch than habituation trials was presented. Haryu and Kajikawa reported no difference in the time spent looking at congruent pitch-size events (e.g. higher-pitch tone paired with smaller object) compared with incongruent events (e.g. higher-pitch tone paired with larger object) for infants. However, in their design, two features (visuospatial height and size) are introduced, each with the potential to impact cross-sensory correspondences related to auditory pitch. One possibility is that some features (e.g. visuospatial height) provoke a stronger cross-sensory effect than others (e.g. size). For instance, if participants were responding to correspondences between auditory pitch and visuospatial height rather than pitch and size, we would expect a preference for objects positioned lower in space when accompanied by a lower-pitch tone. In Haryu and Kajikawa's experiment, a tone was only presented when an object was in contact with the bottom of the screen (i.e. in its lowest visuospatial position) and, for both congruent and incongruent events, 50% of the trials presented a tone which was lower pitch (rather than higher). Should an object's visuospatial position overshadow its size in this enterprise, infants would be expected to be surprised by scenarios in which the tone was higher pitch (irrespective of the object's size), perhaps accounting for the similar looking behaviour reported across congruent and incongruent events in Haryu and Kajikawa's study.

Whilst looking behaviour is a popular measure for research concerned with cross-sensory correspondences in very young infants, the data gathered from these sorts of studies are often open to alternative interpretations. Do participants look at objects that confirm their expectations or violate them? Does a decrease in looking reflect a clear-cut shift in attention? Whilst these sorts of questions might initially concern researchers investigating the presence of correspondences in infants, one advantage of cross-sensory research is that the direction of looking behaviour is, in part, trivial. That is, a preference in either direction (i.e. looking towards a congruent or incongruent stimulus, in cross-sensory terms) indicates sensitivity to a given correspondence.

An alternative method to looking behaviour is to administer tasks that require a verbal and/or manual response to be made. Whilst restricted by the cognitive limitations of the infant, some studies have proven to be successful in this enterprise (for infants see Mondloch & Maurer, 2004; and for young children see Nava, Grassi & Turati, 2016). Mondloch and Maurer presented 30-month-olds with the visual display of two bouncing balls. Each ball differed by size, surface brightness or both and, as a pair, were presented alongside a higher- or lower-pitch tone. When asked which ball they thought made the sound, they found that infants would match brighter and smaller balls with higher-pitch tones. Unlike in Haryu and Kajikawa's (2012) study where one object was dynamic and the other was static, Mondloch and Maurer presented two dynamic objects moving synchronously. In this instance, it was not possible for the infant to attribute the tone to one object over another based on its visuospatial height. Therefore, infants were forced to attend to other (perhaps less salient) features that differentiated the two objects (e.g. brightness and/or size).

The review of the literature in this section demonstrates that infants are sensitive to cross-sensory correspondences very early on in development and, for the most part, prior to language production (Dolscheid, Hunnius, Casasanto & Majid, 2014; Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Lewkowicz & Minar, 2014; Mondloch & Maurer, 2004; Wagner, Winner, Cicchetti & Gardner, 1981; Walker et al., 2010; Walker et al., 2018). Whilst there are some age-related differences reported across studies, often it appears that these can be explained by differences in methodological approaches. For instance, Walker et al. (2010) proposed that by reducing the association between stimuli presented auditorily (e.g.

tone) and visually (e.g. ball), the ability to detect correspondences between auditory pitch and visuospatial height is reduced. A key issue to emerge from this review is the importance of stimuli being unambiguous in terms of their interpretation, especially given that the impact of this on cross-sensory perception is still largely unknown. For example, in order to accurately explore cross-sensory correspondences related to size, objects designed to represent different sizes should not also vary in other cross-sensory features, such as visuospatial height, brightness or pointiness (to name just a few). This issue will be returned to at various stages of this thesis and provides the basis for which stimuli in this thesis are designed.

1.3 Cross-sensory correspondences and language

Although many stimulus features are perceived multimodally, some are assumed to be unisensory: generally speaking, *smoothness* can only be experienced by touch and *brightness* through vision. Why is it then that we refer to melodies as sounding *smooth* or foods as tasting *dull*? Even for stimulus features that are multimodal, such as *sharpness* which can be experienced visually or tactually, language reveals correspondences across other, seemingly unrelated modalities. For instance, *sharp* can be used to describe a sour taste, an unpleasant smell or a higher-pitch sound. Metaphors such as these are just one way in which our appreciation of cross-sensory correspondences is revealed in everyday life, another being that speakers will spontaneously raise or lower their spoken pitch when referring to objects travelling up or down in space, respectively (Shintel, Nussbaum & Okrent, 2006), and when reading aloud narratives that make reference to different sizes (i.e. higher-pitch tone of voice for *smaller*, see Perlman, Clark & Falck, 2015).

Notwithstanding their appearance in language, sensitivity to cross-sensory correspondences can precede language acquisition, with preverbal infants being found to appreciate a range of known cross-sensory correspondences (as outlined in section 1.2 of Chapter 1). To reiterate, four-month-old infants associate higher-pitch sounds with objects that are positioned higher in space (Dolscheid et al., 2014; Walker et al., 2010), as do newborns (Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018). They also associate higher-pitch sounds with objects that are pointier (Walker et al., 2010) and thinner (or as interpreted above, smaller) (Dolscheid et al., 2014). Such findings support theories predicting that cross-sensory correspondences are an unlearned aspect of perception. One proposal (commonly referred to as the 'neonatal synaesthesia' hypothesis) is that infants are born with an interconnected sensory system that allows for an enhanced sensitivity to associations that exist between different sensory modalities (Maurer, 1993; Maurer, Gibson, & Spector, 2013). For instance, the concept of *brightness* might be encoded via vision, touch, taste, smell or hearing very early on in development, but as the brain changes over the course of typical development, it is thought that some of these connections are dissolved. Given that adults are reported to appreciate the same cross-sensory correspondences as infants, one possibility is that we preserve correspondences from infancy that are in some way beneficial to survival, or that the correspondences that persist are those that are reinforced through other means (e.g. language).

If cross-sensory correspondences are independent of language, then illiterate beings might also reveal a sensitivity to them. Ludwig, Adachi and Matsuzawa (2011) have shown that chimpanzees demonstrate similar pitch-luminance (i.e. brightness) mappings as human adults. In a speeded manual discrimination task, participants classified stimuli as being either black or white in colour, whilst simultaneously ignoring a background tone varying in auditory pitch. When the tone's pitch was congruent with the target stimuli's colour (i.e. higher-pitch with brighter), participants performed with higher accuracy in the task. Whilst Ludwig, Adachi and Matsuzawa make the claim that their findings reflect the innateness of cross-sensory correspondences (in the sense that chimpanzees have not learned the association between auditory pitch and brightness), others are less convinced (see Spence & Deroy, 2012). Whilst language can be ruled out as an influencing factor in this case, other experiences with one's sensory environment might facilitate these sorts of mappings (e.g. the sky is both bright and high and smaller animals are more likely to fly than larger animals).

The correspondence between auditory pitch and visuospatial height (i.e. higher-pitch sounds are *higher* than lower-pitch sounds) is one of the most widely tested in cross-sensory research, with preverbal infants (Dolscheid, Hunnius, Casasanto & Majid, 2014; Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson, 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018), children (Nava, Grassi & Turati, 2016) and adults (Bonetti & Costa, 2018; Evans & Treisman, 2010) demonstrating similar pitch-height mappings. These associations are also revealed in many languages (such as English), where auditory pitch is described in terms of its spatial height. Putting the evidence of pitch-height correspondences in preverbal infants to one side momentarily, it is perhaps unsurprising that language users judge higher-pitch sounds as being higher in space when language reinforces this particular correspondence. But are cross-sensory correspondences an artefact of language or are they reflected in the form language takes? Of course, the existence of cross-sensory correspondence in preverbal infants suggests that correspondences preexist language, but a second approach to answering this question is to explore whether pitch-height correspondences exist when languages describe auditory pitch in other ways. In a speeded classification task, monolingual speakers of Kreung (a tribal language originating in Cambodia) classified the auditory pitch of various tones (Parkinson, Kohler, Sievers, & Wheatley, 2012). In Kreung, pitch is described in terms of its *tightness*, with *high* and *low* pitch being replaced by *tight* and *loose* pitch, respectively. During the task, participants also viewed a ball rising and falling on a screen and they found that fewer errors were made classifying auditory pitch when the direction of the moving ball was congruent with the *direction* of pitch. These findings illustrate that the same cross-sensory correspondences can exist across languages, even when, linguistically, they are represented differently.

Converging on the same question but from a different angle, Shayan, Ozturk, Bowerman and Majid (2014) explored how describing pitch in terms of its *thinness* as opposed to its *height* impacts one's sensitivity to pitch-thinness correspondences. They found that for Farsi and Turkish speaking children and adults, who refer to higher-pitch sounds as *thin* and lower-pitch sounds as *thick*, the pitch-thinness correspondence was well established. That is, children and adults associated higher-pitch sounds with thinner objects. However, for German children of the same age (who describe pitch in terms of spatial height), their ability to match auditory pitch and thinness was at chance level (despite German adults doing so successfully). Whilst evidence indicates that cross-sensory correspondences can emerge prior to language acquisition, these findings suggest that consistent metaphorical mappings within languages play a part in moderating the strength of correspondences appreciated by the language user.

Despite the tendency for auditory pitch to be described in terms of its thinness for speakers of Farsi, Turkish and Zapotec, it was not exclusively so (Shayan, Ozturk

& Sicoli, 2011). Instead, auditory pitch was also described in terms of its size, strength, sharpness and visuospatial height, all of which are perceptual qualities regularly evident in cross-sensory correspondences. One possibility is that auditory pitch is experienced similarly across speakers of all languages (that is, higher-pitch sounds are judged by all as being pointier, smaller, brighter, thinner, lighter in weight and higher in visual space than lower-pitch sounds) but the ways in which languages have evolved to describe pitch is environmentally mediated.

When exploring the relationship between cross-sensory correspondences and language, it is also important to consider the transitive nature of correspondences. For example, a sensitivity to pitch-height, pitch-thinness and indeed pitch-tightness correspondences, has the potential to form secondary relationships between visuospatial height, thinness and tightness (see Figure 1.1). In this case, Parkinson, Kohler, Sievers and Wheatley's (2012) observation that participants made fewer errors classifying auditory pitch in terms of its tightness when accompanied by a moving ball that was congruent with the *direction* of pitch, might be explained by a sensitivity to both pitch-tightness and pitch-height correspondences or by a sensitivity to a secondary relationship that exists between height and tightness. Of course, both explanations rely on the assumption that participants associate higher-pitch sounds with objects positioned higher (rather than lower) in space, but a relationship between converging features (such as height and tightness) would be easier to access in this enterprise. For instance, rather than having to identify two separate correspondences (one between pitch and height and the second between pitch and tightness), it is possible to access both by acknowledging the relationship between height and tightness.

Figure 1.1. Demonstration of the transitive nature of cross-sensory correspondences. The features on the left (*tight*, *thin* and *high*) are all independently associated with higher-pitch sounds and the features on the right (*loose*, *thick* and *low*) are associated with lower-pitch sounds. The convergence of these features then allows for secondary relationships to be formed (e.g. between *tight* and *high* or *tight* and *thin*). For demonstrative purposes only, three dimensions are included in this example, but it is important to note that the sample is not intended to be exhaustive.

Whilst research is closer to defining the relationship that exists between cross-sensory correspondences and language, there is still a way to go to understanding its broader implications. The aim of this thesis is to identify if and how an infant's sensitivity to cross-sensory correspondences might be capitalised on to aid language comprehension. Although much of the research evidence points towards the idea that cross-sensory correspondences precede language production, infants are exposed to language from birth (and during gestation, see Moon, Lagercrantz & Kuhl, 2013), so its impact on the formation or retention of correspondences cannot be easily overlooked. In light of this, the next section of this chapter reviews research concerned with *sound-symbolism* (the idea that a word's sound can represent its meaning to some degree) and the evidence suggesting that infants associate words with objects based on their phonetic sound.

1.4 Sound-symbolism

Is the association between a word and its meaning arbitrary or do words carry meaning in and of themselves? How often does a word's sound represent (partially or fully) its referent? These are some of the questions that have guided research concerned with *sound-symbolism*, the idea that nonarbitrary relations exist between a word's sound and the object that it names. Examples of sound symbolism include onomatopoeic words (e.g. *squeak*, *bang* and *pop*) that, when vocalized, phonetically imitate the sounds that they describe. However, many words bear no obvious likeness to their meaning at all (e.g. *tree*, *dog* and *computer*), leaning towards the possibility (at least for English) that sound-symbolism might only encompass a small subset of examples. Despite this, research demonstrates that many speech sounds actually carry subtle meanings, which, when present within a word, have the capacity to impact how appropriate that word is as an object's name. One of the earliest accounts of this phenomenon comes from research by Köhler (1947) who asked adults to pair novel shapes varying in visual pointiness with novel words and found that words such as *maluma* and *takete* were more likely to be associated with rounded and pointed shapes, respectively (see Figure 1.2). More recently, near identical findings have been obtained by Nielsen and Rendall (2011) and also by Ramachandran and Hubbard (2001) when they replaced *maluma* and *takete* with *bouba* and *kiki,* respectively.

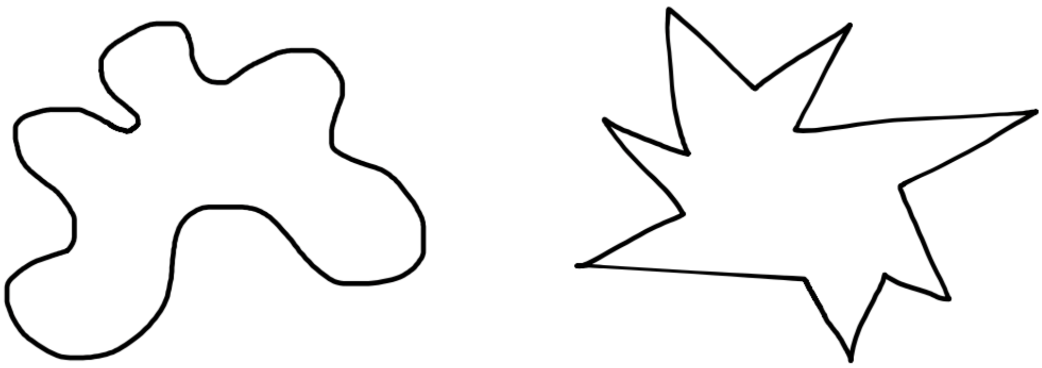


Figure 1.2. Demonstration of *maluma* and *takete,* adapted from Köhler's (1947) original experiment. Participants associated the visually rounded shape on the left with *maluma* and the pointed shape on the right with *takete*.

Research has so far demonstrated that infants as young as four months old perceive similar sound-symbolic relationships to adults (for a review see Fort, Lammertink, Peperkamp, Guevara-Rukoz, Fikkert & Tsuji, 2018; Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011). For instance, Ozturk, Krehm and Vouloumanos found that infants will look longer towards a visual stimulus (e.g. rounded shape) if accompanied by a mismatching auditory label (i.e. *kiki*). This early sensitivity to sound-symbolism has led researchers to question how this enterprise might be capitalised on to aid language development, particularly whether it might facilitate the learning of associations between new words and their referents. In one study, Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada and Kita (2015) habituated fourteen-month-olds to audio-visual stimuli pairs that were either congruent (i.e. visually rounded shape paired with *moma* and pointy shape paired with *kipi*) or incongruent (i.e. rounded shape paired with *kipi* and pointy shape paired with *moma*) with expected sound-symbolic relationships. Infants were then presented with both shapes simultaneously and heard one of the same pseudowords as on the habituation trials (*kipi* or *moma*). In an attempt to establish which of the two shapes infants *preferred* in response to a given pseudoword, looking behaviour towards each shape was recorded. For infants that were habituated to congruent sound-symbolic pairs, Imai et al. found a significant increase in looks towards the correct shape in the presence of its learned name. In contrast, for infants that were habituated to incongruent sound-symbolic pairs, looking towards each object remained around chance level, possibly reflecting a conflict between what is learned during habituation trials (e.g. rounded shape matches *kipi*) and what infants associate naturally (e.g. rounded shape matches *moma*). In summary, these findings demonstrate that a sensitivity to sound-symbolism has the capacity to impact word learning. Specifically, the meaning of a word is more easily recalled if the word conforms with sound-symbolic biases rather than contradicts them. What remains unclear is whether the learning of sound-symbolic words differs to the learning of non-sound-symbolic words, and if sound-symbolic words possess an advantage in this context.

One approach to studying sound-symbolism is to explore the elements that qualify a word as being symbolic in terms of sound. Peña, Mehler and Nespor (2011) presented four-month-old infants with the visual display of two near identical objects, differing only in size, alongside the auditory presentation of a single vowel phoneme. Peña et al. found that for vowel sounds produced nearer the front of the mouth (e.g. /ɪ/ as in *pig* or /e/ as in *egg*) infants looked faster towards and longer at smaller objects, whereas for vowel sounds produced nearer the back of the mouth (e.g. /a/ as in *hat* or /u/ as in *bug*) infants looked preferentially towards larger objects. These findings echo those from similar studies in which adults pair words containing front-closed vowel sounds with pointier or lighter objects and pair words containing back-open vowel sounds with rounder and heavier objects (Monaghan, Mattock & Walker, 2012;

Walker, Barnett & Parameswaran, submitted; Walker & Parameswaran, 2019). Having analysed the intrinsic fundamental frequency[1] of each phoneme included in their stimulus set, Peña et al. found that vowels produced nearer the front of the mouth were higher in auditory pitch than those produced nearer the back. This is an important finding, given the cross-sensory relationship that exists between auditory pitch and visual size, with higher-pitch sounds being judged to be smaller than lower-pitch sounds (Fernández–Prieto et al., 2015; Gallace & Spence, 2006; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004). One possibility is that infants are responding to differences in auditory pitch to guide looking behaviour to either object, rather than any other special characteristics of the vowel sounds themselves.

In one study, Bottini, Barilari and Collignon (2019) explored how visually impaired adults performed in tasks measuring symbolic pointiness compared with healthy controls. In the first of two experiments, participants handled two objects of varying tactile pointiness and were asked to select which object would most appropriately be named *maluma* or *takete*. Mimicking findings obtained by Köhler (1947), both visually impaired and sighted participants paired *maluma* with rounded objects and *takete* with pointy objects (consistently above 73% of the time). In a similar way, advances in cross-sensory research have established that some correspondences are not restricted to specific sensory channels (Barnett, Bremner, & Walker, submitted; Walker, Walker & Francis, 2012). Instead, Barnett, Bremner and

[1] The fundamental frequency is defined as the lowest frequency of a periodic waveform. Pitch is the perceptual quality of a frequency and whilst fundamental frequency and pitch are not identical, the term 'pitch' is often used interchangeably with fundamental frequency.

Walker found that the same relationship existing between auditory pitch and pointiness was induced whether pointiness was encoded through vision or touch.

In a second experiment, Bottini, Barilari and Collignon (2019) investigated how influential a word's graphemic structural features (i.e. its visual appearance in printed form) are on predicting word-object associations (see also Cuskley, Simner, & Kirkby, 2017). In this experiment, early blind participants with no experience of visual reading and healthy controls listened to a list of psuedowords and indicated how *pointy* each pseudoword sounded. They found that various elements of sound (e.g. consonant and vowel types[2]) had significant effects on ratings of pointiness across both groups of participants, but they also found an interaction between orthographic pointiness and sightedness, with the visual appearance of pseudowords affecting ratings of pointiness for sighted participants only. These findings suggest that symbolism in terms of a word's sound and symbolism in terms of a word's visual appearance exist independently from one another, which is supported by the existence of sound-symbolism in pre-reading infants (Imai, et al., 2015; Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011) and in adults from remote populations who do not use written language (Bremner, Caparos, Davidoff, Fockert, Linnell & Spence, 2013). What remains unknown is how the relationship between sound-symbolism and visual-symbolism is defined. One prediction is that a word's sound-symbolism is reflected in the form written language takes, so that words that sound pointier have evolved to be pointier in visual appearance. A second possibility

---

[2] According to the International Phonetic Alphabet (IPA), different consonant and vowel types are defined by their place of articulation (i.e. where in the vocal tract sounds are crafted) and by their manner of articulation (i.e. how the air in the vocal tract interacts with places of articulation to make a sound).

is that, whilst sound-symbolism and visual-symbolism appear to exist independently, sounds that converge in symbolism with their printed form are more salient to language users, perhaps leading to a situation whereby literate beings appreciate sound-symbolic exemplars differently to illiterate beings. This account might explain why research finds a weakened effect of sound-symbolism in visually impaired participants compared with healthy controls (Fryer, Freeman & Pring, 2014), with sighted participants having the advantage of both hearing and visualising words.

The extent to which sound-symbolism impacts word learning is yet to be fully established. A trend in infancy research concerned with language development is to use pseudowords (rather than real words) as stimuli, often with little or no consideration regarding their sound-symbolic potential. This creates a problem when the objects that researchers assign to these pseudowords have one or more of the perceptual qualities regularly evident in sound-symbolic biases. Given that prelinguistic infants as young as four months are found to be sensitive to sound-symbolism (Imai, et al., 2015; Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011), there is a need in the literature for research to identify a selection of unbiased (sound-symbolically speaking) pseudowords for wider use. Whilst this thesis aims to assess whether infants attend to prosody (i.e. the melody of spoken language) to resolve linguistic uncertainty, the aim of Chapter 2 is to identify a list of pseudowords with varying levels of symbolism (including neutral), whilst taking into account both their sound-symbolism and visual-symbolism.

1.5 Prosody as a semantic marker of cross-sensory correspondences

The proposal that cross-sensory correspondences are an unlearned aspect of perception (Maurer, 1993; Maurer, Gibson, & Spector, 2013) is supported by their existence in newborns (Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018) and preverbal infants (Dolscheid et al., 2014; Walker et al., 2010). However, infants are shown to be receptive to language very early on in development, including as fetuses where they demonstrate a preference for their mother's voice over other voices (Kisilevsky et al., 2003; Marx & Nagy, 2015) and discriminate utterances in their native language from those of a novel language at just four days of age (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison, 1988). These findings illustrate that infants have the potential to rely on mechanisms of language other than linguistic content, such as prosody, to make sense of and learn from their environment. Given that speakers reportedly adjust their spoken pitch in accordance with some cross-sensory correspondences (Nygaard, Herold & Namy, 2009; Shintel, Nusbaum & Okrent, 2006), these observations have wide-reaching implications for understanding the onset, development and retention of cross-sensory correspondences, including the possibility that infants learn correspondences from the ways in which others use speech.

In linguistics, prosody is defined as the elements of speech that contribute to speech's acoustic profile (i.e. rhythm) and, traditionally, has been viewed as insignificant in terms of its contribution to the meaning of language. First, and perhaps foremost, prosody contributes to the structural organisation of spoken language. It is characterised by variations in vocal pitch, tempo and loudness, which helps listeners to identify phrase boundaries, locate the prominent parts of an utterance and recognise the correct form of a sentence (e.g. question, statement or command). However, recent research shows that prosody is also recruited to communicate affective information across speakers (Hanuliková & Haustein, 2016; Nygaard & Lunders, 2002). For example, when adults were asked to transcribe ambiguous homophones that were either emotional or neutral in their meaning (e.g. die [emotional] vs dye [neutral]), they were more likely to transcribe homophones correctly if they were spoken in an emotionally congruent tone of voice (Nygaard & Lunders, 2002). Even for non-native language users, research shows that listeners will attend to the prosody of unfamiliar languages to resolve linguistic uncertainty (Hanuliková & Haustein, 2016). In their study, Hanuliková and Haustein presented German monolinguals with English homophones of various affective meanings. When asked to interpret them, participants were more likely to choose a sadder meaning for homophones produced in a sadder tone of voice.

Whilst the relationship between prosody and emotion is well established, a handful of studies have demonstrated associations between prosody and meaning that are unrelated to emotion (Herold, Nygaard, Chicos & Namy, 2011; Herold, Nygaard & Namy, 2011; Hupp & Jungers, 2013; Nygaard, Herold & Namy, 2009; Shintel, Anderson & Fenn, 2014; Shintel, Nusbaum & Okrent, 2006; Tzeng, Duan, Namy & Nygaard, 2018). For example, when speakers were asked to report the direction that a moving object was travelling in, Shintel, Nusbaum and Okrent (2006) found that they would spontaneously adjust their spoken pitch and rate of speech to reflect its movement. That is, participants spoke with a higher pitch for objects travelling upwards and at a faster rate for objects moving more quickly. Other research illustrates the role prosody plays in enhancing memory (Shintel, Anderson & Fenn, 2014), with participants recalling the meaning of a word more easily if, during learning, the word's prosody was congruent with its meaning (e.g. higher-pitch tone of voice for an object positioned higher in space).

Even children as young as four years old have been found to match the rate of their speech to the speed of a moving object (Hupp & Jungers, 2013). Adults and children were asked to verbally indicate which target object (e.g. cat) a moving object (e.g. star) was approaching (e.g. “The star is going to the cat.”). They found that when the target object was moving quickly (compared with slowly), both adults and children adopted a faster rate of speech. One possibility for these findings is that, when presented with a faster moving object, participants rushed their speech so that they could complete the sentence (e.g. “The star is going to the cat.”) before the target object reached its destination. These concerns were addressed by the authors, who demonstrated that sentences were completed well within the movement time of the target object (which is in line with other research showing speakers manipulating their rate of speech for moving objects with no end point, see Shintel & Nusbaum, 2007). If participants were conscious of completing sentences within a desired length of time, Hupp and Juger predicted that the length of spoken sentences would more closely reflect the time available to the speaker.

Hupp and Junger (2013) also found that adults and children were accurate at identifying which of two moving objects a speaker was describing, by matching the speaker’s speech rate with the correct object’s speed. Whilst adults successfully completed the task with subtler prosodic cues, children only did so if the rate of speech and speed of the moving object were presented at an exaggerated rate. In a similar way, Herold, Nygaard, Chicos and Namy (2011) also found that prosody contributes to the interpretation of novel words for young children. Children were presented with picture pairs portraying opposite dimensional adjectives (e.g. small vs big) and asked to identify which picture represented a novel word, spoken in a semantically rich prosody. They found that four-year-olds could only complete the task if training was provided and, while five-year-olds were more confident with the requirements of the task, their success was dependent on being told to attend to a speaker's prosody. Whilst it is tempting to interpret these findings as revealing something about the developmental nature of prosody as a cue to word meaning, another explanation is that these findings reflect age-related differences in task transparency. Children could complete the task with assistive measures, so their initial reluctance to attend to prosody might actually reflect their more general cognitive limitations, rather than an inability to interpret a word's meaning based on the prosody in which it is spoken. In summary, the research presented thus far illustrates that prosody is a mechanism employed by speakers and listeners to communicate paralinguistically (i.e. via the non-lexical elements of speech). However, it is still unknown how infants respond to prosody in this enterprise.

Famously, speakers adopt a prosodically rich speech style (referred to as *infant-directed speech*) when communicating with infants, which is characterised by its greater pitch variation, slowed and deliberate duration of speech, and an overall higher spoken pitch (Lee, Kitamura, Burnham & Todd, 2014) compared with adult-directed speech. Whilst there are some stylistic differences found across cultures, for example, the extent to which infant-directed speech (IDS) is exaggerated (Kitamura, Thanavishuth, Burnham & Luksaneeyanawin, 2001), IDS is regarded as a universal phenomenon (Broesch & Bryant, 2015). Nygaard, Herold and Namy (2009) were amongst the first researchers to investigate the role of prosody in IDS in portraying and inferring word meaning. Participants produced phrases in IDS containing novel words (e.g. "Can you get the *blicket* one?"), which were assigned meanings (e.g. happy) from one of six antonym pairs (i.e. happy/sad, hot/cold, big/small, tall/short, yummy/yucky or strong/weak). They found that speakers modified their spoken pitch, amplitude and rate of speech to reflect opposite dimensions within antonym pairs, for instance, compared with words meaning *big*, they spoke with a relatively higher-pitch, quieter amplitude and at a faster rate for words meaning *small*. When these utterances were presented to adult listeners alongside picture-pairs portraying these same antonym pairs, adults were successful at matching the novel words with their meaning. These findings illustrate that prosody in IDS can be a reliable, paralinguistic tool for communication.

But how precisely do prosodic cues reflect specific meanings, especially when these meanings can be grouped by broader factors (e.g. valence)? Nygaard, Herold and Namy (2009) explored the possibility that listeners respond to a word's positive or negative valence to guide word-object pairings, rather than its precise meaning. To test this, participants were again presented with prosodically meaningful utterances containing a novel word and asked to match the novel word with its meaning. However, rather than presenting participants with picture-pairs that matched the intended meaning of the utterance, participants were presented with alternative picture-pairs that matched in terms of valence only. For example, an utterance containing a novel word that was intended to mean *happy* was presented alongside pictures portraying *yummy* and *yucky* rather than *happy* and *sad*. Overall, participants performed poorly when a word's precise meaning and corresponding picture-pairs mismatched (despite matching in terms of valence), suggesting that participants were relying on prosodic cues that were domain-specific (i.e. belonging to a precise meaning within a given dimension e.g. happy as opposed to sad) rather than domain-general (i.e. belonging to a more general grouping factor e.g. positive as opposed to negative valence) when matching novel words with their correct meaning.

In a dyadic interaction task between mother and infant, Herold, Nygaard and Namy (2012) explored whether mothers use prosody in meaningful ways when reading aloud stories that make reference to different dimensional adjectives (e.g.

small vs big) using real rather than novel words. Unlike Nygaard, Herold and Namy (2009) who found that infant-directed speakers manipulated their spoken pitch for a range of dimensional adjectives, Herold, Nygaard and Namy (2012) found that pitch was only used to differentiate between the dimensional adjectives *strong* (lower-pitch voice) and *weak* (higher-pitch voice). In Herold, Nygaard and Namy's study, the words presented to participants were 'selected based on their conceptual familiarity to young children' (Herold, et al., 2012, p. 426) and, as such, there was an expectation that infant participants would understand (or at least be familiar with) the words included in the study. As demonstrated by Tzeng, Duan, Namy and Nygaard (2018) who found that only for novel words (rather than real words) would speakers manipulate their spoken pitch when referring to varying shades of brightness, one possibility is that the ambiguity of novel words forces speakers to exaggerate their prosody beyond what is found when words are familiar to a speaker and/or listener. In other words, when a word is familiar, the need to use prosodic cues to convey meaning becomes redundant.

The evidence to date suggests that speakers spontaneously recruit prosodic cues to communicate the meaning of words and that adults are accurate at inferring an unfamiliar word's meaning by interpreting the prosody with which it is spoken (Hanuliková & Haustein, 2016; Herold, Nygaard & Namy, 2011; Hupp & Junger, 2013; Nygaard, Herold & Namy, 2009; Shintel, Nusbaum & Okrent, 2006; Tzeng, Duan, Namy & Nygaard, 2018). Whilst a handful of studies have demonstrated that children as young as four years old also use prosody as a cue to word meaning (Herold, Nygaard, Chicos & Namy, 2011; Hupp & Junger, 2013), it remains unclear whether infants respond to prosody to resolve linguistic uncertainty. With research demonstrating that speakers manipulate prosody in a way that reflects known cross-sensory correspondences between pitch and visuospatial height (Shintel, Nusbaum & Okrent, 2006), pitch and brightness (Tzeng, Duan, Namy & Nygaard, 2018) and pitch and size (Nygaard, Herold & Namy, 2009), it is possible that there are other correspondences reflected in speech that are yet to be identified, for example pitch-thinness, -pointiness, -brightness and -weight. One aim of this thesis is to expand the works by Nygaard, Herold and Namy (2009) to cover these correspondences and then to establish if infants respond to prosody in this context.

1.6 General Conclusion and Thesis Objectives

Whilst an early, preverbal sensitivity to cross-sensory correspondences is well-established, research is yet to identify the functional significance of correspondences for infants. Whilst some researchers have identified prosodic correlates to word meaning in infant-directed speech (Nygaard, Herold & Namy, 2009), none have explored this topic exclusively in regard to cross-sensory correspondences, nor have they identified whether, like adults and children, infants attend to prosody in order to resolve linguistic uncertainty (i.e. interpret the meaning of novel words).

A review of the sound-symbolism literature has identified an important issue that requires addressing prior to exploring the functional significance of cross-sensory correspondences as outlined above. With infants demonstrating a sensitivity to sound-symbolism (Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada & Kita, 2015; Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011), and with research related to language development so far neglecting the impact this might have on their findings, there is a need in the literature for a set of pseudowords that vary in terms of their sound-symbolic potential. Prior to achieving this, a more comprehensive review of the related literature is required (see section 2.1.1 of Chapter 2). The series of experiments that then follow in Chapter 2 combine to explore the relationship between a word's sound and its symbolic pointiness, whilst also taking into consideration its visual-symbolism at a surface level (i.e. typeface) and at a graphemic-structural level (i.e. a word's visual appearance in printed form).

The aim of Chapter 3 is to extend works by Nygaard, Herold and Namy (2009) to identify prosodic correlates to word meaning that are related to known cross-sensory correspondences in infant-directed speech, and then to explore whether infants attend to these cues to interpret the meaning of unfamiliar words (Chapter 4). Finally, in Chapter 5 the main findings are summarised and discussed in relation to their wider implications, with suggestions for future studies and more general lines of investigation.

CHAPTER 2

Establishing neutrally sound-symbolic pseudowords

2.1 Introduction

Despite a prominent linguistic principle that a word's sound and its meaning are arbitrarily related, research demonstrates that some speech sounds are more or less likely to be associated with objects of a particular size (Thompson & Estes, 2011), shape (Köhler, 1947; Nielson & Rendall, 2011; Ramachandran & Hubbard, 2001) and taste (Ngo, Velasco, Salgado, Boehm, O'Neill & Spence, 2013). As also outlined in section 1.4 of Chapter 1, some of the most prominent examples of sound-symbolism are concerned with an object's pointiness and demonstrate how varying consonant and vowel sounds impact the appropriateness of names for pointy and rounded shapes. This sensitivity to sound-symbolism appears to emerge very early in development (Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada & Kita, 2015; Fort, Lammertink, Peperkamp, Guevara-Rukoz, Fikkert & Tsuji, 2018; Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011), with infants as young as four months old associating words such as *kiki* and *kipi* with pointier shapes and *bubu* and *moma* with rounder shapes.

Particularly in infancy research related to language development, there is a growing trend to present participants with novel pseudowords during experimental investigations. Typically, pseudowords are selected without acknowledging their sound-symbolic potential and, whilst the research in question might not be related to sound-symbolism per se, this has been shown to have profound effects on performance in these sorts of studies. For example, Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada and Kita (2015) demonstrated that when pseudowords sound-symbolically mismatched with the objects they named, infants failed to learn them. In many cases, sound-symbolism is an extension or subcategory of cross-sensory correspondences, with the features related to one sensory modality (in this case, words encoded auditorily) being mapped to the features of another (e.g. pointiness encoded visually or tactually). As a result, research records similar correspondences across these two fields, with vowel sounds produced nearer the front of the mouth being higher in auditory pitch and associated with smaller objects compared with vowel sounds produced nearer the back of the mouth, which are lower in auditory pitch and associated with larger objects (Peña, Mehler & Nespor, 2011).

The broader aim of this thesis is to identify the functional significance of cross-sensory correspondences in infant-directed speech, by exploring how prosody is used by adult speakers and infant listeners to convey and interpret a word's meaning. In part, this will be achieved by presenting infants with object pairs representing some of the perceptual qualities regularly evident in cross-sensory correspondences (e.g. size) and with pseudowords spoken in a meaningful style of prosody (e.g. higher-pitch tone of voice to refer to smaller as opposed to larger objects). In order to isolate the effect that prosody is having on word-object pairings in this context, it appears imperative to first eliminate any potential impact of sound-symbolism. Without excluding sound-symbolism from our investigations, relationships formed between words and objects could be the result of infants detecting sound-symbolic biases, cross-sensory correspondences related to prosody, or to some unknown combination of these two factors. With this in mind, the aim of this chapter is to identify a list of pseudowords with varying levels of sound-symbolism, but particularly neutral (i.e. pseudowords that are neither associated with pointy nor rounded objects but with objects that sit somewhere between these two extremes). It is these neutral (sound-symbolically speaking) pseudowords that will be used in all experiments reported in this thesis. To begin, a review of the current perspectives on sound-symbolism is presented, which is used to guide the research strategy adopted in this chapter, including the decision to explore the relationship between sound-symbolism and visual-symbolism in the context of our objectives (i.e. to identify neutrally sound-symbolic pseudowords).

2.1.1 Current perspectives on sound-symbolism

Despite evidence indicating that humans operate with some sort of sound-symbolic bias (one that causes them to inherently associate particular words with particular types of objects), it is still unclear what qualifies a word to be *sound-symbolic*. This is largely the result of the variability found in visual and auditory stimuli across research. Inconsistencies in word stimuli, such as length and speech sounds (i.e. consonant or vowel) make it difficult to pinpoint the elements of a word that contribute to its symbolism. For instance, compared with *kiki*, *takete* contains a greater number of phonemes and introduces different consonant and vowel sounds, yet both words are associated with visually pointier shapes over visually rounder shapes (Ramachandran & Hubbard, 2001; Köhler, 1947). These findings lead to the conclusion that it is the relative rather than absolute coding of a stimulus that is imperative for a word to be judged as being sound-symbolic (see also Brunetti, Indraccolo, Del Gatto, Spence & Santangelo, 2018). That is, compared with *bouba* and *maluma*, *kiki* and *takete* are judged to be more appropriate labels for pointier objects.

Much of the recent sound-symbolism literature has attempted to identify whether consonant or vowel sounds are more fundamental in the prediction of word-object pairings. However, evidence increasingly suggests that both consonant and vowel sounds play a shared role (but in what capacity is still unclear) in sound-symbolic associations (Monaghan, Mattock, & Walker, 2012; Nielson & Rendall, 2011; Peña, Mehler & Nespor, 2011; Westbury, 2005). For instance, by categorising vowel sounds in terms of tongue position during articulation (specifically, the tongue's proximity relative to the front [front/back] and roof [closed/open] of the mouth), Monaghan, Mattock and Walker (2012) found that participants paired visually pointier objects with words containing front-closed vowel sounds and paired visually rounder objects with words containing back-open vowel sounds. For consonant sounds, Westbury (2005) asked participants to identify real (as opposed to artificial) letters or letter strings presented in visually pointy or rounded borders and found that reaction times were slower when *stops* (i.e. consonant sounds that are achieved by blocking and then releasing the flow of air in the vocal tract during articulation, e.g. *t*, *k* and *d*) were presented in rounded frames and *continuants* (i.e. consonant sounds that are achieved by keeping the vocal tract partly open during articulation, e.g. *m*, *l* and *s*) were presented in pointy frames. Here, sensitivity to this type of sound-symbolism (stops sound pointier than continuants) interfered with performance in the task, despite it having no relevance to classifying letters or letter strings as real or not. These findings demonstrate the interfering nature of sound-symbolism and, since this thesis is concerned with prosody as opposed to sound-symbolism, reiterate the importance of identifying neutrally sound-symbolic words for our research and for wider use.

But do different words (and indeed phonemes) carry varying degrees of sound-symbolism? That is, are some words judged to be strongly symbolic whereas others are only weakly symbolic? In a series of experiments, Thompson and Estes

(2011) identified that sound-symbolism is not exclusively dichotomous in nature (i.e. distinguishing one value from another) but can mark degrees along a continuum between two opposite extreme values. In their study, Thompson and Estes presented participants with five novel objects of varying sizes, accompanied by five novel pseudowords and asked them to match each pseudoword with one of the five objects. Each pseudoword varied in the number of small- and large-sounding phonemes it contained, which were distributed so that each 6-letter string contained a systematic ratio of small- and large-sounding letters. Thompson and Estes selected voiceless stop consonants and front-closed vowels to represent smaller-sounding phonemes and voiced stop consonants and back-closed vowels to represent larger-sounding phonemes. This meant that pseudowords designed to represent smaller objects contained a larger ratio of voiceless stop consonants and front-closed vowels than their larger counterparts, and pseudowords representing larger objects contained a larger ratio of voiced stop consonants and back-closed vowels. In order to test the linearity between visual size and sound, pseudowords representing medium-sized objects contained an equal number of voiceless/voiced stop consonants and front-/back-closed vowels. Thompson and Estes found that as the number of large-sounding letters increased in pseudowords, participants were more likely to match pseudowords with larger objects.

Research is yet to establish if, like size (Thompson & Estes, 2011), words that evoke symbolic pointiness are also perceived as a graded function. The closest evidence to date comes from research by Tzeng, Nygaard and Namy (2017) who asked children between the ages of three and seven years to pair sound-symbolic foreign words and pseudowords with pointy and rounded shapes. Although all children demonstrated a sensitivity to sound-symbolism for pseudowords, only older children identified relationships between sounds and shapes for foreign words. Tzeng, Nygaard and Namy conclude that the foreign words revealed subtler examples of sound-symbolism than pseudowords designed to be sound-symbolic. This might explain why older children, who have had more experience with language, identified the sound-symbolic properties of words more readily than younger children. In terms of symbolic pointiness being appreciated as a graded function, the detection of sound-symbolic biases was dependent on the quality of segments that provided sound-symbolic cues. That is, words containing subtler symbolic cues were harder to identify as being sound-symbolic.

2.1.1.1 Sound-symbolism and visual-symbolism

One approach to understanding sound-symbolism is concerned with the relationship between a word's visual appearance in printed form (i.e. its graphemic structural features) and the visual appearance of its corresponding object/s. For instance, not only is *kiki* considered to be a more appropriate name for visually pointier objects than *bouba* is, when printed in identical typefaces, it is also visually pointier in appearance. In Thompson and Estes's (2011) research, they found a significant correlation between a pseudoword's visual width and the number of large-sounding phonemes it contained: pseudowords containing a greater number of large-sounding phonemes were significantly larger in size when printed than their small-sounding counterparts, which corresponded with the size of object the pseudowords were assigned to (see Figure 2.1). More recently, Cuskley, Simner and Kirby (2015) confirmed that the visual appearance of graphemes contained within words can strongly determine the shape of objects they appear associated with. In their research, native English speakers were required to rate the appropriateness of novel pseudowords when paired with visually pointy or rounded shapes. Here, Cuskley, Simner and Kirby generated a list of pseudowords that differed in terms of their phonological and orthographic pointiness. As was predicted, they found that pseudowords containing visually pointy/rounded graphemes were more likely to be associated with pointy/rounded shapes, respectively, and that participants relied more heavily on differences in orthographic pointiness than phonological pointiness to identify appropriate word-object pairings.

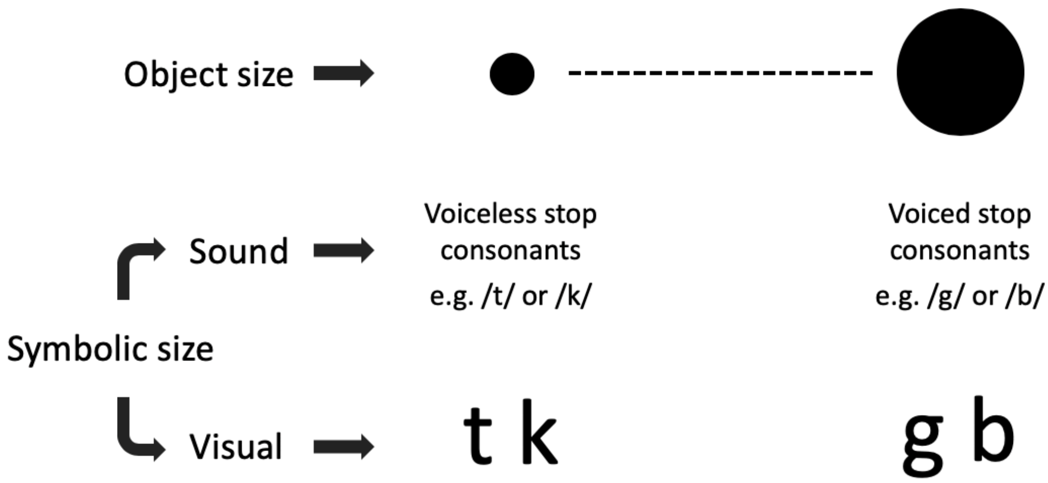


Figure 2.1. Demonstration of the association between sound-symbolism and visual-symbolism adapted from Thompson and Estes' (2011) findings. Sounds that were symbolic in terms of their size also corresponded with their visual width in printed form (i.e. sounds associated with smaller objects were smaller in size when represented graphemically). For demonstrative purposes, only consonants are displayed here, but Thompson and Estes found that the relationship between sound- and visual-symbolism for vowels also behaved in this way.

The association between sound-symbolism (i.e. symbolism in terms of a word's sound) and visual-symbolism (i.e. symbolism in terms of a word's appearance in printed form) is still largely unspecified. One possibility introduced in section 1.4 of Chapter 1 is that sound-symbolism and visual-symbolism exist independently, which is supported by the existence of sound-symbolism in pre-reading infants (Imai, et al., 2015; Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011). Perhaps for literate beings, sound-symbolism and visual-symbolism combine interactively to predict a word's symbolic potential (i.e. when sound-symbolism and visual-symbolism converge they induce a stronger symbolic impact than they otherwise do apart) or perhaps they combine in a relatively simple and additive manner. What remains clear is that a word's potential for sound-symbolism cannot be fully appreciated without also acknowledging its visual-symbolism. Therefore, in the quest to identify neutral (sound-symbolically speaking) pseudowords, the association between sound-symbolism and visual-symbolism is addressed in the series of experiments reported in this chapter.

2.1.2 Overview of Experiments 1 - 4

The collective aim of the series of experiments reported in this chapter is to identify a list of pseudowords that are judged to be neutrally sound-symbolic (i.e. they are neither associated with pointy nor rounded objects but with objects that sit somewhere between these two extremes). Given that one of the most popular examples of sound-symbolism is concerned with an object's pointiness (see Bottini, Barilari & Collignon, 2019; Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada & Kita, 2015; Köhler, 1947; Nielsen & Rendall, 2011; Ozturk, Krehm & Vouloumanos, 2013; Ramachandran & Hubbard, 2001), this chapter will explore if, like size

(Thompson & Estes, 2011), symbolic pointiness is also experienced as a graded function. We predict that by varying the number of pointy-sounding and rounded-sounding letters words contain, words can be manipulated so that they are more or less symbolic of pointiness. This work will also explore how sound-symbolism and visual-symbolism combine (i.e. interactively or additively) to define a word's symbolic potential.

To begin, a list of novel pseudowords, graded according to their expected phonological pointiness, was created (for details see section 2.2.1.2 of Chapter 2). In Experiment 1 and Experiment 2, participants were presented with the visual display of each pseudoword in printed form and asked to rate its visual appearance on a 7-point Likert-type scale, ranging from visually 'very pointy' to visually 'very rounded'. In Experiment 1 only, the verbal recoding of pseudowords was blocked (eliminating the availability of phonological features, see Figure 2.2), which, when compared with Experiment 2 (see Figure 2.3), provided the opportunity to explore the factors that predict symbolic pointiness (e.g. sound or vision) when phonological and visual features were differentially available.

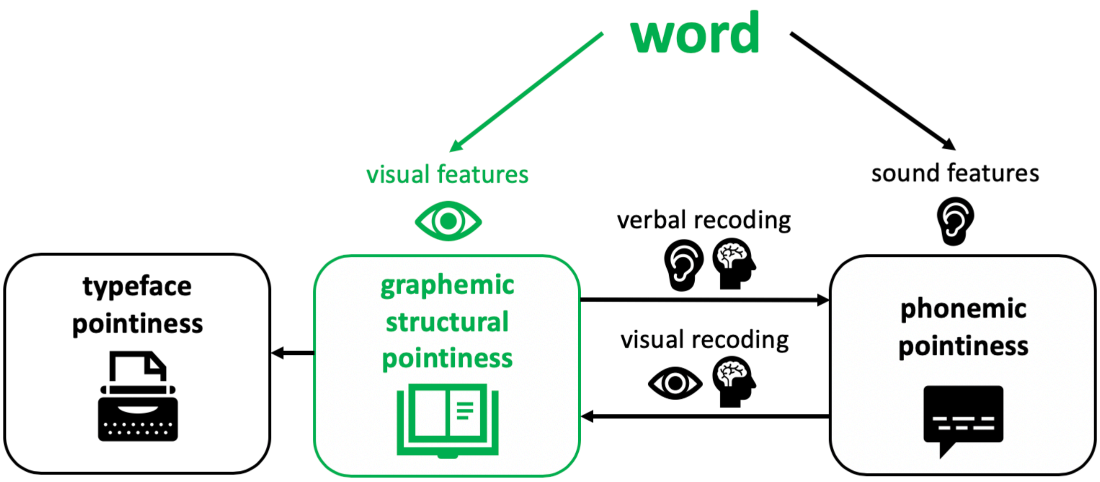


Figure 2.2. Demonstration of the visual features and sound features available (displayed in green) to participants in Experiment 1. Participants were presented with the visual display of each pseudoword in printed form and in a single typeface. The verbal recoding of pseudowords presented visually was blocked so that participants were unable to access a pseudoword's sound features (i.e. its phonemic pointiness).

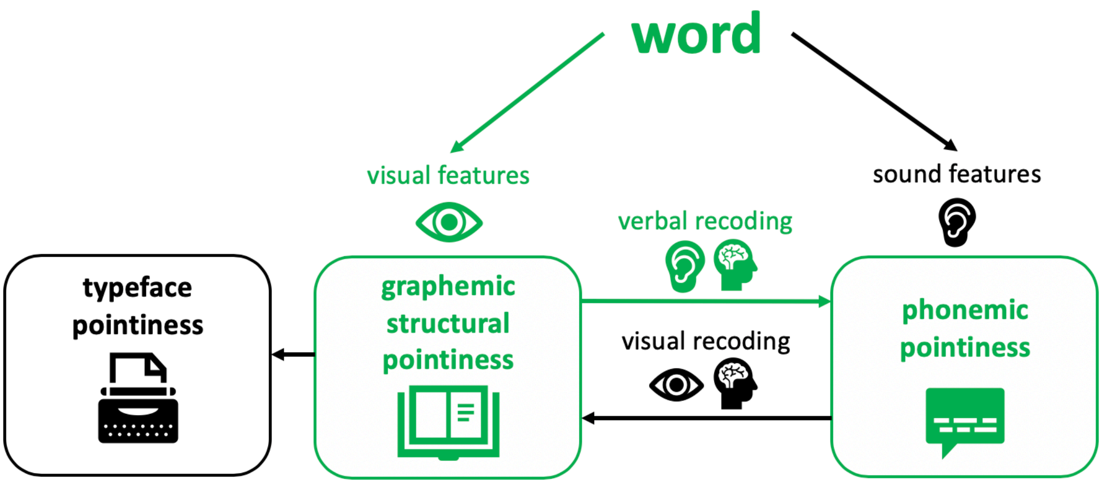


Figure 2.3. Demonstration of the visual features and sound features available (displayed in green) to participants in Experiment 2. Participants were presented with the visual display of each pseudoword in printed form and in a single typeface. Participants were free to verbally recode the pseudowords presented visually and so had access a pseudoword's sound features (i.e. its phonemic pointiness).

In order to identify any graded effects of symbolic pointiness, in Experiment 3 (see Figure 2.4) and Experiment 4 participants listened to the spoken presentation of each pseudoword and matched pseudowords with one of five objects representing a progression of visual pointiness (from pointy to rounded or vice-versa). Having identified the impact of a pseudoword's visual appearance on symbolic pointiness at a graphemic-structural level in Experiment 1 and Experiment 2, in Experiment 4 we presented pseudowords visually as well as auditorily and explored the impact of a pseudoword's typeface (also contributing to a pseudoword's visual appearance) on symbolic pointiness (see Figure 2.5).

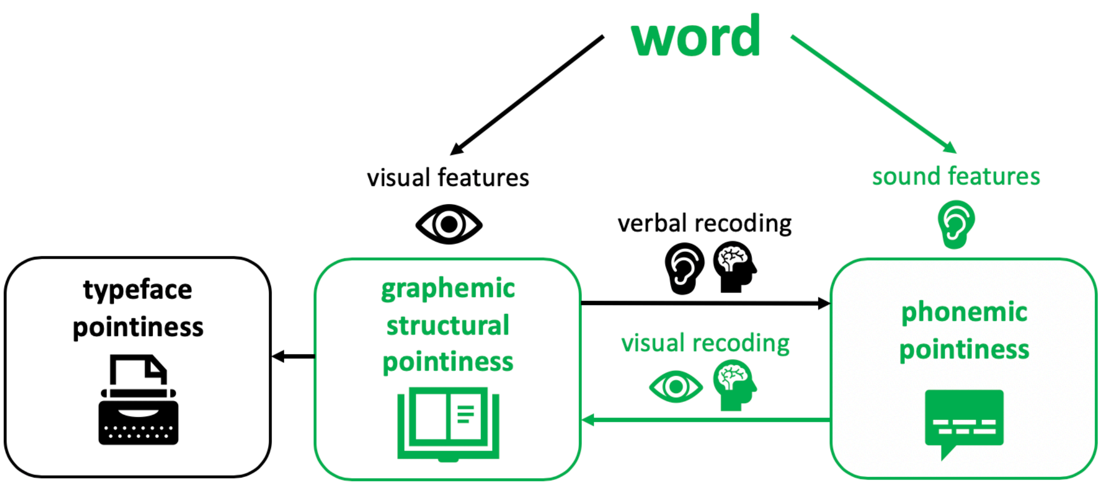


Figure 2.4. Demonstration of the visual features and sound features available (displayed in green) to participants in Experiment 3. Participants were presented with the spoken presentation of each pseudoword. Participants were free to visually recode the pseudowords presented auditorily and so had access a pseudoword's visual features (i.e. its graphemic structural pointiness).

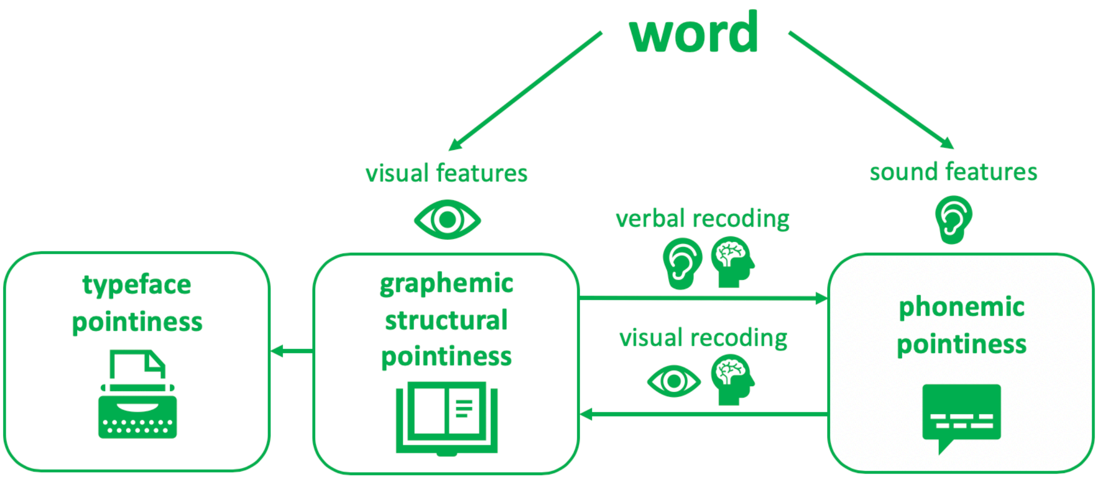


Figure 2.5. Demonstration of the visual features and sound features available (displayed in green) to participants in Experiment 4. Participants were presented with the visual display of each pseudoword in printed form, in either a visually pointy or rounded typeface, and with the spoken presentation of each pseudoword. Participants were free to verbally recode the pseudowords presented visually and visually recode the pseudowords presented auditorily, so had full access a pseudoword's sound features and visual features.

2.2 Experiment 1: Relationship between phonology and graphemes (verbal recoding blocked)

The objective of Experiment 1 was to explore whether sound-symbolism predicted visual-symbolism by identifying which psuedowords in our stimulus set were judged as being visually pointier in terms of their graphemic-structure (i.e. visual appearance in printed form). This was achieved by presenting pseudowords visually to participants and having them rate the visual appearance of each pseudoword on a 7-point Likert-type scale ranging from 1 (visually very pointy) to 7 (visually very rounded). In this experiment, the verbal recoding of pseudowords was blocked (i.e. phonological features were made unavailable) by asking participants to listen to a radio interview in preparation for a questionnaire at the end of the study and to count aloud from 1 to 5 repeatedly throughout the duration of the experiment.

2.2.1 Method

2.2.1.1 Participants

Eighteen Lancaster University students (3 males and 15 females) aged between 18 and 22 years (M = 18.89, SD = 1.02) volunteered to take part in this study in exchange for payment of £3.50 or course credit. All participants were fluent in English, though not necessarily as their first language, with self-reported normal or corrected-to-normal hearing and vision. All experimental procedures were approved by the FST Research Ethics Committee, Lancaster University, on 13th February 2017 (FST16083).

2.2.1.2 Materials

Pseudowords. In total, 98 pseudowords were constructed using a two-syllable, CVCV (C = consonant, V = vowel) form (see Table 2.1 in Appendix A for a full list of pseudowords). Each pseudoword consisted of varying numbers of pointy-sounding and rounded-sounding phonemes. In order to test whether sound-symbolism related to pointiness is appreciated as a graded function, five categories of pseudowords were created: *extreme rounded-*, *rounded-*, *neutral-*, *pointy-* and *extreme pointy-sounding*. Each category was distinguishable by the ratio of pointy- and rounded-sounding phonemes pseudowords contained (see Table 2.2). Recall in Thompson and Estes (2011) experiment, for pseudowords containing an equal number of small- and large-sounding phonemes, participants were more likely to associate them with objects that were considered ‘medium’ in size. It is possible that the combination of these phonemes caused the cross-sensory features of each to compete, leading to participants failing to associate a given pseudoword with one size over another. Therefore, pseudowords designed to be *neutrally* sound-symbolic in this study were created by pairing two pointy-sounding phonemes with two rounded-sounding phonemes.

Table 2.2

Number of rounded-sounding and pointy-sounding phonemes contained within pseudowords according to their assumed phonological pointiness (i.e. extreme rounded, rounded, neutral, pointy, and extreme pointy).

Phonological Pointiness	Number of rounded-sounding and pointy-sounding phonemes		Example /IPA transcription[3]/
	Rounded	Pointy	
Extreme-rounded	4	0	*moru* /mɒruː/
Rounded	3	1	*muri* /mʌriː/
Neutral	2	2	*roti* /rɒtiː/
Pointy	1	3	*betu* /betuː/
Extreme-pointy	0	4	*tebi* /tebiː/

[3] The IPA (International Phonetic Alphabet) provides a standardised representation of the sounds of English language, which has been achieved by assigning each sound its own symbol (e.g. the vowel sound in *loss* and *nod* is represented by the symbol /ɒ/). It is the combination of these symbols which informs the standard pronunciation of a given word.

Research suggests that words containing stop consonants (i.e. sounds that are achieved by blocking and then releasing the flow of air in the vocal tract during articulation) are more likely to be paired with pointier shapes than words containing continuant consonants (i.e. sounds that are achieved by keeping the vocal tract partly open during articulation), which are more likely to be paired with rounder shapes (Westbury, 2005). For the pseudowords in this study, four consonant sounds were included, two *stops* (i.e. /b/ and /t/) and two *continuants* (i.e. /r/ and /m/).

Each pseudoword also contained two of four vowels. Due to the nature of English, vowels can be expressed by a number of different phonemes depending on preceding and succeeding consonant and vowel sounds. That is, the same vowel (e.g. *a*) can be pronounced very differently for different words (e.g. /eɪ/ as in *crane,* /æ/ as in *cat* or /ɑː/ as in *cart*) and, as a result, produced in varying places in the vocal tract (e.g. /eɪ/ is produced nearer the front of the mouth, /æ/ is produced centrally and /ɑː/ is produced towards the back). With research showing that vowels produced nearer the front (as opposed to back) of the mouth are more likely to be associated with pointier (Monaghan, Mattock & Walker, 2012) and smaller (Peña, Mehler & Nespor, 2011; Thompson & Estes, 2011) objects, this variability in vowel phonemes was an important consideration when selecting which vowels to include in our pseudowords. Due to the increased articulatory variability for the vowel *a* compared with other vowels, only *e*, *i*, *o* and *u* are included in our stimulus set. In total, these four vowels produced seven vowel sounds, which are displayed according to their place of articulation (i.e. tongue position relative to the front and roof of the mouth) in Table 2.3.

Table 2.3

Vowel sounds included within pseudowords, identified by tongue positioning during articulation: proximity relative to the front (front/back) and roof (close-open) of the mouth.

Proximity to roof	Front	Back
Close	/iː/ as in *sheep*	/uː/ as in *boot*
Close-mid	/ɪ/ as in *pig*; /e/ as in *bed*	/oʊ/ as in *show*
Open-mid		/ʌ/ as in *cup*
Open		/ɒ/ as in *con*

Fort and Alexander-Peperkamp (2015) explored whether a word's starting letter determined the shape of object that it was associated with. They found that consonant sounds continued to be a significant predictor of pointiness when a word began with either a consonant or vowel. For this reason, the current study uses only a CVCV format. Although seemingly irrelevant to a word's onset, it is still possible that letter positioning might interact with object pointiness in a way that is currently unknown. To control for this, letter positioning within pseudowords was counterbalanced. That is, for every four-letter string generated (e.g. *roti*), three additional pseudowords employing the same letters but in a new sequence (e.g. *tori* [counterbalancing consonants], *rito* [counterbalancing vowels], and *tiro* [counterbalancing consonants and vowels]) were also created. As introduced above, for vowel sounds, counterbalancing in this way can cause phonological changes across words containing the same letters. Importantly for our research, changes to vowel sounds were modest across our pseudowords and, whilst vowels changed their position, their place of articulation (i.e. front or back) was preserved in all cases. For example, when counterbalancing vowels so that *roti* (/rɒti:/) became *rito* (/rɪtoʊ/), in both instances, *o* is produced nearer the back of the mouth and *i* is produced nearer the front. Finally, any pseudowords that ended with the letter *e* (due to its influence on the pronunciation of preceding vowel sounds and its ability to create a single syllable word) and any four-letter string that corresponded (visually or phonetically) with a real word in English (e.g. *tumi* pronounced /tʌmi:/) were eliminated from our stimulus set. During the experiment, pseudowords were presented visually to participants, in lowercase Helvetica typeface.

Rating scale. A Likert type rating scale was composed of 7 points, ranging from 1 (visually 'very pointy') to 7 (visually 'very rounded'). The midpoint on the scale was labelled 'neutral' and successively more extreme points were labelled 'slightly pointy/rounded' and 'pointy/rounded', respectively. The rating scale was presented in Helvetica typeface and scale points were spaced equally, 1cm apart.

Blocking verbal recoding. Participants listened to a BBC 1 radio interview[4] between radio presenter Nick Grimshaw and actor Eddie Redmayne, OBE., dated November 2016. Participants listened to the interview through a pair of Philips SHP1900/00 Stereo Headphones. A short questionnaire containing seven questions about the interview's content was issued at the end of the experiment.

[4] Interview sourced from www.youtube.com: https://www.youtube.com/watch?v=kqKBqQ2mzTU&t=602s

2.2.1.3 Procedure

Participants were presented with a 7-point rating scale (ranging from ‘very pointy’ to ‘very rounded’), which was presented horizontally and centrally on an Apple MacBook (2008) 13-inch computer screen. For each trial, an individual pseudoword appeared directly above the rating scale and participants were asked to rate the pseudoword’s visual pointiness by selecting the integer scale position judged to be most appropriate. Responses were made by clicking on the corresponding scale position with an Apple Mouse cursor. Participants were advised to respond to the visual appearance of the pseudoword only and asked not to dwell on its sound. Once participants had marked their decision, the pseudoword disappeared from the screen and, after a short delay of 2 seconds, a new pseudoword was presented. Throughout the experiment, participants listened to a radio interview and were asked to pay close attention to its content in preparation for a short questionnaire at the end of the study. Participants were also required to count aloud, from 1 to 5, at a steady pace (approximately 2 words/second) and to repeat until the study had finished. Once completed, participants were asked seven questions about the interview’s content. Participants rated the visual appearance of all pseudowords contained in the stimulus set, with the order of pseudowords randomly generated for each participant. Each participant completed 98 trials and the whole procedure took approximately 20 minutes. The experiment was executed using PsyScript (Slavin, 2014).

2.2.2 Results and discussion

2.2.2.1 Overview of analyses

For this and all other experiments reported in this chapter, the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) in R (R Core Team, 2013) was used to perform linear mixed effects analyses exploring the relationship between a given dependent variable and controlled (fixed) and uncontrolled (random) independent variables. The general strategy for creating the models reported in this chapter was to include all fixed and random effects as a first instance and then to eliminate those that reduced the model's overall goodness of fit. This was achieved by comparing each model's Akaike Information Criterion (AIC) using the *AICcmodavg* package (Mazerolle, 2019) in R, which estimates the quality of a statistical model for a given set of data (with a lower AIC value indicating a higher quality model). Once a model reached its optimum (i.e. lowest) AIC, we identified which independent variable/s were having a significant effect on the dependent variable. To do this, *p*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. For all statistical tests, an alpha level of .05 was used as a significance criterion. Unless reported, visual inspections of residual plots did not reveal any obvious deviations from homoscedasticity or normality in any of the linear mixed effects analyses reported in this chapter.

2.2.2.2 Results

The dependent measure was the rating of each pseudoword's visuostructural pointiness, ranging from 1 (very pointy) to 7 (very rounded). In order to explore the association between sound-symbolism and visual-symbolism, a score between 1 and 5 was assigned to each pseudoword, which corresponded with the phonological pointiness category pseudowords were assumed to belong to (1 = extreme-rounded, 2 = rounded, 3 = neutral, 4 = pointy, and 5 = extreme-pointy). In the models reported in this and all other experiments in this chapter, this score is represented by the fixed effect, *phonological pointiness.*

As a fixed effect, *phonological pointiness* was included in the model, with an intercept for participants included as a random effect. Analysis revealed that *phonological pointiness* significantly affected judged *visuostructural pointiness*, $\chi^2(1) = 239.46$, $p < .001$, with each step increase in *phonological pointiness* raising ratings of *visuostructural pointiness* by .57 ($SE = .04$). That is, as pseudowords increased in phonological pointiness, participants rated pseudowords as being progressively pointier in visuostructural appearance (for means and standard error of the means see Table 2.4).

2.2.2.3 Discussion

Whilst the verbal recoding of pseudowords was blocked, the results illustrate the linear relationship that exists between a pseudoword's sound and its graphemic structural appearance. Pseudowords designed to be symbolic of pointiness in terms of their sound, were also judged to be symbolic of pointiness in terms of their visual appearance: pointier sounding pseudowords were rated as being visually pointier in appearance. These findings might be interpreted in one of three ways: 1. pointier sounds in pseudowords are represented visually as being pointy in appearance, 2. our attempt to block verbal recoding in the present experiment was not entirely effective (i.e. participants were verbally recoding pseudowords), or 3. judged symbolic pointiness is based on a combination of sound and visual properties, so that pseudowords that both sound and look pointier (or rounder) are judged as being symbolic. In this case, successfully blocking the verbal recoding of pseudowords would continue to yield a significant relationship between phonemic sound features and graphemic structural features. The effectiveness of our efforts to block verbal recoding in the present experiment is addressed in section 2.3.2.1 of Chapter 2.

2.3 Experiment 2: Relationship between phonology and graphemes (verbal recoding not blocked)

Similar to Experiment 1, the objective of Experiment 2 was to explore whether the presence of sound-symbolism predicts the presence of visual-symbolism by identifying which psuedowords in our stimulus set are judged as being visually pointier in terms of their graphemic-structure (i.e. visual appearance in printed form). Again, participants in this experiment were required to rate the visual appearance of each pseudoword on a 7-point Likert-type scale ranging from 1 (very pointy) to 7 (very rounded). However, unlike in Experiment 1, participants were free to verbally recode the pseudowords presented to them. By comparing the results of Experiment 1 with the results of the present experiment, one aim of this experiment is to explore how the cross-sensory features of words (i.e. their sound and visual properties) combine (additively or interactively) to predict symbolic pointiness.

2.3.1 Method

2.3.1.1 Participants

Twenty-two Lancaster University students (3 males and 19 females) aged between 18 and 27 years (M = 20.39, SD = 2.3) volunteered to take part in this study in exchange for payment of £3.50 or course credit. All participants were fluent in English, though not necessarily as their first language, with self-reported normal or corrected-to-normal hearing and vision.

2.3.1.2 Materials

The same pseudowords and rating scales from Experiment 1 were used in this experiment.

2.3.1.3 Procedure

Participants completed the same rating task as Experiment 1, during which they rated all pseudowords for their visual appearance on a scale ranging from 1 visually ‘very pointy’ to 7 visually ‘very rounded’. In this experiment, participants were free to verbally recode the pseudowords. Therefore, participants were not required to listen to a radio interview throughout the experiment, nor count aloud from 1 to 5. All other aspects of the procedure remained the same as Experiment 1.

2.3.2 Results and discussion

The dependent measure was the ratings of each pseudoword’s visuostructural pointiness, ranging from 1 (very pointy) to 7 (very rounded). To begin, linear mixed effects analyses were performed to explore 1. the association between phonological pointiness and ratings of visuostructural pointiness in the present experiment, 2. whether ratings of visuostructural pointiness from Experiment 1 predicted ratings of visuostructural pointiness in the present experiment, and 3. whether the effects of phonological pointiness interacts with the effects of visuostructural pointiness. Later in this section, the effectiveness of our efforts to block verbal recoding in Experiment 1 is explored (see section 2.3.2.1 of Chapter 2). The strategy for creating models in this experiment was the same as Experiment 1 (see section 2.2.2.1 of Chapter 2 for details). As fixed effects, *phonological pointiness, Experiment 1 visuostructural* *pointiness ratings,* an interaction term for *phonological pointiness*Experiment 1 visuostructural pointiness ratings* and *pseudoword presentation order*[5] were entered into the model, including an intercept for participants as a random effect.

First, analysis revealed that *phonological pointiness* significantly affected *visuostructural pointiness*, $\chi^2(2) = 33.02$, $p < .001$, with each step increase in *phonological pointiness* raising ratings of *visuostructural pointiness* by .16 ($SE = .15$). That is, as a pseudoword increased in phonological pointiness, participants rated pseudowords as being progressively pointier in visuostructural appearance (for means and standard errors see Table 2.4).

---

[5] Whilst *pseudoword presentation order* improved the overall fit of the model, pseudowords were presented randomly to participants and so its effect on the dependent variable is not reported in this or any of the following experiments included in this chapter.

Table 2.4

For each level of phonological pointiness (extreme rounded, rounded, neutral, pointy, extreme pointy), mean ratings for visuostructural pointiness in Experiment 1 and Experiment 2 (min = 1 ['very pointy'], max = 7 ['very rounded']) are reported (with standard error in parenthesis).

Phonological pointiness	Experiment 1	Experiment 2
	M (*SE*)	*M* (*SE*)
(1) Extreme-rounded	5.93(.16)	5.8(.14)
(2) Rounded	5.10(.06)	5.05(.05)
(3) Neutral	4.44(.05)	4.19(.04)
(4) Pointy	4.04(.08)	3.79(.07)
(5) Extreme-pointy	3.58(.23)	3.23(.2)

Next, we explored the extent to which ratings of visuostructural pointiness were similar across Experiment 1 and Experiment 2. Analysis revealed that *Experiment 1 visuostructural pointiness ratings* significantly affected *visuostructural pointiness*, $\chi^2(2) = 657.55$, $p < .001$, with each step increase in *Experiment 1 visuostructural pointiness ratings* raising ratings of *visuostructural pointiness* by .81 ($SE = .11$). That is, as ratings of visuostructural pointiness in Experiment 1 increased in pointiness, participants rated pseudowords as being progressively pointier in visual appearance in Experiment 2. A Spearman's rank-order correlation confirmed a moderate, positive correlation between these two variables, $r_s(2154) = .53$, $p < .001$.

Finally, we identified whether the association between phonological pointiness and visuostructural pointiness was interactive in nature (i.e. the effect of a pseudoword's sound on judged symbolic pointiness depends on its visual appearance or vice-versa). Analysis revealed that the interaction between *phonological pointiness* and *Experiment 1 visuostructural pointiness ratings* on ratings of *visuostructural pointiness* in the present experiment was not significant, $\chi^2(1) = .04$, $p = .85$.

2.3.2.1 Effectiveness of verbal recoding blocking

One way of identifying the effectiveness of our efforts to block verbal recoding in Experiment 1 is to isolate the impact of each phoneme on ratings of visuostructural appearance. In order to do this, all pseudowords containing a particular vowel phoneme[6] were grouped and included as separate fixed effects each in their own linear mixed effect model. Each model also included a fixed effect for *pseudoword presentation order* and an intercept for participants as a random effect. The results of each linear mixed effect analysis are displayed collectively in Table 2.5. Unlike in Experiment 1 where individual vowel phonemes had no effect on a pseudoword's likelihood of being rated as pointier or rounder in visual appearance, for the present experiment, pseudowords containing the vowel phonemes /ɪ/ and /i:/ were rated as being significantly pointier in visual appearance and pseudowords containing all other vowel phonemes (i.e. /e/, /u:/, /ʌ/, /oʊ/ and /ɒ/) were rated as being significantly rounder in visual appearance.

[6] For vowel sounds in the English language, letters represent a number of different phonemes (e.g. /i/ can be pronounced /ɪ/ as in *pig*, /aɪ/ as in *child*, /ə/ as in *pupil*, /i:/ as in *physique*, /j/ as in *onion*, or /ʃ/ as in *anxious*). However, for nearly all consonants, letters represent a single phoneme. As a result, it is not possible to study the impact that a consonant's phoneme has on ratings of *visuostructural pointiness* separately from it visuostructural appearance (i.e. letter). Therefore, only vowel phonemes are reported here.

Table 2.5

Parameter estimates (b) (with standard error in parenthesis) and chi-square (χ^2) values for fixed effects in the linear mixed effects analysis related to vowel phonemes for Experiment 1 and Experiment 2.

Vowel Phonemes	Experiment 1		Experiment 2	
	b(SE)	χ^2	*b(SE)*	χ^2
/ɪ/	.12(.07)	2.85	-.9(.07)	183.98***
/e/	-.07(.07)	1.07	.18(.06)	8.74**
/u:/	-.001(.001)	.37	.15(.06)	6.06*
/ʌ/	-.04(.08)	.26	.23(.07)	10.48**
/i:/	.03(.07)	.24	-.74(.06)	141.88***
/oʊ/	.009(.07)	.02	.55(.06)	80.35***
/ɒ/	-.009(.08)	.01	.51(.07)	53***

* $p < .05$, ** $p < .01$, *** $p < .001$, otherwise non-significant

2.3.2.2 Discussion

In summary, the results echo the same linear relationship between a pseudoword's sound and its graphemic structural appearance that was observed in Experiment 1, with pointier sounding pseudowords being rated as pointier in visual appearance. The relationship between sound-symbolism and visual-symbolism was found not to be interactive in nature, suggesting that these features combine additively to predict a pseudoword's symbolic pointiness. Unlike Experiment 1 in which the verbal recoding of pseudowords was blocked, in the present experiment participants were free to verbally recode the pseudowords presented visually. In order to identify whether verbal recoding was effectively blocked in Experiment 1, the impact of individual vowel phonemes on ratings of visual appearance were assessed for Experiment 1 and the present experiment. In the present experiment, participants rated pseudowords containing vowel sounds produced nearer the front of the mouth[7] as being visually pointier in appearance than pseudowords containing vowel sounds produced nearer the back of the mouth. This is similar to research showing a correspondence between vowel sounds produced nearer the front of the mouth and objects that were smaller in size (Peña, Mehler & Nespor, 2011). These findings indicate that ratings of visuostructural pointiness in the present experiment were mediated (but to what degree is unknown) by a pseudoword's sound. However, for Experiment 1, individual vowel phonemes had no effect on ratings of a pseudoword's visual appearance, suggesting that our efforts to block verbal recoding were effective in this case. For this reason, we conclude that Experiment 1 provides a more reliable measure of visuostructural pointiness than the present experiment, despite the results from both experiments being highly correlated.

2.4 Preparation for Experiment 3 – 4

In Experiment 3 and Experiment 4, participants were required to pair pseudowords with objects representing varying degrees of visual pointiness. But how appropriate were these objects for representing gradations of this kind? For another purpose, tangible versions of the same objects were rated on verbal scales representing a selection of the perceptual qualities comprising the core set of cross-sensory correspondences in the literature (i.e. sharpness, pitch, thinness, speed,

---

[7] The exception of this was for the vowel phoneme /e/, which, when present within a pseudoword increased the likelihood of a pseudoword being rated as rounder (rather than pointier) in visual appearance.

visuospatial height, weight, brightness and size). Whilst the objects were presented visually to participants in Experiment 3 and Experiment 4, in this experiment participants explored the objects by touch alone. The advantage of gathering ratings of cross-sensory perceptual qualities through touch rather than through vision, is that touch allows for ratings of perceived brightness as a cross-sensory feature, in addition to all others. Had the objects been presented visually during this rating task, participants would have been informed of their brightness and so expectations of an object's brightness based on its pointiness would have been void. Importantly, research concerned with touch as the encoding channel for cross-sensory correspondences has found that the same correspondences related to pointiness (Barnett, Bremner & Walker, submitted), size (Walker, Walker & Francis, 2012) and height (Nava, Grassi & Turati, 2016) can exist regardless of source modality. In other words, correspondences are experienced similarly across different encoding channels (e.g. vision or touch). Therefore, participants' ratings of these objects (despite being encoded tactually) can be considered synonymous to visual versions of these same objects. In this section, the details of the rating study are presented in full. In the original experiment, it was predicted that pointier objects experienced by touch, relative to rounder objects experienced by touch, would be judged as being sharper, higher in pitch, thinner, faster, higher in visual space, lighter in weight and brighter.

2.4.1 Method

2.4.1.1 Participants

Thirty undergraduate students (2 males, 28 females) aged between 18 and 25 years (*M* = 18.72, *SD* = 1.39) from Lancaster University volunteered to take part in the study in exchange for course credit. They were all fluent in English, though not necessarily as their first language. All experimental procedures were approved by the Department of Psychology Ethics Committee of Lancaster University on 28th July 2015.

2.4.1.2 Materials

Objects. Three novel objects were designed using Autodesk Inventor Professional (2015) software and manufactured using a Modela MDX-15 3-dimensional printer. The objects were coded from the pointiest (object 1) to the roundest (object 3) (see Figure 2.6). Each object had five 'points', whose edges were altered to yield different degrees of tactile pointiness. This was achieved by adjusting and matching the corner radiuses on all edges (convex and concave) of the objects, which increased from 0mm (object 1) to 8mm (object 3), in steps of 4mm.

Figure 2.6. The novel objects attached to individual bases. In order from the pointiest object (object 1) on the left to the roundest object (object 3) on the right.

Increasing the corner radius of objects like these, whilst maintaining their volume (i.e. amount of material), results in a reduction of object size, such that objects with a larger corner radius have a smaller footprint. Given the corner radius ranges from 0mm (object 1) to 8mm (object 3), the objects in the stimulus set decreased in size, from object 1 to object 3. To address this issue and avoid a confound between overall size and pointiness, each object was scaled up so that its footprint fit snugly in to an 80mm (width) x 80mm (length) x 20mm (depth) box. One consequence of increasing an object's size in this way, is that the volume (and, consequently, mass) of the object is increased also (see Table 2.6). It is likely that an object's size would be more easily detected by touch than would an object's volume, hence the decision to ensure that the objects were matched in terms of size despite variations in volume. To ensure that mass could not be encoded directly, the objects were fixed to the table throughout the experiment, preventing participants from lifting and manipulating them. The objects were also covered with an opaque piece of material, measuring 73 x 47 cm.

Table 2.6

The corner radius in millimeters (mm), volume in cubic centimeter (cm^3) and mass in grams (g), for each object.

Object	Corner Radius (mm)	Volume (cm^3)	Approx. Mass (g)
1 (pointiest)	0	48.66	23
2	4	58.27	26
3 (roundest)	8	64.69	28

Rating scales. Eight scales defined by antonym word pairs were used to represent the perceptual qualities regularly evident in cross-sensory correspondences, that is, *blunt-sharp*, *low-pitch-high-pitch*, *thick-thin*, *slow-fast*, *low in space-high in* *space*, *heavy-light*, *dark-bright* and *big-small*. The six major points on each scale were labelled with VERY, QUITE, or SLIGHTLY, so that, for instance, the sequence was VERY BLUNT, QUITE BLUNT, SLIGHTLY BLUNT, SLIGHTLY SHARP, QUITE SHARP, and VERY SHARP. An additional 5 unlabelled minor points were included to allow for ratings positioned between two major points.

2.4.1.3 Procedure

After giving consent, participants were allowed to familarise themselves with the three objects by placing their dominant hand underneath the cloth covering the objects and feeling each object. The objects could not be lifted and their position relative to the participant was indicated by the numbers 1 to 3 running left to right on the table. These numbers also served as an indicator to participants as to which object they should rate on a given scale. Once participants were familiarized with the objects, they proceeded to rate them on the full set of scales. The scales were presented to participants in the form of a paper booklet and participants indicated their decision by manually marking the most appropriate scale position. Participants were informed that they could return to the objects at any point throughout the experiment, and as many times as necessary to complete the task. In a fully counterbalanced, within-participants design, participants rated all objects on all scales, providing a total of 24 judgements. Whether the left-right positioning of the objects went from roundest to pointiest, or pointiest to roundest, was counterbalanced across participants, as was the scale's direction defined by the left-right positioning of the antonym pair. Participants rated the objects with a randomly determined order for both object and scale. The whole procedure took approximately 15 minutes.

2.4.2 Results and discussion

The dependent measure was the rating of each object, ranging from 1 (very blunt, low-pitch, thick, slow, low in space, heavy, dark and big) to 6 (very sharp, high-pitch, thin, fast, high in space, light, bright and small). For purpose of statistical analysis, each object (rounded, neutral and pointy) was assigned a value of 1, 2 or 3, respectively. Kendall's *tau* was employed as a non-parametric index of the degree to which two variables were similarly ordered according to known cross-sensory correspondences (i.e. pointier objects are expected to be sharper, higher in auditory pitch, thinner, faster, higher in visual space, lighter in weight, brighter and smaller in size). Tied values were permitted and taken into account. Mean values for *tau* could range from 0 to 3, with 0 reflecting a perfectly negative association (i.e., pointier objects being rated as rounder, lower in auditory pitch, thicker, slower, lower in visual space, heavier in weight, darker and bigger in size) and 3 reflecting a perfectly positive association (i.e. pointier objects being rated in accordance with known cross-sensory correspondences as outlined above). Were participants to fail completely to associate the two variables, the expected (null) value for *tau* would be 1.5. The ratings of each object on each scale were also submitted to a repeated measures ANOVA. The results are displayed in Table 2.7.

For the top 7 antonym pairs in Table 2.7, the effect of object pointiness was significant. In every case, a positive association between object pointiness and ratings was recorded, with pointier objects judged to be sharper, higher-pitch, thinner, faster, higher in space, lighter, and brighter. For the antonym pair *big – small*, the effect of object pointiness was not significant. That is, the ratings for each object did not differ significantly from one another on the scale referring to perceived size, despite objects varying in volume. These findings confirm the appropriateness of these objects as representing visual pointiness in Experiment 3 and Experiment 4, whilst also illustrating the transitive nature of cross-sensory correspondences as discussed in section 1.3 of Chapter 1. That is, visual and tactile features that are typically associated with auditory pitch, converge with one another (and in the same direction) to reveal relationships between tactile pointiness and sharpness, thinness, speed, visuospatial height, weight and brightness.

Table 2.7

The mean rating of each object (with standard error in parenthesis) on each of the 6-point scales. Scores below and above 3.5 reflect an association with the left and right antonym within a pair (as displayed in the table), respectively. In the final two columns, mean values for Kendall's tau (min = 0, max = 3, null = 1.5) and F are reported.

Scale	Object			Mean *tau*	*F*
	Rounded	Neutral	Pointy		
blunt - sharp	5.77(.08)	4.2(.15)	1.67(.11)	2.97***	706.51***
low-pitch - high-pitch	4.65(.21)	3.77(.17)	2.2(.21)	2.53***	53.46***
thick - thin	4.58(.18)	3.98(.14)	3.03(.25)	2.37***	28***
slow - fast	4.28(.27)	3.48(.15)	3.03(.31)	2.08**	6.62*
dark - bright	4.15(.28)	3.3(.18)	3(.31)	2.13**	5.37*
heavy - light	3.52(.27)	3.08(.19)	2.53(.26)	2.07**	5.54*
low in space - high in space	4.18(.26)	3.47(.17)	3.08(.3)	2.05*	6.01*
big - small	3.65(.24)	3.55(.18)	3.35(.21)	1.78	1.06

* $p < .05$, ** $p < .01$, *** $p < .001$, otherwise not significant

2.5 Experiment 3: Graded effects of symbolic pointiness

Having established the additive (rather than interactive) nature of phonemic and graphemic structural features that combine to predict a word's symbolic pointiness in Experiment 2, the purpose of Experiment 3 was to explore whether these features allow for graded effects in sound-symbolism. In other words, do pseudowords in our stimulus set symbolise a full range of intermediate levels of pointiness (perhaps with a ‘neutral’ midway level), rather than just the two extremes (i.e. pointy or rounded)? To test this, by listening to the pronunciation of our pseudowords only, participants paired each pseudoword based on its sound with one of five visually presented objects, ranging from pointiest to roundest.

2.5.1 Method

2.5.1.1 Participants

Thirty Lancaster University students (15 males and 15 females) aged between 18 and 36 years (M = 21.32, SD = 3.78) volunteered to take part in this study in exchange for payment of £3.50 or course credit. All participants were fluent in English, though not necessarily as their first language, with self-reported normal or corrected-to-normal hearing and vision.

2.5.1.2 Materials

Pseudowords. The pseudowords were the same as in Experiment 1. In this experiment, pseudowords were only presented auditorily to participants. In some instances, pseudowords deviated from their standard pronunciation, for example, *roti* might more commonly be pronounced /rəʊti:/ (rhymes with *floaty*) but, for the purpose of this research, was pronounced /rɒti:/ (rhymes with *potty*) by our speaker. Here, the latter pronunciation ensures a better example of back-open (/ɒ/) and front-close (/i:/) vowel sounds within a single pseudoword for our *neutral-sounding* category. Thus, by choosing a less obvious (but acceptable) pronunciation, a tighter control is achieved. Each pseudoword was transcribed phonetically and presented to a speaker, who was blind to the research aims, to read aloud. The speaker was female, aged 51 years and spoke English as a first language with a slight northern English accent. The speaker was instructed to read each pseudoword once, in a natural and clear manner. These utterances were recorded in a silent room using a MacBook laptop (2008) microphone and clipped into individual sound files using PRAAT sound analysis software (Boersma & Weenink, 2017). Clipped sound files varied in duration from .56 seconds to .79 seconds (M = .68, SD = .05). Due to the nature of our stimulus set (including only 8 letters yielding a total of 11 phonemes), there was the potential for participants to have recognised that pseudowords were comprised of a finite number of consonant and vowel sounds in a new configuration each time. To reduce the detection of any such patterns, the full list of pseudowords was divided equally so that each participant rated half of the words in our stimulus set[8].

Objects. The novel objects representing varying degrees of visual pointiness were 2-dimensional, visual versions of those presented in section 2.4 of Chapter 2. The objects were designed using Autodesk Inventor Professional (2015) software and coded from the pointiest to the roundest (see Figure 2.7). To provide a greater range of visual pointiness for the present experiment, an additional two objects were included (object 2 and object 4). The design of the objects was identical to that presented in section 2.4 of Chapter 2, except the corner radiuses on all edges of the

[8] Across all five categories of words, there was an uneven distribution of pseudowords. For instance, for the *neutral-sounding* category 46 pseudowords were generated, whereas, for the *extreme rounded-sounding* category, it was only possible to generate a total of 4. This is due to the limited quantity of rounded-sounding and pointy-sounding letters, of which 4 of each were included. In order to generate pseudowords that are considered 'extreme' examples of roundedness and pointiness, only phonemes that were typically associated with one shape (e.g. pointy) over another (e.g. rounded) were included within a given word. As such, our stimulus set was limited by a finite number of letter combinations. For this reason, all *extreme rounded-sounding* and *extreme pointy-sounding* pseudowords were presented to all participants. The remaining 3 categories: *rounded-*, *neutral-* and *pointy-sounding* pseudowords were divided equally between two groups of participants as discussed above.

objects increased from 0mm (object 1) to 8mm (object 5) in steps of 2mm (rather than 4mm). Again, the objects were scaled so that their footprint fit snugly in to an 80mm (width) x 80mm (length) x 20mm (depth) box. The objects were coloured pale grey (RGB value[9]: 227, 226, 226; HSB value[10]: 0°, 1%, 89%) and presented against a plain, white background.

Figure 2.7. The five novel shapes representing a progression of visual pointiness, in order from object 1 on the left (pointiest) to object 5 on the right (roundest).

2.5.1.3 Procedure

Each participant was presented with the visual display of the five novel objects, which were positioned in order, horizontally and centrally on an Apple MacBook (2008) 13-inch computer screen. Across participants the ordering of the five objects (from pointiest to roundest and vice-versa) was counterbalanced. Participants listened to the pronunciation of each pseudoword via a pair of Philips

---

[9] RGB (stands for Red Green Blue). A precise shade of colour is represented by the accumulation of individual Red, Green and Blue values ranging from 0 to 255.
[10] HSB (stand for Hue Saturation Brightness). A colour's *hue* is represented by a value ranging from 0 to 360 degrees and its *saturation* and *brightness* are represented as a percentage from 0 – 100%. For *saturation*, 0% represents no colour and 100% represents full colour. For *brightness*, 0% represents black and 100% represents white. Thus, the closer a *brightness* value is to 100% the brighter the shade.

SHP1900/00 Stereo Headphones. For each participant, the presentation order of the pseudowords was randomly generated. Across each trial, the five novel objects remained fixed and the presentation of each pseudoword marked the onset of a trial. Participants were instructed to pair each pseudoword with one of the five objects presented and advised to respond to the *sound* of each word in order to make their selection. Responses were made by clicking on the appropriate object with an Apple Mouse cursor. Once a choice had been made, the trial ended and, after a short delay of 2 seconds, a new pseudoword was presented. Participants completed a total of 52 trials and the whole procedure took approximately 15 minutes.

2.5.2 Results and discussion

The dependent measure was the visual pointiness of the object selected to be paired with each pseudoword, ranging from 1 (pointiest) to 5 (roundest), and for the model reported in this experiment is represented by the fixed effect, *object pointiness*. Linear mixed effects analyses were performed to explore 1. whether phonological pointiness predicted pseudoword-object pairings (with pointier sounding pseudowords being paired with visually pointier objects), 2. whether ratings of visuostructural pointiness from Experiment 1 predicted pseudoword-object pairings (with visually pointier pseudowords being paired with visually pointier objects), and 3. whether phonological pointiness interacts with ratings of visuostructural pointiness to predict pseudoword-object pairings. The ratings of visuostructural pointiness included in the analyses reported here are taken from Experiment 1, where participants rated the visual appearance of each pseudoword on a 7-point scale ranging from 1 visually ‘very pointy’ to 7 visually ‘very rounded’ whilst the verbal recoding of pseudowords was blocked (i.e. phonological features were made unavailable). The results of

Experiment 2 showed that our efforts to block verbal recoding was effective in Experiment 1 (see section 2.3.2.1 of Chapter 2) and so ratings of visuostructural pointiness were considered uncontaminated by phonological features, hence the decision to include visuostructural pointiness ratings from Experiment 1 rather than Experiment 2 (where participants were free to verbally recode pseudowords). The strategy for creating models in this experiment was the same as Experiment 1 (see section 2.2.2.1 of Chapter 2 for details). As fixed effects, *phonological pointiness, visuostructural pointiness ratings,* an interaction term for *phonological pointiness*visuostructural pointiness ratings* and *pseudoword presentation order* were entered into the model, including an intercept for participants as a random effect.

First, we identified whether phonological pointiness predicted pseudoword-object pairings. Analysis revealed that *phonological pointiness* significantly affected *object pointiness*, $\chi^2(2) = 32.78$, $p < .001$, with each step increase in *phonological pointiness* raising ratings of *object pointiness* by .28 (*SE* = .16). That is, as a pseudoword's sound increased in pointiness, participants paired pseudowords with objects that were pointier in visual appearance (for means and standard errors see Table 2.8).

Next, we explored the relationship between visuostructural pointiness ratings in Experiment 1 and pseudoword-object pairings. Analysis revealed that *visuostructural pointiness ratings* significantly affected *object pointiness*, $\chi^2(2) = 56.43$, $p < .001$, with each step increase in *visuostructural pointiness ratings* raising ratings of *object pointiness* by .33 (*SE* = .11). That is, as ratings of a pseudoword's visuostructural pointiness in Experiment 1 increased in pointiness, participants paired pseudowords with objects that were pointier in appearance.

Finally, we explored whether an interaction between phonological pointiness and visuostructural pointiness ratings in Experiment 1 predicted pseudoword-object pairings. Analysis revealed that the interaction between *phonological pointiness* and *visuostructural pointiness ratings* on pseudoword-object pairings in the present experiment was not significant, $\chi^2(1) = .19$, $p = .66$.

In summary, these findings illustrate the linear relationship that exists between a pseudoword's sound and the visual pointiness of objects that it is associated with. In short, pointier sounding pseudowords were more likely to be paired with visually pointier objects, demonstrating that words can possess cross-sensory features in gradations. In this experiment, pseudowords were presented auditorily only, yet we found a strong correlation between the pointiness of objects paired with pseudowords and the pseudowords' previous ratings of visuostructural pointiness from Experiment 1 (with pseudowords rated as pointier in visual appearance being paired with visually pointier objects in this experiment). It is possible that the phonemes included in our set of pseudowords naturally correlated with their visual appearance. For example, the letters *t* and *i* were judged as being pointier than the letters *m* and *o* in terms of their sound and their visual appearance. This proposal is discussed further in section 2.7 of Chapter 2.

2.6 Experiment 4: The impact of typeface on symbolic pointiness

The experiments presented thus far have demonstrated the influence a word's sound and its visual appearance can have on ratings of symbolic pointiness, with pointier sounding pseudowords being rated as pointier in visual appearance and paired with visually pointier objects. The findings suggest that phonemic and visual properties of a word combine additively (rather than interactively) and that the resulting symbolic pointiness is a graded function, with the number of pointy-sounding phonemes present within words linearly predicting the shape (i.e. pointiness) of objects that words are associated with. In Experiment 1 and Experiment 2, participants rated the visual appearance of each pseudoword in our stimulus set at a graphemic-structural level. However, words can vary in their visual appearance in ways other than their graphemic-structure, such as typeface. Whilst considered a surface feature of written language (in the sense that it does not contribute to a word's structural appearance or indeed its strict meaning), some typefaces have been found to possess cross-sensory features which can interfere with how participants respond to printed words (Walker, 2016). For example, the typeface *Palatino Italic* was judged to be higher-pitch, pointier, thinner, lighter in weight, smaller, faster and brighter than the typeface **Cooper Black**, and when participants were asked to classify words (e.g. *squeal*) according to the auditory pitch of their referents, Walker found that participants performed faster and more accurately in the task when words were presented in a pitch-congruent typeface (e.g. *Palatino Italic* for higher-pitch and **Cooper Black** for lower-pitch) rather than a pitch-incongruent typeface (i.e. *Palatino Italic* for lower-pitch and **Cooper Black** for higher-pitch).

The aim of Experiment 4 is to explore whether the cross-sensory features of different typefaces predicts a word's symbolic pointiness. The advantage of exploring the effect of typeface in this context, is that, unlike the graphemic-structure of a word, typeface can be fully detached from a word's sound. In other words, the sound of a letter is uncontaminated by the typeface in which it is presented. For instance, *pig* is pronounced /pɪg/ whether it is presented in Palatino Italic (*pig*) or Cooper Black (**pig**). So far, the findings from Experiment 1 - 3 have been uncertain as to whether participants are responding to a pseudoword's visual or sound features (or both) when making judgements about a pseudoword's symbolic pointiness. By introducing varying typefaces to our investigation, we will be able to confirm whether a pseudoword's visual appearance is having an effect in this context. We predict that pseudowords presented in a visually pointier typeface (as opposed to a visually rounder typeface) will be associated with visually pointier objects.

In the present experiment, participants paired each pseudoword in our stimulus set with one of five visually presented objects, ranging from pointiest (object 1) to roundest (object 5). Unlike in Experiment 3, where pseudowords were presented auditorily only, in this experiment pseudowords were presented both auditorily and visually, half in a visually pointy typeface (*Palatino Italic*) and half in a visually rounded typeface (**Cooper Black**).

2.6.1 Method

2.6.1.1 Participants

Sixty-four Lancaster University students (19 males and 45 females) aged between 18 and 36 years ($M = 21.32$, $SD = 3.78$) volunteered to take part in this study in exchange for payment of £3.50 or course credit. All participants were fluent in English, though not necessarily as their first language, with self-reported normal or corrected-to-normal hearing and vision.

2.6.1.2 Materials

Pseudowords. The pseudowords were the same as in Experiment 1. In this experiment, participants listened to the pronunciation of each pseudoword as well as being presented, visually, with each pseudoword in printed form. For auditory presentation, the pseudowords were the same as in Experiment 3. For visual presentation, pseudowords were presented in lowercase and displayed in either **Cooper Black** or *Palatino Italic* typeface. Each typeface differed in overall size (despite font size being consistent across words), so that the same pseudoword was wider when presented in **Cooper Black** than when it was presented in *Palatino Italic* (e.g. **mebu** vs *mebu*) (see Table 2.9 for a full list of character width and heights). Given the convergence of size and pointiness in cross-sensory research (for example, smaller and pointier objects are judged to be higher in auditory pitch, for a review see Spence, 2011), it was imperative that the presentation of pseudowords in each typeface was counterbalanced. To achieve this, pseudowords that shared the same starting letter were divided into two equal groups. For one group of participants, half of the pseudowords starting with the letters *b*, *m*, *r* and *t* were presented in **Cooper Black** and the other half presented in *Palatino Italic*. The presentation of **Cooper Black** and *Palatino Italic* was then counterbalanced across participants, so that a pseudoword presented in **Cooper Black** to one group of participants was presented in *Palatino Italic* to another. For each trial, pseudowords appeared individually and centered on a screen.

Objects. The objects were the same as in Experiment 3.

Table 2.9

*For each typeface (**Cooper Black** and Palatino Italic), the height and width of each character (measured in millimeters [mm]) when displayed in identical font sizes are reported.*

Character	**Cooper Black**		*Palatino Italic*	
	Height (mm)	Width (mm)	Height (mm)	Width (mm)
b	13	12	14	7
m	9	17	9	15
r	9	10	9	7
t	12	9	12	5
e	9	9	9	6
i	13	9	12	7
o	9	11	9	7
u	9	12	9	10

2.6.1.3 Procedure

Participants completed the same pseudoword-object matching task as used in Experiment 3, during which they paired individual pseudowords with one of five objects based on the pseudoword's sound. In this experiment, pseudowords were presented auditorily and visually. For visual presentation, pseudowords were displayed centrally on the screen, directly above the five shapes. Visual and auditory presentation of each pseudoword was simultaneous, so that when a pseudoword appeared on the screen, participants listened to its pronunciation via a pair of Philips SHP1900/00 Stereo Headphones. Across participants, the ordering of the five shapes (from pointiest to roundest and vice-versa) was counterbalanced, along with the typefaces pseudowords were presented in. All other aspects of the procedure remained the same as in Experiment 3.

2.6.2 Results and discussion

The dependent measure was the object selected to be paired with each pseudoword, ranging from 1 (pointiest) to 5 (roundest), and for the model reported in this experiment is represented by the fixed effect, *object pointiness*. Linear mixed effects analyses were performed to explore 1. whether a pseudoword's typeface predicted pseudoword-object pairings (with pseudowords presented in a pointier typeface being paired with visually pointier objects), 2. whether phonological pointiness predicted pseudoword-object pairings (with pointier sounding pseudowords being paired with visually pointier objects), 3. whether ratings of visuostructural pointiness from Experiment 1 predicted pseudoword-object pairings (with visually pointier pseudowords being paired with visually pointier objects), and 4. whether a pseudoword's typeface interacts with ratings of visuostructural pointiness from Experiment 1 to predict pseudoword-object pairings. Like Experiment 3, the ratings of visuostructural pointiness included in the analyses reported here are taken from Experiment 1. The strategy for creating models in this experiment was the same as Experiment 1 (see section 2.2.2.1 of Chapter 2 for details). As fixed effects, *phonological pointiness, typeface, visuostructural pointiness ratings,* an interaction term for *typeface*visuostructural pointiness ratings* and *pseudoword presentation order* were entered into the model, including an intercept for participants as a random effect.

First, we identified whether typeface predicted pseudoword-object pairings. Analysis revealed that *typeface* significantly affected *object pointiness*, $\chi^2(2) = 38.72$,

$p < .001$, with each step increase in *typeface* raising ratings of *object pointiness* by .56 (SE = .2). That is, for pseudoword's presented in a visually pointier typeface (i.e. *Palatino Italic*, M = 2.95, SD = 1.21) compared with a visually rounder typeface (i.e. **Cooper Black**, M = 3.18, SD = 1.15) participants paired pseudowords with objects that were pointier in appearance.

Next, we identified whether phonological pointiness also predicted pseudoword-object pairings as demonstrated by Experiment 3. Analysis revealed that *phonological pointiness* significantly affected *object pointiness*, $\chi^2(1) = 10.75$, p = .001, with each step increase in *phonological pointiness* raising ratings of *object pointiness* by .08 (SE = .03). That is, as a pseudoword's sound increased in pointiness, participants paired pseudowords with objects that were pointier in visual appearance (for means and standard errors see Table 2.8).

Table 2.8

For each level of phonological pointiness (extreme rounded, rounded, neutral, pointy, extreme pointy), mean ratings for object pointiness in Experiment 3 and Experiment 4 (min = 1 [most pointy], max = 5 [least pointy]) are reported (with standard error in parenthesis).

	Experiment 3	Experiment 4
Phonological pointiness	*M* (*SE*)	*M* (*SE*)
Extreme-rounded	3.81(.10)	3.71(.07)
Rounded	3.42(.05)	3.33(.04)
Neutral	2.98(.04)	2.93(.03)
Pointy	2.65(.07)	2.82(.05)
Extreme-pointy	2.52(.14)	2.50(.10)

As was also demonstrated in Experiment 3, analysis revealed that *visuostructural pointiness ratings* significantly affected *object pointiness*, $\chi^2(2) = 208.79$, $p < .001$, with each step increase in *visuostructural pointiness ratings* raising ratings of *object pointiness* by .26 ($SE = .07$). That is, as ratings of a pseudoword's visuostructural pointiness in Experiment 1 increased in pointiness, participants paired pseudowords with objects that were pointier in appearance.

Finally, we explored whether an interaction between typeface and visuostructural pointiness ratings in Experiment 1 predicted pseudoword-object pairings. Analysis revealed that the interaction between *typeface* and *visuostructural pointiness ratings* on pseudoword-object pairings in the present experiment was not significant (albeit marginal), $\chi^2(1) = 2.87$, $p = .09$.

As also illustrated by the findings in Experiment 3, a pseudoword's sound and its visual appearance predicted the visual pointiness of its corresponding object. This appears to be the case for the graphemic-structural appearance of pseudowords and for their typeface, with pseudowords presented in a visually pointier typeface being paired with visually pointier objects. Despite approaching significance, a lack of an interaction between a typeface's visual pointiness and the visuostructural pointiness of pseudowords, suggest that these features of visual symbolism predict symbolic pointiness independently from one another. In other words, pseudowords of varying graphemic-structural pointiness can be judged as being *even more* pointy or rounded depending on the typeface in which they are presented (or vice-versa).

2.7 General Discussion

The primary aim of this chapter was to identify a list of sound-symbolically neutral pseudowords for use in all other experiments reported in this thesis. The full list of pseudowords can be found in Table 2.1 in Appendix A, complete with their individual mean ratings for visuostructural pointiness from Experiment 1 and Experiment 2 and their mean ratings for phonological pointiness from Experiment 3 and Experiment 4. It is these individual ratings that then guided decisions regarding which pseudowords in our stimulus set were *neutrally* sound-symbolic with regard to pointiness (for details see section 3.2.1.2 of Chapter 3). Whilst the aim of this chapter was relatively simple, a review of the literature highlighted the importance of acknowledging a word's visual features (e.g. graphemic structure and typeface) when assessing its overall symbolism. Until now, it was unclear how sound-symbolism and visual-symbolism combined to predict a word's symbolic potential and so, rather than assume that these features combine additively (as opposed to interactively), it was deemed necessary to address this question as part of our overall objectives. The main findings are summarised below, with suggestions as to how this work could be usefully extended.

In Experiment 1 and Experiment 2, participants rated the visuostructural appearance of our pseudowords on a Likert-type scale ranging from *very pointy* to *very rounded* and we found that a pseudoword's phonological pointiness linearly predicted ratings of its visual appearance, with pseudowords containing a larger number of pointy-sounding phonemes, compared with rounded-sounding phonemes, rated as also having a pointier visual appearance. This was the case whether the verbal recoding of pseudowords was blocked (Experiment 1) or not blocked (Experiment 2), suggesting that the visual appearance of letters in our stimulus set naturally correlated with their sound (e.g. *t* was judged as being pointy in terms of its visual appearance and its sound). We also found that the relationship between sound-symbolism and visual-symbolism was additive rather than interactive in nature.

In Experiment 3 and Experiment 4, participants paired each pseudoword with one of five objects representing a progression of visual pointiness. The results demonstrate that pointier sounding pseudowords were more likely to be associated with visually pointier objects, irrespective of whether the pseudowords were encoded auditorily only (Experiment 3) or via a combination of audition and vision (Experiment 4). Recall in Thompson and Estes' (2011) research that the ratio of small-sounding and large-sounding phonemes words contained linearly predicted the size of object that words were associated with. In a similar way, the results of Experiment 3 and Experiment 4 confirm that, like size, symbolic pointiness is also experienced as a graded function, allowing for words to be *neutrally* sound-symbolic (i.e. they are neither associated with pointy nor rounded objects but with objects that sit somewhere between these two extremes) or to possess varying levels of symbolic-pointiness (e.g. very pointy/rounded or somewhat pointy/rounded). For pseudowords presented visually in Experiment 4, typeface was manipulated to explore if, like a word's graphemic structural appearance, surface visual features also predicted a word's symbolic potential. We found that when pseudowords were presented in a pointier typeface, participants were more likely to pair them with visually pointier objects.

The idea that a word's visual appearance predicts the type of object it is associated with is not a new one. In fact, research demonstrates that some symbolic relationships might be best explained by visual similarity across printed word and object, quite separate from any phonemic influences (Cuskley, Simner & Kirby, 2015; Thompson & Estes, 2011). That is, the pointy appearance of letters within a word mirrors the pointy appearance of the object that it names. However, with pre-reading infants as young as four months old also demonstrating a sensitivity to sound-symbolism (Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011), a word's appearance cannot be the only predictor of the object (or type of object) it is associated with, or at least not for all cases. One possibility is that a word's visual appearance enhances a sensitivity to sound-symbolism that exists prior to reading, in the sense that once language can be read, there is an increased attraction to sound-symbolic exemplars that correspond both visually and phonetically with one another. As a result, some sound-symbolic relationships appreciated early on in development might be lost over the course of language development, with literate beings retaining a sensitivity only to those that correspond both visually and phonetically. Future research should explore whether pre-reading infants and children appreciate sound-symbolic relationships differently to literate children and adults, and whether these differences can be explained by a lack of experience with the visual appearance (i.e. letters) of symbolic phonemes.

A second possibility is that the relationship between sound-symbolism and visual-symbolism reported in the present chapter actually reflects inadequacies in stimulus design. A standard practice in sound-symbolism research is to source symbolic consonant and vowel sounds from previous, related studies. Having adopted this approach, the set of pseudowords created for the experiments reported in this chapter were graded according to their *assumed* phonological pointiness and assigned to one of five sound-symbolic categories. These categories then guided our understanding of the relationship between phonemic pointiness and visual pointiness. But what if the phonemes selected were in fact poor examples of sound-symbolic pointiness? For instance, what if phonemes were visually symbolic but not phonetically symbolic? Given that a large proportion of the research our assumptions are based upon was carried out with literate, adult participants who have the capacity to activate visual representations of words upon hearing them, it possible that earlier studies have wrongly interpreted visual-symbolism as sound-symbolism (especially if not properly controlled for). Whilst the findings from Experiment 2 suggest that this was not the case for our research, with a phoneme's sound only impacting ratings of a pseudoword's visuostructural appearance when participants were free to verbally recode the pseudowords, this concern should not go unaddressed by future researchers exploring sound-symbolism.

CHAPTER 3

Cross-sensory prosodic correlates to word meaning in infant-directed speech

3.1 Introduction

In section 1.5 of Chapter 1, a review of the evidence demonstrating how speakers manipulate their prosody (i.e. melody of speech) to communicate semantic information was presented (see Herold, Nygaard, Chicos & Namy, 2011; Herold, Nygaard & Namy, 2011; Hupp & Jungers, 2013; Nygaard & Lunders, 2002; Nygaard, Herold & Namy, 2009; Tzeng, Duan, Namy & Nygaard, 2018; Shintel, Anderson & Fenn, 2014). It appears that for both emotional stimuli (e.g. compared with *sadness*, *joy* is expressed by speech that is higher in pitch, louder and shorter in overall duration, Sbattella, Colombo, Rinaldi, Tedesco, Matteucci & Trivilini, 2014) and non-emotional stimuli (e.g. speakers employ a higher spoken pitch for objects moving up in space and match their rate of speech with the speed of a moving object, Shintel, Nusbaum & Okrent, 2006), speakers spontaneously recruit prosody in meaningful ways. Some prosodic cues to meaning appear to reflect known cross-sensory correspondences (key findings are summarised in Figure 3.1), suggesting that speakers are capitalising on a preestablished cross-sensory sensitivity to communicate paralinguistically (i.e. via the non-lexical elements of speech).

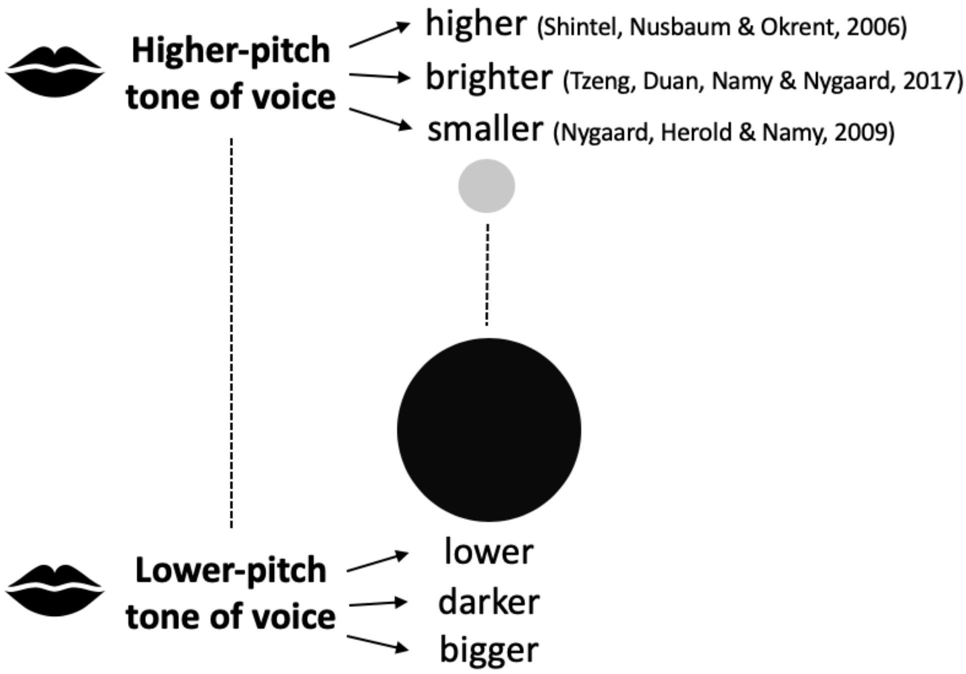


Figure 3.1. Summary of the research showing speakers employing a higher-pitch tone of voice for objects that are higher in visual space, brighter and smaller, reflecting known cross-sensory correspondences between auditory pitch and visuospatial height, brightness and size.

In the company of infants, infant-directed speakers also appear to employ prosodic cues to meaning (including size, happiness, temperature and strength, Nygaard, Herold & Namy, 2009). But other than for objects of varying sizes, it remains unknown if infant-directed speakers will manipulate their prosody in a way that capitalises on other known cross-sensory correspondences. For instance, like smaller objects, do infant-directed speakers employ relatively higher spoken pitch to refer to objects that are visually higher in space, brighter, pointier and thinner? Contributing to achieving the aim of this thesis and identifying the functional significance of cross-sensory correspondences in infant-directed speech, the experiment reported in this chapter explores how speakers with an infant audience in mind communicate the cross-sensory meaning of novel pseudowords at a prosodic-level. To begin, a brief review of the literature defining the relationship between infant-directed speech and language development is presented, followed by an overview of Experiment 5.

3.1.1 Infant-directed speech and language development

It has long been demonstrated that infants are more attracted to infant-directed speech (IDS) than they are to adult-directed speech (ADS). In brief, research shows that infants as young as two days old prefer to listen to IDS over ADS (Cooper & Aslin, 1990; Dunst, Gorman & Hamby, 2012) and that activity in the left and right temporal regions of the brain increases when infants listen to IDS compared with when they listen to ADS (Naoi, Minagawa-Kawai, Kobayashi, Takeuchi, Nakamura, Yamamoto & Kojima, 2012; Zangl & Mills, 2007). In terms of language acquisition, research shows that 21-month-olds learn novel words presented in IDS but, under similar conditions, fail to learn novel words presented in ADS (Ma, Michnick Golinkoff, Houston & Hirsh-Pasek, 2011). In Ma et al.'s study, infant participants were taught novel word-object pairs via IDS or ADS. Later, these same words were auditorily presented alongside the visual display of two objects (one of which matched the word's correct meaning) and they found that infants would only look longer towards the correct object if they had learned its label via IDS.

But why are infants successful at learning words presented in IDS but appear to be unsuccessful at learning words presented in ADS? One possibility is that infants are attracted to the general acoustic properties of IDS which, compared with ADS, is typically higher in spoken pitch, slower in duration and includes greater pitch variation (Lee, Kitamura, Burnham & Todd, 2014). Perhaps it is this attraction that causes infants to attend to IDS's linguistic content, allowing for new words to be acquired. This theory is supported with research by Estes and Hurley (2013), who habituated 17-month-olds to novel word-object pairs under the manipulation of one of three conditions: 1. novel words were presented in ADS; 2. multiple instances of a novel word were presented in IDS; or 3. single instances of a novel word were presented in IDS. They found that infants could only later detect violations of learned word-object pairs if they had previously been exposed to multiple instances of a single word in IDS (Condition 2). Considering that many characteristics of IDS are relative in nature (e.g. pitch variation), Estes and Hurley conclude that variation across spoken words was necessary to ensure that the use of IDS was detected by infants, which in turn secured their attention to its content (i.e. words).

Beyond securing attention, the feature-specific prosodic cues to meaning that exist in IDS (Nygaard, Herold & Namy, 2009) indicate that this register of speech[11] might be being used in more sophisticated ways, perhaps with the capacity to promote language comprehension (i.e. interpret word meaning). As demonstrated in Figure 3.1, some prosodic cues to meaning actually reflect known cross-sensory correspondences that infants are found to be sensitive to, such as the correspondences between auditory pitch and size (Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004), brightness (Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004) and visuospatial height (Dolscheid, Hunnius, Casasanto & Majid, 2014; Walker et al., 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018). This observation indicates that there might be other correspondences

[11] A register of speech is a style of spoken language used in a particular communicative context. For instance, speech directed to infants vs speech directed to adults.

(e.g. pitch-pointiness and pitch-thinness) that are also being capitalised on by speakers but, until now, only prosodic correlates to size have been established in IDS[12] (Nygaard, Herold & Namy, 2009). Therefore, the aim of this chapter is to extend the work of Nygaard et al. and identify whether the range of cross-sensory correspondences that infants are found to be sensitive to are revealed at the prosodic-level of infant-directed speech.

3.1.2 Overview of Experiment 5

The aim of Experiment 5 was to identify prosodic cues to word meaning that are related to a range of cross-sensory features (i.e. visuospatial height, size, brightness, pointiness and thinness) in infant-directed speech. To do this, adult speakers with regular experience communicating with infants were presented with the visual display of two novel objects representing opposite poles on dimensions relating to five antonym pairs (i.e. high/low, small/big, bright/dark, pointy/rounded and thin/thick). Imagining that they were talking to an infant, participants were asked to read aloud ambiguous sentences containing a novel pseudoword (e.g. "Look at the *rebo* one.") each time associating the pseudoword with just one of the objects within a pair. For each sentence, measures of fundamental frequency (i.e. pitch), duration (i.e. rate of speech) and amplitude (i.e. loudness) were obtained and analysed in order to identify the acoustic signatures of each cross-sensory feature.

Whilst we anticipated that speakers would manipulate their pitch of voice, rate of speech and loudness to refer to various cross-sensory features, it is important to

---

[12] Prosodic correlates to visuospatial height and brightness outlined in Figure 3.1 were established for ADS only.

note that these dimensions are not all scaled in the same way. For instance, compared with speed and amplitude, which are magnitude-based dimensions, pitch is generally considered to be a metathetic dimension[13] (Spence, 2011, Stevens, 1975). That is, higher-pitch sounds are not 'more' or 'less' than lower-pitch sounds but are qualitatively different[14]. For this reason, whilst the relationships between auditory pitch and the range of cross-sensory features included in this experiment is well-established (i.e. higher-pitch sounds are judged as being higher in space, smaller, brighter, pointier and thinner), it was unclear whether rate of speech and amplitude would align in consistent or contradictory directions for all or some cross-sensory features.

To ensure that prosodic cues to meaning recorded in the present experiment were uncontaminated by the sound-symbolic potential of the words themselves, the novel pseudowords included in this experiment were taken from the bank of pseudowords created in Chapter 2 and selected as being examples of pseudowords which were judged to be *neutrally* sound-symbolic (i.e. they were neither pointy or rounded in visuostructural appearance, nor were they associated with visually pointy or rounded objects). Without excluding sound-symbolism from our investigation in

---

[13] The distinction between *metathetic* and *prothetic* (i.e. magnitude-based) dimensions is considered to be a distinction between quality and quantity. That is, opposite dimensions that are metathetic (such as high- vs low-pitch) are experienced as being qualitatively different from one another, whereas opposite dimensions that are prothetic (such as quiet vs loud or fast vs slow) are experienced as being more or less in amount than one another.

[14] Whilst pitch is widely accepted as being a metathetic dimension, a physical analysis of pitch would conclude that higher-pitch is 'more' in terms of vibrations per unit of time than lower-pitch, and so this definition is doubtful.

this way, a speaker's prosody could be influenced by a pseudoword's inherent sound-symbolic potential.

3.2 Experiment 5: Acoustic profiles of visuospatial height, size, brightness, pointiness and thinness in infant-directed speech

3.2.1 Method

3.2.1.1 Participants

Twenty-four adults (20 females and 4 males) aged between 20 and 49 years (M = 29.25, SD = 7.75) were recruited from Lancaster University and volunteered to take part in the study in exchange for payment of £5. All participants spoke English as a first language, with self-reported normal or corrected-to-normal hearing and vision. All participants had regular and recent experience with one or more infants under the age of 24 months; thirteen were parents, five worked with infants, and six had a close family member or friend with a child. An additional six participants were tested but were excluded from the analysis due to experimental error (n = 2) or a failure to complete all trials during testing (n = 4). All experimental procedures were approved by the FST Research Ethics Committee, Lancaster University, on 14th August 2017 (FST16187).

3.2.1.2 Materials

The pseudowords included in this and all following experiments presented in this thesis were taken from the list of pseudowords generated and then rated according to their visual- and sound-symbolic pointiness in Chapter 2 (see Table 2.1 in

Appendix A for a full list of pseudowords). For the present experiment, five pseudowords were selected from the list as those judged as being *neutrally* symbolic in terms of pointiness (i.e. they were neither pointy or rounded in visuostructural appearance, nor were they associated with visually pointy or rounded objects). The selection criteria were as follows: for ratings of visual-symbolic pointiness (Experiment 1 and Experiment 2) each pseudoword had a mean rating of 4 (± 2) and for ratings of sound-symbolic pointiness (Experiment 3 and Experiment 4) each pseudoword had a mean rating of 3 (± .5). The five pseudowords were: *temu* (pronounced /temu:/ and rhymes with a combination of *stem* and *moo*), *bori* (/bɒri:/, rhymes with *lorry*), *rebo* (/reboʊ/, rhymes with *fellow*), *ribo* (/rɪboʊ/, rhymes with *widow*), and *timu* (/tɪmu:/, rhymes with *igloo*).

Sentences in the following format: 'Look at the (novel pseudoword) one.' were presented visually, in the centre of an Apple MacBook (2008) 13-inch computer screen. Sentences were accompanied by the visual display of two novel, 3-dimensional objects positioned directly above the sentence. One object within a pair was highlighted by the indication of an arrow (see Figure 3.2). Object pairs were designed using Autodesk Inventor Professional (2015) and represented opposite feature values within one of five antonym pairs: *small/big*, *high/low*, *pointy/rounded*, *bright/dark* and *thin/thick* (see Table 3.1). Objects within a pair were matched closely for all other features (e.g. profile, perceived size and brightness) outside of the dimension by which they contrasted (e.g. visuospatial height). The objects designed to represent opposite values of pointiness were identical to those used in Experiments 1 – 4 (for details see section 2.4 of Chapter 2). To reduce the perception that objects within a pair were positioned at varying distances from the viewer, the objects were presented as if sitting on (or, for the object representing *high*, floating above) a plain, untextured, horizontal surface.

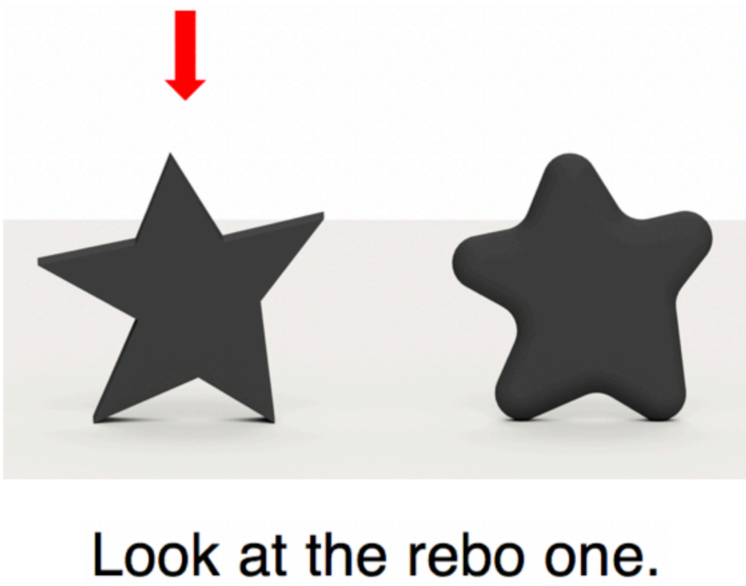


Figure 3.2. An example of the display as shown to participants. Participants were asked to assign the meaning of the novel pseudoword presented (i.e. rebo) to whichever object (e.g. pointy) was indicated to by the arrow on a given trial and to repeat the sentence aloud, in an infant-directed style of speech.

A questionnaire including six cross-sensory related questions, was issued at the end of the experiment and included to assess the degree to which participants were sensitive to key cross-sensory correspondences (whilst the full questionnaire can be found in Appendix B, see Figure 3.3 for an example). For each question, participants were first required to read a short statement about a hypothetical object and then, based on this information, answer a question about an unrelated feature the object also possessed. For example, ‘This object makes a high-pitched sound. What shape do you think it is?’. Responses were made on 7-point Likert scales, defined by one of five antonym word pairs: *pointy-rounded*, *thin-thick*, *bright-dark*, *small-big* and *high-*

pitched-low-pitched. The seven points on each scale were either unlabelled or labelled with VERY, SLIGHTLY or NEITHER so that, for instance, the sequence was VERY ROUNDED, ROUNDED, SLIGHTLY ROUNDED, NEITHER ROUNDED NOR POINTY, SLIGHTLY POINTY, POINTY, and VERY POINTY. The left-right direction of the antonym word pairs (i.e. pointy-rounded or rounded-pointy) was randomly generated so that they were not aligned consistently with the correspondences themselves (e.g. pointy, thin, bright, small and high-pitched were not always on the left or always on the right).

Participant's utterances were audio recorded using a Roland R-09HR 24 bit 96 kHz WAVE/MP3 Recorder.

1. This object makes a high-pitched sound. What shape do you think it is?

VERY ROUNDED	ROUNDED	SLIGHTLY ROUNDED	NEITHER ROUNDED NOR POINTY	SLIGHTLY POINTY	POINTY	VERY POINTY

Figure 3.3. An example of a questionnaire item assessing the extent to which participants were sensitive to some known cross-sensory correspondences.

3.2.1.3 Procedure

Individual sentences containing one of five novel pseudowords (e.g. 'Look at the rebo one.') were presented to participants. Participants were required to read aloud each sentence and instructed to employ infant-directed speech appropriate for an infant under the age of 24 months. Next, the same sentences were presented individually and below the display of two novel objects, one of which was indicated by an arrow on the screen. Participants were informed that "the arrow is pointing towards the object that represents the unfamiliar word in the sentence". Again, imagining that they were talking to an infant, participants were required to repeat the sentence aloud. Importantly, participants were asked to imagine that the infant in mind was looking at the same display of two objects but without any visual cue identifying the meaning of the novel word. Participants were instructed to help the infant work out which of the two objects they were referring to but were given no indication of how this might be achieved.

For each antonym pair (e.g. small/big), participants read aloud two sentences containing the same pseudoword, one for each opposite value. The pseudoword that was used to represent an antonym pair was randomly selected per participant, as was the order in which sentences were presented. In total, participants read aloud 10 sentences, which were audio recorded by the experimenter. In a fully counterbalanced within-subjects design, whether an object representing one value within an antonym pair was positioned on the left or right of the screen was counterbalanced, along with the left and right positioning of the arrow. In order to ensure participants correctly identified the feature by which objects within a pair contrasted, at the end of the experiment participants completed a short questionnaire asking them to indicate the feature (e.g. size) that differed between each object within an object pair. The questionnaire also included a series of six cross-sensory related questions designed to assess the extent to which participants were sensitive to cross-sensory correspondences. The study was conducted in a silent room and the whole procedure took approximately 30 minutes.

3.2.2 Results

3.2.2.1 Questionnaire Data

To begin, we present the results of the questionnaire data because some of these findings are included in the subsequent analyses reported in this section. Participants' responses as to which feature they believed to have varied across objects within a pair are displayed in Table 3.1. Of the 24 participants, 23 (96%) correctly identified size, 20 (83%) correctly identified visuospatial height, 24 (100%) correctly identified pointiness and 23[15] (96%) correctly identified brightness as the feature differentiating objects within these pairs. For objects that differed by thinness, 11 participants (46%) identified size as the distinguishing feature, 8 (33%) identified thinness and 5 (21%) identified shape. Whilst the difference in proportions between size and thinness was not significant, $\chi^2(1) = .35$, $p = .56$, analysis revealed that the majority of participants reported either size or shape rather than thinness as being the distinguishing feature across these objects, $\chi^2(1) = 4.08$, $p = .04$. These findings and their potential to impact the results of the present experiment are discussed further in section 3.3.1 of Chapter 3.

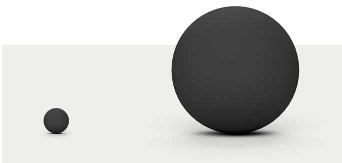
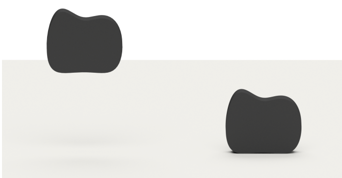
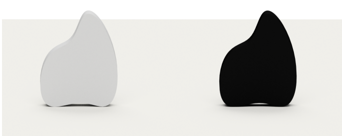
A unique score for sensitivity to cross-sensory correspondences was also recorded for each participant and included as a fixed effect (named 'cross-sensory sensitivity score') in the linear mixed effects models reported in section 3.2.2 of Chapter 3. For purpose of statistical analysis, sensitivity scores were achieved by assigning values (ranging from 1 'sensitive to correspondences in the unexpected direction' to 7 'sensitive to correspondences in the expected direction') to individual questionnaire responses. The mean of these responses was then calculated for each participant. Should participants display an insensitivity to correspondences in either

[15] For brightness, 13 participants referenced to colour, but this was accepted as correct identification.

direction, a null value of 4 would be expected. Overall, participants demonstrated a sensitivity to cross-sensory correspondences in the expected direction (*M* = 5.32, *SD* = .77), which was significant compared to chance, $t(23) = 6.98$, $p < .001$.

Table 3.1

Antonym pairs and their corresponding objects. In the final two columns, the features reported by participants as varying across objects within a pair are displayed, along with the number of participants (max = 24) reporting the same feature.

Dimensional Adjective (intended)	Object pair	Dimensional Adjective (reported)	*N*
Size (small vs big)		***size***	23
		weight	1
Visuospatial height (high vs low)		***height***	20
		weight	1
		shadow	1
		distance (from viewer)	1
		softness	1
Pointiness (pointy vs rounded)		***pointiness***	24
Brightness (bright vs dark)		***brightness***	23
		weight	1
Thinness (thin vs thick)		***thinness***	8
		size	11
		shape	5

3.2.2.2 *Acoustic Analyses*

In the present experiment, it was possible that participants would manipulate their prosody across an entire sentence or just in relation to its target word (i.e. novel pseudoword). To check, utterances were analysed at both their sentence-level and target word-level (reported in that order). Whilst the focus of this thesis is concerned with cross-sensory correspondences related primarily to auditory pitch, other research has found that amplitude can also correspond with some object properties. For instance, objects that are positioned higher in space are judged to be louder than objects positioned lower in space (Puigcerver, Rodríguez-Cuadrado, Gómez-Tapia & Navarra, 2019). The potential for speaking rate to also reflect different dimensions of meaning is likely, especially given the relationship between speed, size and weight (larger objects are generally heavier and move more slowly than smaller objects). For this reason, using PRAAT (Boersma & Weenink, 2017), mean fundamental frequency, duration and mean amplitude values were extracted for each sentence-level and target word-level utterance.

Fundamental frequency (Hz) is a measure of how many times the vocal folds vibrate per second. The sensation of a frequency is commonly referred to as the pitch of a sound, with higher-pitch sounds containing a greater number of vibrations per second (e.g. 2500Hz) compared with lower-pitched sounds (e.g. 300Hz). Variations in the duration (or length) of utterances serves as an index of speaking rate, with shorter durations (measured in seconds) indicating a faster rate of speech compared with longer durations. Amplitude (measured in decibels [dB]) refers to the overall energy of an utterance and indicates how loudly a sentence is spoken. Prior to extracting values for amplitude, amplitude was normalised to account for any variation across participants and within experimental sessions that could not be controlled for (e.g. distance from microphone, age or physiology), whilst preserving those variations that were made in response to the stimuli. Normalising in this way does not affect values for fundamental frequency or duration.

3.2.2.3 Overview of analysis

For the present experiment, the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) in R (R Core Team, 2013) was used to perform linear mixed effects analyses exploring the relationship between a given dependent variable and controlled (fixed) and uncontrolled (random) independent variables. The general strategy for creating the models reported in this chapter was to include all fixed and random effects as a first instance and then to eliminate those that reduced the model's overall goodness of fit. This was achieved by comparing each model's Akaike Information Criterion (AIC) using the *AICcmodavg* package (Mazerolle, 2019) in R, which estimates the quality of a statistical model for a given set of data (with a lower AIC value indicating a higher quality model). Once a model reached its optimum (i.e. lowest) AIC, we identified which independent variable/s were having a significant effect on the dependent variable. To do this, *p*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. For all statistical tests, an alpha level of .05 was used as a significance criterion. Unless reported, visual inspections of residual plots did not reveal any obvious deviations from homoscedasticity or normality in any of the linear mixed effects analyses reported in this chapter.

Prior to performing linear mixed effects analyses, the data were subsetted by antonym pair (e.g. *small/big*, *high/low*, *pointy/rounded*, *bright/dark* and *thin/thick*) so that we could identify which antonym pairs participants manipulated their prosody for and in which direction (e.g. higher-pitch voice for smaller as opposed to bigger objects). The intended meaning of a pseudoword on a given trial (e.g. small or big) was also recorded for each utterance and, in the models reported in this chapter, is represent by the fixed effect, *pseudoword meaning*.

3.2.2.4 Sentence-level Analysis: Fundamental Frequency

The dependent measure was the mean fundamental frequency (measured in Hz) of each sentence-level utterance. The optimum model included a fixed effect for *pseudoword meaning* and *cross-sensory sensitivity score*, including an intercept for participants as a random effect. Analysis revealed that only for the antonym pair *high/low* did *pseudoword meaning* significantly affect *fundamental frequency*, $\chi^2(1) = 7.98$, $p = .005$, with each step increase in visuospatial height raising *fundamental frequency* by 18.08 Hz ($SE = 6$) (for means and standard error of the means see Figure 3.4). That is, participants employed a relatively higher spoken pitch when referring to objects positioned higher as opposed to lower in visual space. A marginally significant effect of *pseudoword meaning* on *fundamental frequency* was also recorded for the antonym pair *small/big*, $\chi^2(1) = 2.84$, $p = .09$, with each step increase in size raising *fundamental frequency* by 18.51 Hz ($SE = 10.89$).

3.2.2.5 *Sentence-level Analysis: Duration*

The dependent measure was the duration (measured in seconds) of each sentence-level utterance. The optimum model included a fixed effect[16] for *pseudoword meaning* and an intercept for participants as a random effect. Analysis revealed that only for the antonym pair *small/big* did *pseudoword meaning* significantly affect *duration*, $\chi^2(1) = 6.05$, $p = .01$, with each step increase in size shortening *duration* by .14 seconds ($SE = .05$) (for means and standard error of the means see Figure 3.4). That is, participants employed a relatively faster rate of speech (i.e. shorter duration) when referring to objects that were smaller as opposed to bigger in size.

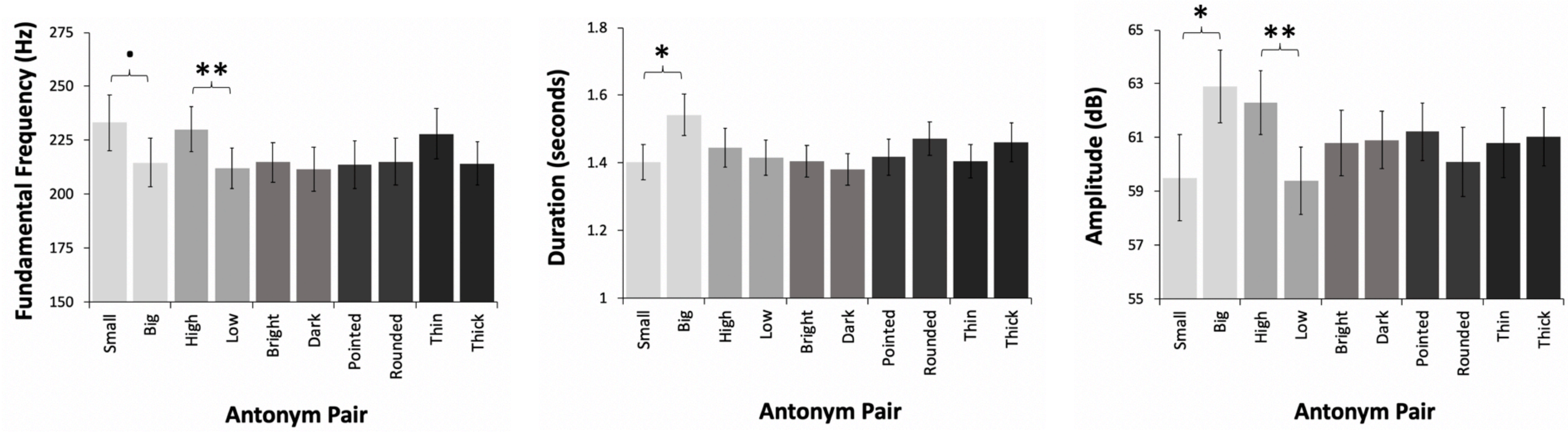
3.2.2.6 *Sentence-level Analysis: Amplitude*

The dependent measure was the mean amplitude (measured in dB) of each sentence-level utterance. The optimum model included a fixed effect for *pseudoword meaning* and *cross-sensory sensitivity score* and an intercept for participants as a random effect. Analysis revealed that for the antonym pair *high/low, pseudoword meaning* significantly affected *amplitude*, $\chi^2(1) = 7.77$, $p = .005$, with each step increase in visuospatial height raising *amplitude* by 2.46 dB ($SE = .83$). A significant effect of *pseudoword meaning* on *amplitude* was also recorded for the antonym pair *small/big*, $\chi^2(1) = 5.49$, $p = .02$, with each step increase in size lowering *amplitude* by 2.13 Hz ($SE = .87$). In summary, participants employed a relatively louder voice (i.e.

---

[16] For this model, including *cross-sensory sensitivity score* as a fixed effect increased the model's AIC value, which was interpreted as lowering the model's overall goodness of fit. Therefore, cross-sensory sensitivity scores were excluded from this model.

amplitude) when referring to objects positioned higher as opposed to lower in visual space and for objects that were bigger as opposed to smaller in size (for means and standard error of the means see Figure 3.4).



. $p < .1$, * $p < .05$, ** $p < .01$, otherwise non-significant

Figure 3.4. Sentence-level mean values for fundamental frequency measured in hertz(Hz), duration measured in seconds and amplitude measured in decibels(dB) according to opposite values within antonym pairs (*small/big*, *high/low*, *bright/dark, pointy/rounded* and *thin/thick*). Bars represent standard errors of the means.

3.2.2.7 Target Word-level Analysis: Fundamental Frequency

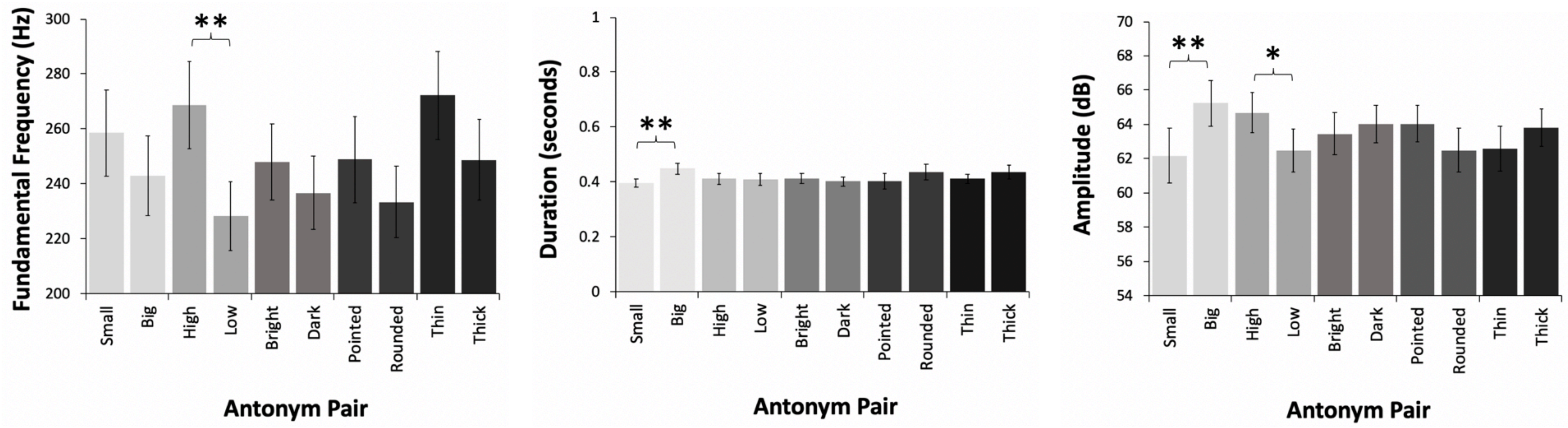
The dependent measure was the mean fundamental frequency (measured in Hz) of each target word-level utterance. The optimum model included a fixed effect for *pseudoword meaning* and *cross-sensory sensitivity score*, including an intercept for participants as a random effect. Analysis revealed that only for the antonym pair *high/low* did *pseudoword meaning* significantly affect *fundamental frequency*, $\chi^2(1) = 9.23$, $p = .002$, with each step increase in visuospatial height raising *fundamental frequency* by 40.37 Hz ($SE = 12.29$) (for means and standard error of the means see Figure 3.5). That is, participants employed a relatively higher spoken pitch when referring to objects positioned higher as opposed to lower in visual space.

3.2.2.8 Target Word-level Analysis: Duration

The dependent measure was the duration (measured in seconds) of each target word-level utterance. The optimum model included a fixed effect for *pseudoword meaning* and *cross-sensory sensitivity score*, including an intercept for participants as a random effect. Analysis revealed that only for the antonym pair *small/big* did *pseudoword meaning* significantly affect *duration*, $\chi^2(1) = 7.66$, $p = .006$, with each step increase in size shortening *duration* by .05 seconds ($SE = .02$) (for means and standard error of the means see Figure 3.5). That is, participants employed a relatively faster rate of speech (i.e. shorter duration) when referring to objects that were smaller as opposed to bigger in size.

3.2.2.9 Target Word-level Analysis: Amplitude

The dependent measure was the mean amplitude (measured in dB) of each target word-level utterance. The optimum model included a fixed effect for *pseudoword meaning* and *cross-sensory sensitivity score* and an intercept for participants as a random effect. Analysis revealed that for the antonym pair *high/low, pseudoword meaning* significantly affected *amplitude*, $\chi^2(1) = 6.14$, $p = .01$, with each step increase in visuospatial height raising *amplitude* by 2.23 dB ($SE = .86$). A significant effect of *pseudoword meaning* on *amplitude* was also recorded for the antonym pair *small/big*, $\chi^2(1) = 7.32$, $p = .007$, with each step increase in size lowering *amplitude* by 3.07 Hz ($SE = 1.07$). In summary, participants employed a relatively louder voice (i.e. amplitude) when referring to objects positioned higher as opposed to lower in visual space and for objects that were bigger as opposed to smaller in size (for means and standard error of the means see Figure 3.5).



* $p < .05$, ** $p < .01$, otherwise non-significant

Figure 3.5. Target word-level mean values for fundamental frequency measured in hertz(Hz), duration measured in seconds and amplitude measured in decibels(dB) according to opposite values within antonym pairs (*small/big*, *high/low*, *bright/dark, pointy/rounded* and *thin/thick*). Bars represent standard errors of the means

3.2.3 Summary of main findings

In brief, the results of Experiment 5 demonstrate that infant-directed speakers manipulated their spoken pitch in accordance with known cross-sensory correspondences between auditory pitch and visuospatial height, at both a sentence-level (i.e. across the entire length of a sentence) and at a target word-level (i.e. across the length of a pseudoword). In other words, speakers employed a relatively higher-pitch tone of voice when referring to objects positioned higher rather than lower in visual space. Despite being marginally significant (and for sentence-level analyses only), speakers also employed a higher-pitch tone of voice for smaller (as opposed to bigger) objects. Again, this finding is in line with the cross-sensory correspondence between auditory pitch and size (higher-pitch sounds are *smaller*). A number of feature-specific prosodic correlates to meaning were also observed, for example, an object's visuospatial height was represented by distinctions in amplitude (i.e. *higher* was louder) and size represented by distinctions in duration and amplitude (i.e. *smaller* was faster and quieter). Some of the findings summarised here are in line with earlier research by Nygaard, Herold and Namy (2009), who recorded similar relationships between size and a speaker's fundamental frequency (also only marginally significant), speech rate and amplitude.

3.3 General Discussion

The aim of this chapter was to identify prosodic cues to word meaning that are related to a range of cross-sensory features (i.e. visuospatial height, size, brightness, pointiness and thinness) in infant-directed speech. The main findings are discussed in more detail below, including the observation that infant-directed speakers do not employ prosodic cues to meaning for objects that vary in terms of their brightness, pointiness or thinness. In short, we found that only for objects of contrasting visuospatial heights and sizes did infant-directed speakers manipulate their prosody in accordance with known cross-sensory correspondences (i.e. higher-pitch tone of voice for objects that were higher in visual space and smaller in size). The distinction between metathetic dimensions (i.e. scaled in terms of their quality e.g. auditory pitch) and prothetic dimensions (i.e. scaled in terms of their quantity e.g. speed and amplitude) is also revisited in light of our findings.

3.3.1 What about brightness, pointiness and thinness?

One surprising finding to emerge from Experiment 5 was the lack of prosodic cues employed by speakers to represent brightness, pointiness and thinness, especially given the wealth of research demonstrating cross-sensory correspondences between auditory pitch and each of these object features (for pitch-brightness see Haryu & Kajikawa, 2012; Marks, 1974; Marks, 1989; Mondloch & Maurer, 2004; Zeljko, Kritikos, & Grove, 2019, for pitch-pointiness see Walker et al., 2010; Walker, 2012, for pitch-thinness see Dolscheid, Hunnius, Casasanto & Majid, 2014; and for a review see Spence, 2011). In this section, these findings are discussed in the context of other research related to cross-sensory correspondences and/or paralinguistic communication. As the first study to explore prosodic correlates to brightness, pointiness and thinness in infant-directed speech, suggestions for extending this work are also offered.

In Tzeng, Duan, Namy and Nygaard's (2018) study, adult-directed spoken words for brighter shades of colour were relatively higher-pitch, faster and louder than darker shades. However, in the present experiment, participants did not distinguish between contrasting brightnesses based on spoken pitch, rate of speech or amplitude. Whilst there is an expectation that infant-directed speakers will exaggerate their prosody beyond what is already found in adult-directed communication, one possibility is that these registers of spoken language vary in more ways (rather than just to their degree) than is currently understood. In other words, perhaps only in the company of adults (rather than infants) will speakers employ prosodic cues to brightness. A more likely possibility is that the stimuli employed in Tzeng, Duan, Namy and Nygaard's research provoked a different response from participants than those employed in our experiment. In their study, *brightness* was achieved by adjusting the amount of white a primary colour (i.e. red) contained in gradual increments. Whilst this method does alter the perceived brightness of a colour, it also has the potential to change a colour entirely. For instance, adding or removing white from red results in colour changes that more closely resemble pink or maroon, respectively. This is an important consideration given the varying connotations that different colours (despite originating from the same shade) possess. For instance, pink signifies softness, sweetness and is considered to be gender-specific, whereas maroon is considered to be warm, deep and passionate. The decision to represent brightness as grayscale in the present experiment limits this confound and reduces its potential to impact perception (not forgetting that participants in Experiment 5 correctly identified brightness as the primary feature distinguishing these objects). These observations lead to the tentative conclusion that (at least for infant-directed speech) brightness in its explicit form is not represented at a prosodic level.

As discussed in section 1.5 of Chapter 1, research has shown that when a speaker's prosody is congruent with a word's meaning, the retention of newly acquired words is improved (Shintel, Anderson & Fenn, 2014). That is, for listeners, congruent prosody can be instrumental in successful word learning. For objects that differed in terms of their pointiness (or sharpness), Shintel, Anderson and Fenn defined 'congruent prosody' as being higher-pitch, louder and faster for pointier(sharper) objects. Like *brightness*, the findings from the present experiment showed that infant-directed speakers were not spontaneously employing these same prosodic cues when talking about objects of varying pointiness. It is possible that these differences illustrate that listeners are more receptive to meaningful prosody than speakers are at producing it, introducing a potential avenue for future research. Rather than assume that infants are only receptive to the same prosodic cues that infant-directed speakers produce, the aim of Chapter 4 is to identify whether infants attend to prosody that reflects the full range of cross-sensory correspondences infants are sensitive to, including those related to brightness, pointiness and thinness.

In the present experiment, questionnaire responses revealed that only 33% of participants correctly identified thinness as the feature differentiating the thin-thick objects. The decision to represent thinness as wider or narrower rods (or tubes), occurred after their success in revealing pitch-thinness correspondences in infants (Dolscheid, Hunnius, Casasanto & Majid, 2014). However, thinness is unavoidably conflated with size (thicker objects are also bigger) and, by association, weight (bigger objects are usually heavier). It also appears that representing thinness as rods can impact perceptions of shape, with 21% of participants in the present experiment identifying shape as the distinguishing feature across thin-thick objects. A challenge for future researchers is to identify other ways of representing thinness whilst limiting these confounding variables (size, weight and shape as just three examples). One approach to this is employed in the series of experiments presented in Chapter 4, where thinness represented by rods is replaced by skeletal cubes whose frames increase or decrease in thinness but overall sizes (i.e. footprints) remain the same.

3.3.2 The convergence of pitch and amplitude

The convergence of cross-sensory features related to auditory pitch (i.e. pointy, thin, bright, small and light vs rounded, thick, dark, big and heavy) is well-established, such that, smaller objects are judged to be brighter (Walker, Walker & Francis, 2015), just as we might expect thinner objects to be brighter or pointier objects to be lighter in weight. However, for amplitude, which is magnitude-based (i.e. prothetic), it appears that these features do not converge in the same way. Instead, we found that participants spoke relatively louder for objects positioned higher in space and quieter for objects that were smaller in size. These findings provide evidence that magnitude-based associations can be distinct from cross-sensory correspondences, and whilst they are interesting to study in their own right, they have the potential to impact research concerned with the cross-sensory nature of prosody. For instance, an association that exists between spoken pitch and visuospatial height or size may be extinguished if accompanied by incongruent amplitudes. For this reason, when exploring whether infants respond to prosody by way of resolving linguistic uncertainty in Chapter 4, variations in amplitude will be eliminated (where possible) from our investigations.

CHAPTER 4

Infants' attention to prosody in conditions of linguistic uncertainty

4.1 Introduction

The detection of cross-sensory correspondences shortly after birth, and at least before language acquisition, indicates that some correspondences are not linguistically mediated (despite their notable presence in language, for a review see section 1.3 of Chapter 1). In demonstrating that infant-directed speakers manipulate their pitch of voice to reflect changes in visuospatial height (i.e. higher-pitch tone of voice for objects positioned higher in visual space), the findings reported in Chapter 3 indicate that speakers might be capitalising on this early sensitivity in a way that promotes language development (i.e. helping infant listeners identify a word's meaning by mapping their pitch of voice to a referent's spatial location). Research with children shows that they require more obvious prosodic cues to meaning than adults (Hupp & Jungers, 2013) and only attend to prosody if instructed to or provided with training (Herold, Nygaard, Chicos & Namy, 2011). But do infants extract semantic information from prosody? Whilst research is yet to answer this question in regard to non-emotional stimuli, Friend (2001) found that infants are sensitive to changes in emotional prosody, but that this sensitivity diminishes once language learning accelerates. One possibility is that children experience a developmental decline in their responsiveness to prosody (emotional or otherwise) around the onset of formal education, in which a focus towards language learning at a linguistic rather than paralinguistic level is instilled. As such, related research findings with children should not discourage the exploration of infants' attention to prosody and the potential it has to impact language development.

The present chapter presents three experiments with the shared aim of identifying whether infants attend to non-emotional prosody to resolve linguistic uncertainty (i.e.

interpret word meaning), and whether they do so in a way that capitalises on the known cross-sensory correspondences that they are found to be sensitive to (for a review see section 1.2 of Chapter 1). The general approach adopted in the series of experiments reported here was to present participants with novel object pairs that differed by a single cross-sensory feature (e.g. size, visuospatial height, pointiness, brightness, thinness or weight), alongside sentences containing novel pseudowords spoken in a prosodically deliberate way (e.g. in a higher- or lower-pitched voice and/or in a faster or slower rate of speech). It was assumed that if participants were attending to prosody and matching pseudowords with objects based on this information, they would demonstrate a preference for (i.e. look longer towards or manually select) a particular object within an object pair. If participants were capitalising on their sensitivity to cross-sensory correspondences in this exercise, we would expect them to prefer objects that were smaller, higher in visual space, pointier, brighter, thinner or lighter in weight when prosody was higher in pitch and/or faster in speaking rate.

4.2 Experiment 6: Prosody as a cue to meaning for 13-month-olds

To address these aims, Experiment 6 first investigated whether infants at the onset of language production interpreted the cross-sensory meaning of novel pseudowords by attending to the prosody in which the pseudowords were spoken. In one study, Friend (2001) found that 15-month-old infants were more likely to engage with a toy when a speaker's prosody was approving rather than disapproving, even when the linguistic content of speech was contradictory (e.g. "Don't touch."). However, for infants that were more advanced at understanding the linguistic content of speech, this effect was reversed: highly receptive infants responded to a speaker's lexical rather than prosodic message. Friend's observation that infants are more sensitive to affective prosody (or paralanguage) earlier on in development motivated the decision to explore our aims with 13-month-olds. Whilst at 13 months of age infants have acquired some receptive vocabulary and are able to produce a handful of words, it would appear that they are more likely to rely on prosody to interpret a speaker's intentions. Using a screen-based, preferential looking procedure, participants in this experiment were shown two objects (differing by a single cross-sensory feature, e.g. size) and listened to ambiguous sentences describing one of the two objects (e.g. "Ooh look, a rebo one! Where's the rebo one?"), spoken in either a higher- or lower-pitch tone of voice and/or a faster or slower rate of speech.

4.2.1 Method

4.2.1.1 Participants

Fifteen male and 15 female infants (M = 13.13 months; range: 13.03 – 13.27 months) completed the study. A further 7 infants were excluded from the analysis because of excessive restlessness and for looking at the screen for less than 70% of the time. All experimental procedures were approved by the FST Research Ethics Committee, Lancaster University, on 6th March 2018 (FST17079).

4.2.1.2 Materials

Auditory stimuli. A female speaker (aged 24 years) was recorded producing short sentences containing either a novel pseudoword or familiar word. The speaker, blind to the research aims, was instructed to employ infant-directed speech throughout the recordings and to imagine that they were talking to an infant aged 13 months. Having worked with infants under the age of 36 months for 3 years, the speaker had extensive experience communicating with children of this age.

The pseudowords included in this experiment were the same as those used in Experiment 5 (see section 3.2.1.2 of Chapter 3 for details). Sixteen familiar words were sourced from The Oxford Communicative Development Inventory database (Hamilton, Plunkett & Schafer, 2000), which provides an index of receptive and productive vocabulary for English-speaking (British) children at different ages. The familiar words were all nouns and selected from the database as being amongst the first 100 words produced by 13- and 14-month-old infants in the United Kingdom (see Table 4.1).

The novel pseudowords were all embedded within the same phrase: "Ooh look, a (pseudoword) one! Where's the (pseudoword) one?" The speaker was instructed to read the phrase aloud for all pseudowords six times, each time adjusting her prosody for the whole utterance[17] in one of six ways: speaking with a higher-pitch tone of voice (high), a lower-pitch tone of voice (low), a faster rate of speech (fast), a slower rate of speech (slow), a higher-pitch tone of voice and a faster rate of speech (high+fast), or a lower-pitch tone of voice and a slower rate of speech (low+slow). Phrases were audio recorded in a silent room using a Roland R-09HR 24 bit 96 kHz WAVE/MP3 Recorder. Once clipped into individual utterances, each recording was modified[18] using PRAAT (Boersma & Weenink, 2017) so that, within a given prosody style (e.g. higher-pitch tone of voice), all utterances were acoustically similar in terms of their fundamental frequency, duration and amplitude. For

---

[17] In Chapter 3 we found that infant-directed prosody was similar across the length of an utterance and across the length of a target-word. For this reason, in the present experiment, the speaker was instructed to manipulate her prosody at a sentence-level (rather than at a target word-level).

[18] Representing varying prosody styles of speech could have been achieved in one of three ways: 1. Manipulating prosody via natural speech only (no electronic modification), 2. Manipulating prosody via electronic modification only (no manipulation of natural speech), or 3. Manipulating prosody via a combination of natural speech and electronic modification. Having attempted all three options, the latter proved to be the most appropriate at controlling the acoustic properties of speech whilst also allowing speech to sound natural.

instance, sentences spoken in a higher-pitch tone of voice had a sentence-level mean fundamental frequency of around 289 Hz, duration of 3 seconds and amplitude of 75dBs. Utterances were also modified so that across prosody styles (e.g. higher-pitch vs lower-pitch), acoustic features by which utterances unintentionally differed (e.g. duration and amplitude) were also similar. For instance, for higher- vs lower-pitch utterances, the sentence-level mean duration and amplitude were around 3 seconds and 74.5 dBs, respectively. Table 4.2 provides a full list of the mean values for fundamental frequency, duration and amplitude (post-modification) for each prosody style.

Familiar words were each embedded in one of four phrases (which word was presented in which phrase was randomly determined for each participant):

"Ooh a (familiar word)! Where's the (familiar word)?";

"Look at the (familiar word). Can you see the (familiar word)?";

"Wow a (familiar word)! Look at the (familiar word)!";

and, "Where is the (familiar word)? Can you find the (familiar word)?" Again, in an infant-directed style of speech, the speaker was instructed to read each phrase aloud for each familiar word once, which were audio recorded and clipped into individual utterances as described above.

Table 4.1

Sixteen familiar words and their corresponding picture pair representations.

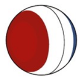
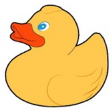
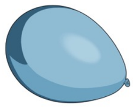
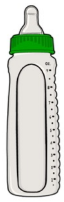
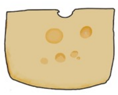
Familiar Words	Picture Pair	
apple vs banana		
ball vs car		
duck vs teddy bear		
sock vs shoe		
balloon vs book		
drink vs dummy		
bubble vs hat		
cheese vs clock		

Table 4.2

Sentence-level mean values (with standard deviation in parenthesis) for fundamental frequency measured in Hz, duration measured in seconds and amplitude measured in dB according to each prosody style: higher-pitch (high), lower-pitch (low), faster rate (fast), slower rate (slow), higher-pitch and faster rate (high+fast), or lower-pitch and slower rate (low+slow).

	M (*SD*)		
Prosody style	Fundamental frequency	Duration	Amplitude
high	**288.59(12.6)**	3.23(.05)	75.44(.91)
low	**231.44(6.01)**	3.26(.06)	74.46(1.47)
fast	243.27(1.54)	**2.59(.05)**	76.2(.99)
slow	238.06(2.59)	**4.06(.05)**	75.87(1.69)
high+fast	**276.24(10.71)**	**2.75(.24)**	76.41(.79)
low+slow	**215.54(9.1)**	**4.19(.07)**	74.47(1.39)

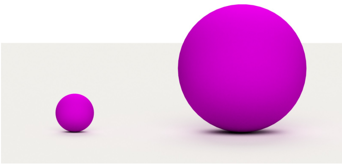
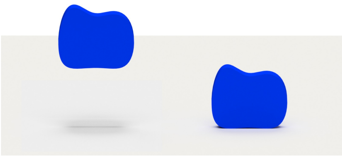
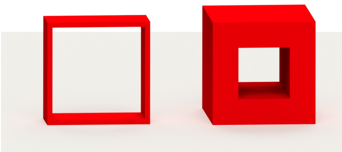
Note. Values for the target features within a prosody style appear in bold.

Visual stimuli. For trials whereby target words were familiar to participants (referred to as *familiar trials* hereafter), participants were presented with side-by-side images of two objects, each representing a noun (e.g. apple and banana). Clip-art images of these objects were sourced from the Picture Perfect stimulus set (Saryazdi, Bannon, Rodrigues, Klammer & Chambers, 2018) (see Table 4.1). For trials whereby target words were novel to participants (referred to as *novel trials* hereafter), the objects were largely the same as those used in Experiment 5 (see section 3.2.1.2 of Chapter 3 for details), with the exception of those designed to represent the antonym pair *thin-thick* being replaced by two skeletal cubes of varying frame thicknesses but of equal overall size (i.e. footprint) (see Table 4.3). For this experiment (with the exception of objects designed to represent varying brightnesses), the objects were also coloured rather than grayscale. To ensure that objects representing each antonym pair (again, with the exception of *bright-dark*) were matched in terms of their perceived brightness, chosen colours were converted temporarily to grayscale, positioned side-by-side and compared and adjusted for their relative brightness prior to being converted back to their original colours (see Figure 4.1).

Figure 4.1. On the left, the four colours chosen to colour the object pairs: (clockwise from top left) *pointy-rounded* (RGB value: 145, 54, 0; HSB value: 22°, 100%, 57%), *high-low* (RGB value: 0, 47, 222; HSB value: 227°, 100%, 87%), *small-big* (RGB value: 181, 0, 181; HSB value: 300°, 100%, 71%), and *thin-thick* (RGB value: 203, 0, 3; HSB value: 359°, 100%, 80%). On the right, the same four colours (in the same configuration) converted to grayscale.

Table 4.3

Antonym pairs and their corresponding objects.

Antonym pair	Objects
small vs big	
high vs low	
pointy vs rounded	
bright vs dark	
thin vs thick	

4.2.1.3 Procedure

Prior to testing, caregivers of participants were issued with a complete list of the familiar words and asked to indicate their child's familiarity with them. This was achieved by caregivers marking the words they believed their child would know the meaning of. The study took part in a quiet, dimly lit room. Infants sat on a caregiver's lap, facing forward, and were positioned approximately 65cm from a computer screen with a resolution of 1280 x 720 pixels, the middle of which was in line with the infants' eyes. A Tobii X60 eye-tracker (Tobii Pro, Stockholm, Sweden) located beneath the screen recorded participants' gaze location at

17ms intervals and a video camera above the screen recorded participants throughout the procedure. Two speakers connected to the computer were positioned either side of the screen, facing the infant. Caregivers were asked not to comment on the stimuli being presented visually or auditorily.

Participants were first shown a short, age-appropriate video clip to fixate their visual attention towards the computer screen. The eye-tracker was then calibrated using a five-point infant calibration procedure. Once calibration accuracy was confirmed, participants were shown a series of three trials, each of which presented images of familiar objects in pairs (e.g. apple and banana) on the screen and were accompanied by short, infant-directed sentences presented auditorily. Sentences were designed to direct the participant's gaze towards one of the two objects presented, for example, "Look it's a banana! Can you see the banana?". Objects were presented on the screen for 6 seconds and sentences presented once. The novel trials followed in a similar format: participants were shown pairs of novel objects and listened to ambiguous sentences describing one of the objects presented. In these trials, the pseudoword used to describe the object (e.g. *temu* pronounced /temu:/) was also novel to participants. In total, participants were presented with five novel trials and a further five familiar trials, which were organised so that all novel trials were followed by at least one familiar trial. The inclusion of familiar trials served to demonstrate and reinforce the task (i.e. to visually locate an object that a speaker refers to). Between each trial, a visual and auditory attention-getter was presented to attract gaze towards the center of the screen. This was manually controlled by the experimenter so that it could be presented for the desired length of time (i.e. until the infant looked at the screen if they had looked away).

In a fully counterbalanced within-subjects design, which novel pseudoword was used to represent a given cross-sensory feature was counterbalanced, along with the prosodic manipulation of the utterance. For instance, *temu* was used to represent size, pointiness, visuospatial height, thinness and brightness for different participants and, within each cross-sensory feature, utterances containing *temu* were manipulated to have an overall higher- or lower-pitch, faster or slower rate, or a higher-pitch and faster rate or lower-pitch and slower rate, yielding thirty possible test trials for a given novel pseudoword. Whether the target object was positioned on the left or right of the screen was randomly generated. The whole procedure took approximately 15 minutes and afterwards participants were given an age-appropriate book for taking part in the study.

4.2.2 Results

4.2.2.1 Caregiver Reported Word Familiarity

Of the 16 familiar words included in the present experiment, caregivers reported their child to be familiar with an average of 10.46 (*SD* = 3.5, min = 4, max = 16).

4.2.2.2 Overview of analysis

The main analyses were run in R (R Core Team, 2015) using the eyetrackingR package (Dink & Ferguson, 2015). Two 640 x 720-pixel areas of interest (AOIs) circumscribed the objects on the screen. Non-AOI looks were not included in the analysis. The dependent measure was the log-transformed proportional looking data (= time spent looking at the congruent object / time spent looking at both the congruent object and the incongruent object). For familiar trials, *congruent object* refers to the object named by the speaker. For novel trials, *congruent object* refers to the object that was in line with the core set of correspondences in the literature for a given prosody style (i.e. higher-pitch sounds are smaller, higher in visual space, pointier, brighter and thinner). Each trial was presented for a duration of 6000ms. Within a trial, the target word occurred twice. The onset of the first and second target word was manually tagged in Tobii Pro Studio and the timestamp of each extracted. For each participant, data from 13 trials (8 familiar and 5 novel) were included in the final analyses.

We began our analysis by averaging looking behaviour across the whole trial and, later, by comparing the difference in looking behaviour between pre-target word onset and post-target word onset time windows. To do this, using the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) in R, we performed linear mixed effects analyses to explore the relationship between proportional looking behaviour and controlled (fixed) and uncontrolled (random) independent variables. The general strategy for creating the models reported in this experiment was to include all fixed and random effects as a first instance and then to eliminate those that reduced the model's overall goodness of fit. This was achieved by comparing each model's Akaike Information Criterion (AIC) using the *AICcmodavg* package (Mazerolle, 2019) in R, which estimates the quality of a statistical model for a given set of data (with a lower AIC value indicating a higher quality model). Once a model reached its optimum (i.e. lowest) AIC, we identified which independent variable/s were having a significant effect on proportional looking behaviour. To do this, *p*-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question. For all statistical tests, an alpha level of .05 was used as a significance criterion. Unless reported, visual inspections of residual plots did not reveal any obvious deviations from homoscedasticity or normality in any of the linear mixed effects analyses reported in this experiment.

4.2.2.3 Whole Trial Analysis

We began our analyses by identifying whether participants looked proportionately longer towards congruent objects (rather than incongruent objects) for familiar trials and novel trials (indicated by the fixed effect of *trial type*). The optimum model included a fixed effect for *trial type* (familiar vs novel) and an intercept for participants as a random effect. Whilst the model revealed a significant difference in proportional looking behaviour for familiar vs novel trials, $\chi^2(1) = 9.31$, $p = .002$, average looking times to the congruent object for familiar trials ($M = .53$, $SD = .12$) and novel trials ($M = .46$, $SD = .13$) were no different from chance, $p = .41$ and $p = .12$, respectively. Mean proportional looks to the congruent object for familiar and novel trials are displayed in Figure 4.2.

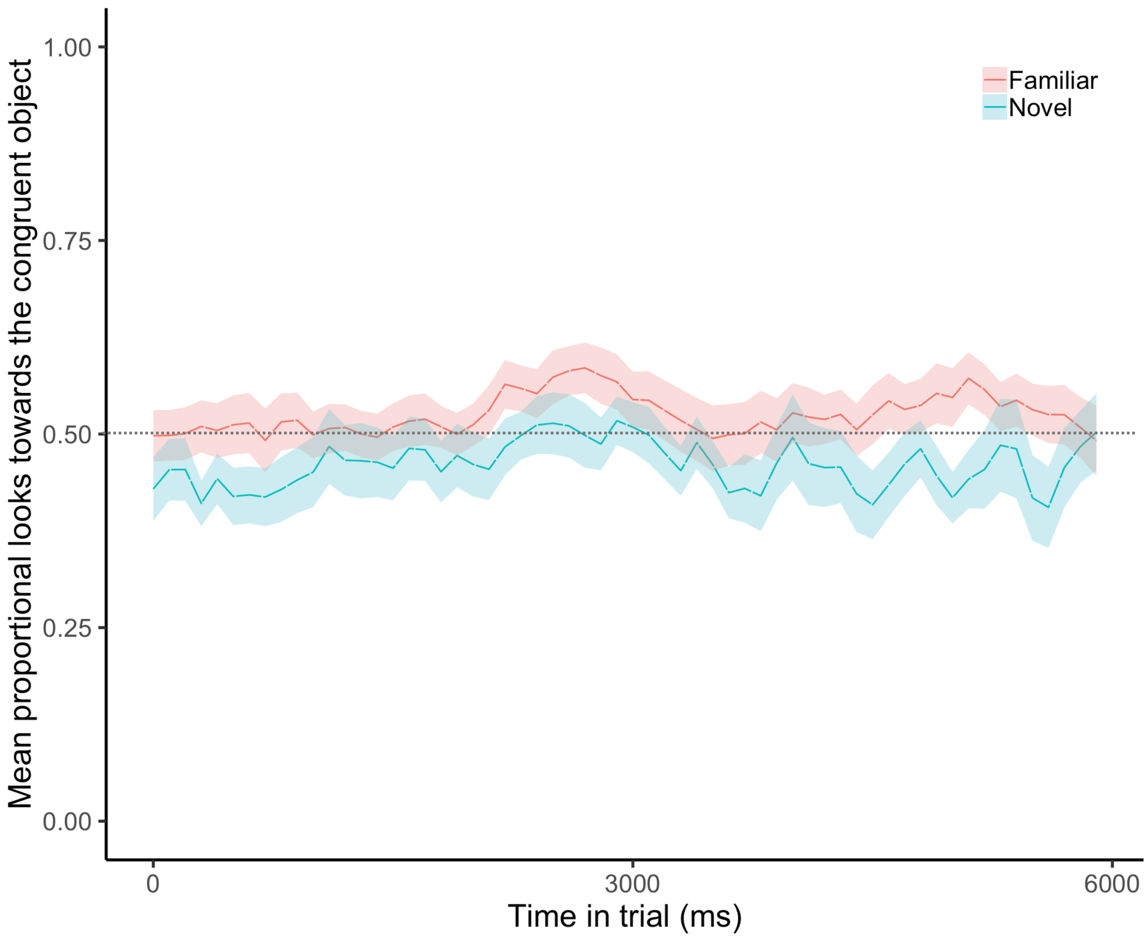


Figure 4.2. Mean proportional looks towards the congruent object for familiar and novel trials, with 95% confidence intervals.

To explore how infants responded to the different visual and auditory stimuli in novel trials, a second linear mixed effects analysis was performed, with *antonym pair* (i.e. small/big, high/low, pointy/rounded, bright/dark and thin/thick), *prosody* (i.e. high, low, fast, slow, high+fast and low+slow) and an interaction term for *antonym pair*prosody* entered into the model as fixed effects, including an intercept for participants as a random effect. The model revealed a significant difference in proportional looking behaviour across levels of *antonym pair*, $\chi^2(4) = 4.27$, $p = .002$, and for the interaction between *antonym pair* and *prosody*, $\chi^2(20) = 2.03$, $p = .007$. Proportional looking behaviour across levels of *prosody* were not significant.

Using the emmeans package in R, the estimated marginal means[19] were obtained for the model and indicated a decrease in looks towards the congruent object for the antonym pairs *high/low*, *thin/thick*, and *small/big*, all of which (apart from *small/big* which was marginally significant) were significantly below chance; $p = .001$, $p = .02$, and $p = .07$, respectively. These findings suggest that infants looked longer towards objects that were higher in space, thinner and smaller in response to hearing a novel word spoken in a relatively lower-pitch tone of voice and/or a slower rate of speech. For the interaction between *antonym pair* and *prosody*, the model-estimated marginal means are displayed in Table 4.4.

---

[19] Model-estimated marginal means provide a mean response for each predictor variable in a given model, adjusted for all other covariates in the model.

Table 4.4

Model-estimated marginal means, standard error (SE) and significance values (p) for the interaction between antonym pair and prosody in novel trials for Experiment 6.

Antonym Pair	Prosodic Feature	Estimated marginal mean	*SE*	*p*
bright-dark	fast	-0.42	.70	.55
	high	0.25	.73	.73
	high+fast	0.17	.68	.81
	low	1.37	.68	.05
	slow	1.97	.76	.01*
	low+slow	0.22	.68	.75
high-low	fast	-0.52	.68	.45
	high	-0.32	.68	.64
	high+fast	-0.86	.68	.21
	low	-2.12	.68	.002**
	slow	-2.36	.76	.002**
	low+slow	0.60	.73	.42
pointy-rounded	fast	2.10	.68	.003**
	high	-0.40	.68	.56
	high+fast	-0.83	.68	.23
	low	-1.58	.68	.02*
	slow	-0.14	.68	.84
	low+slow	0.82	.68	.23
small-big	fast	0.06	.68	.92
	high	-0.45	.68	.51
	high+fast	0.29	.68	.67
	low	-0.78	.68	.25
	slow	-0.36	.68	.60
	low+slow	-1.79	.68	.009**
thin-thick	fast	-1.12	.68	.10
	high	-0.91	.68	.18
	high+fast	-0.39	.68	.56
	low	-0.68	.68	.32
	slow	-0.61	.68	.37
	low+slow	-0.17	.68	.81

Note. Model specified in R as Proportion Looking ~ Antonym Pair * Prosody + (1|Participants)

For ease of interpretation, prior to extracting the estimated marginal means for the model, the data were corrected so that the chance level was set at a value of 0. Estimated marginal

means were then compared against the corrected chance value, with positive values representing a higher proportion of looks towards the congruent object and negative values represent a lower proportion of looks towards the congruent object (because non-AOI looks were not included in the analysis, these are interpreted as a higher proportion of looks towards the incongruent object).
* $p < .05$, ** $p < .01$, otherwise non-significant

In summary, for objects that differed by their brightness, participants looked more towards the congruent object (i.e. darker) when prosody was lower-pitch or slower. For *visuospatial height*, participants looked more towards the incongruent object (i.e. higher) when prosody was both lower-pitch and slower. For *pointiness*, participants looked more towards the congruent object (i.e. pointier) when prosody was faster and looked more towards the incongruent object (i.e. pointier) when prosody was lower-pitch. Finally, for *size*, participants looked more towards the incongruent object (i.e. smaller) when prosody was both lower-pitch and slower.

4.2.2.4 Pre- and Post-Target Word Analysis

The results presented in section 4.2.2.3 of this chapter show that, averaged across the length of a trial, proportional looking behaviour was not significantly different from chance for familiar trials or novel trials. However, looking behaviour might have changed over the course of a trial, perhaps around the onset or offset of a target word (for which there were two instances in a given trial e.g. “Ooh look, a *rebo* one! Where’s the *rebo* one?”). To explore this possibility, linear mixed effects analyses were performed to identify whether proportional looking behaviour differed across time windows immediately before and after a target word was presented.

For the first target word in a trial, the pre-target word time window was defined by the period of time between the start of a trial and the onset of the first target word, and the post-target word time window defined by the period of time between the onset of the first target word and 3000ms (i.e. the middle of the trial). For the second target word in a trial, the pre-target word time window was defined by the period of time between 3000ms and the onset of the second target word, and the post-target word time window defined by the period of time between the onset of the second target word and the end of the trial (see Figure 4.3 for an example).

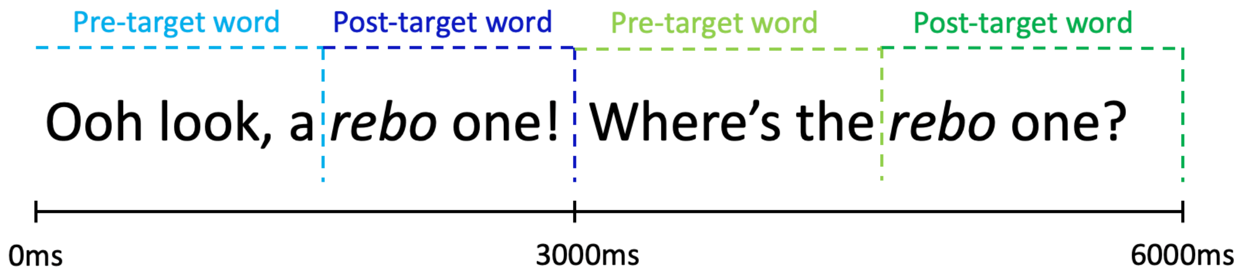


Figure 4.3. Demonstration of pre- and post-target word (i.e. rebo) time windows for the first and second instance of a target word in a trial.

For both models, *pre-/post-target word time windows* was entered as a fixed effect, including an intercept for participants as a random effect. For familiar trials and novel trials, the effect of *pre-/post-target word time windows* was not significant for either the first or second target word in a trial, suggesting that infants looked equally often at congruent and incongruent objects before and after a target word was presented.

4.2.3 Discussion

In response to hearing a novel word spoken in a prosodically deliberate way (e.g. higher- or lower-pitch tone of voice), the results of this experiment showed that 13-month-old infants looked longer towards incongruent objects (in cross-sensory terms) for object-pairs that differed by visuospatial height, thinness and size. However, the interaction between a speaker's prosody and antonym pair (i.e. small/big, high/low, pointy/rounded, bright/dark and thin/thick) on looking behaviour revealed that only for some levels of prosody did infants look longer towards incongruent objects. For instance, infants looked proportionately longer towards higher, smaller and pointier objects in response to speech that was lower in pitch and, for smaller objects only, slower in rate (findings which are incongruent with cross-sensory correspondences: lower-pitch sounds are associated with lower, bigger and rounder objects). However, infants also looked longer towards darker objects in response to hearing a slower rate of speech and looked longer towards pointier objects in response to hearing a faster rate of speech. It appears that for some antonym pairs (i.e. *high/low*, *small/big*, and *pointy/rounded*), infants looked longer towards incongruent objects in response to varying spoken pitches and, for other antonym pairs (i.e. *pointy/rounded* and *bright/dark*), looked longer towards congruent objects in response to varying rates of speech.

These findings might be interpreted in one of two ways: 1. infants were attending to prosody to interpret word meaning but not always in expected directions according to known cross-sensory correspondences (i.e. they associated novel pseudowords spoken in a lower-pitch with objects that were higher in space, smaller or pointier), or 2. infants were not processing the meaning of sentences, so that their looking behaviour reflected conceptual novelty preferences for incongruent objects[20] in response to various pitches of sound rather than pseudoword-object associations. That is, longer looking at incongruent objects indicated a cognitive recognition of a mismatch between prosody and one of the objects within an antonym pair. The latter interpretation is supported by two key findings of this experiment,

---

[20] Given the cross-sensory focus of this thesis, only our findings relating to spoken pitch are interpreted here. In response to various rates of speech, it remains unknown why infants looked longer towards congruent rather than incongruent objects for some antonym pairs (i.e. bright/dark and pointy/rounded).

both of which suggest that infants were not processing sentences' linguistic content: 1. During familiar trials, infants were unable to visually locate familiar objects in response to hearing their labels (despite caregivers advocating their familiarity with these words), and 2. Looking behaviour did not differ immediately before or after a target word was presented (i.e. infants were not triggered by the presence of a familiar or novel word to look towards the word's referent).

A further observation was that infants did not respond (by looking proportionately longer towards one object over another) to sentences spoken in a higher-pitch tone of voice for any of the five antonym pairs. Given the already higher-pitch nature of infant-directed speech, it is possible that utterances spoken in a higher-pitch tone of voice in the present experiment were not high enough (relatively speaking) in order to be detected as meaningful by infants. Just as children were found to require more obvious prosodic cues to infer meaning than adults (Hupp & Junger, 2013), future research should strive to identify at what point infants identify higher-pitch prosody as actually being *higher-pitch*.

4.3 Experiment 7: Prosody as a cue to meaning for adults

The findings of Experiment 6 suggest that 13-month-old infants were not processing sentences' linguistic content and, therefore, were not matching novel pseudowords with their intended referents. Instead, these findings were interpreted as infants demonstrating a novelty preference for incongruent objects in response to the prosodic properties of speech (i.e. pitch). For this reason, these findings can be treated similarly to other cross-sensory research revealing the extent to which correspondences exist in infancy (e.g. infants associate higher-pitch sounds with objects that are *higher* in space, Dolscheid, Hunnius, Casasanto & Majid, 2014; Walker et al., 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018, *smaller*, Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch &

Maurer, 2004, *pointier*, Walker et al., 2010, *brighter*, Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004 and *thinner*, Dolscheid et al., 2014).

The findings of Experiment 6 are somewhat surprising, given that, without training, infants as young as six months have been found to match familiar words with their referents in a similar looking time study (Bergelson & Swingley, 2012). Perhaps, like Friend (2001) demonstrated with 15-month-olds, 13-month-olds are more receptive to prosody than they are to the linguistic content of speech, so much so that they overlook the presence of a familiar or novel word in speech when it is highly prosodic. Alternatively, these findings might be related to the complexity of the task. For instance, target words (novel or familiar) were not presented in isolation (like Bergelson & Swingley, 2012) but embedded within longer speech streams. Despite research showing that infants as young as 6 (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005) and 7.5 (Jusczyk & Aslin, 1995) months possess the ability to segment and extract familiar words from fluent speech, the inclusion of novel pseudowords (without any prior exposure to these pseudowords) both embedded in fluent speech and presented in a prosodically rich style of speech, might have proved to be too cognitively demanding for this age group.

Prior to running a similar study with older infants in Experiment 8, the aim of the present experiment was to identify whether adults use prosodic information to interpret word meaning when presented with the same stimuli as Experiment 6. This decision was motivated by related research, which found that when adult listeners were presented with ambiguous sentences containing a novel pseudoword and spoken in infant-directed speech, they correctly identified the meaning of the pseudoword by attending to the meaningful prosody in which the pseudoword was spoken (Nygaard, Herold & Namy, 2009). Similar to Experiment 6, participants in the present experiment were shown two objects (differing by a single cross-sensory feature, e.g. size) and listened to ambiguous sentences describing one of the two objects (e.g. "Ooh look, a rebo one! Where's the rebo one?"), spoken in either a higher- or lower-pitch tone of voice and/or a faster or slower rate of speech. Rather than relying on looking behaviour to measure preferences for one object over another within a pair, in this experiment participants were asked to manually select the object to which they believed the speaker was referring.

4.3.1 Method

4.3.1.1 Participants

Thirty Lancaster University students and staff (8 males and 22 females) aged between 19 and 62 years ($M = 27.73$, $SD = 9.47$) volunteered to take part in this study in exchange for payment of £3.50. All participants were fluent in English, though not necessarily as their first language, with self-reported normal or corrected-to-normal hearing and vision. All experimental procedures were approved by the FST Research Ethics Committee, Lancaster University, on 11th June 2019 (FST18121).

4.3.1.2 Materials

With the exception of stimuli included in familiar trials, the auditory and visual stimuli used in this experiment were the same as those used in Experiment 6. The experiment was executed using PsychoPy3 (Peirce, Gray, Simpson, MacAskill, Höchenberger, Sogo, Kastman & Lindeløv, 2019).

4.3.1.3 Procedure

Participants were introduced to the study via written instructions presented on an Apple MacBook Pro (2017) 13-inch computer screen. Once participants were confident with the requirements of the task, they indicated their readiness to proceed by pressing the space bar on a QWERTY keyboard and the experiment began. During the study, participants were shown novel object pairs and listened to ambiguous sentences describing one of the objects presented (e.g. "Ooh look, a *temu* one! Where's the *temu* one?"). In these trials, the pseudoword used to describe the object (e.g. *temu* pronounced /temu:/) was also novel to participants. Participants were instructed to pair each pseudoword with one of the two objects presented. Responses were made by pressing either the left or right arrow key on the keyboard, which corresponded with the left-right positioning of the objects (i.e. left arrow key for the object positioned on the left of the screen). Once a choice has been made, the trial ended and, after a short delay, a new pair of novel objects were presented accompanied by another ambiguous sentence, containing a new unfamiliar pseudoword.

In a fully counterbalanced within-subjects design, which pseudoword was used to represent a given cross-sensory feature was counterbalanced, along with the prosodic manipulation of the utterance (e.g. higher-pitch vs lower-pitch). For instance, *temu* was used to represent size, pointiness, visuospatial height, thinness and brightness for different participants and, within each cross-sensory feature, utterances containing *temu* were manipulated to have an overall higher- or lower-pitch, faster or slower rate, or a higher-pitch and faster rate or lower-pitch and slower rate, yielding thirty possible test trials for a given pseudoword. Whether an object was positioned on the left or right of the screen was randomly determined for each participant. The whole procedure took approximately 15 minutes and participants completed a total of 5 trials. At the end of the study, participants were shown an image of the *thin-thick* objects for a second time and asked to indicate the feature by which the two objects differed.

4.3.2 Results

4.3.2.1 Overview of analysis

Using R (R Core Team, 2013), logistic regression analyses were performed to identify the independent variables that were predicting the dependent variable, object choice. For purpose of statistical analysis, choices that were in line with the core set of correspondences in the literature (i.e. higher-pitch sounds are higher in space, smaller, pointier, brighter and thinner) were coded as *congruent* and given a value of 1, with contradictory choices coded as *incongruent* and given a value of 0. Failure to make a choice was coded as missing data.

The general strategy for creating the model reported in this experiment was to include all known variables that might be having an effect on our dependent variable as a first instance and then to eliminate those that reduced the model's overall goodness of fit. This was achieved by comparing each model's Akaike Information Criterion (AIC)[21] using the *AICcmodavg* package (Mazerolle, 2019) in R. Once the model reached its optimum (i.e. lowest) AIC value, a series of Wald tests were performed in order to identify whether independent variables were having a significant effect on object choice. For all statistical tests, an alpha level of .05 was used as a significance criterion. Unless reported, visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

4.3.2.2 Main analyses

The optimum model included two predictor variables: *prosody* (i.e. how sentences were spoken and included 6 levels: higher-pitch, lower-pitch, faster rate, slower rate, higher-

[21] AIC provides an estimate of the quality of a statistical model for a given set of data (with a lower AIC value indicating a higher quality model).

pitch and faster rate, and lower-pitch and slower rate) and *objects* (i.e. which antonym pair was presented visually to participants on a given trial and included 5 levels: *small/big*, *high/low*, *pointy/rounded*, *bright/dark* and *thin/thick*). The output of the model is summarised in Table 4.5 and shows that levels of *prosody* and levels of *objects* were not having a significant effect on whether an object chosen was congruent or incongruent with cross-sensory correspondences. Wald test confirmed that the overall effect of *prosody* was not significant on object choice, Wald $\chi^2(5) = 3.7$, $p = .59$, nor was the overall effect of *objects*, Wald $\chi^2(4) = 1.9$, $p = .76$.

4.3.2.3 Thin-thickness identification

When asked to indicate which feature differed across *thin-thick* objects, of the 30 participants, 25 (83%) correctly identified thinness and 5 (17%) identified size. The difference in proportions between thinness and size was significant, $\chi^2(1) = 24.07$, $p < .001$, showing that the majority of participants reported thinness rather than size as being the distinguishing feature across these objects.

Table 4.5

Summary of logistic regression analysis, including Regression Coefficient (B), standard error (SE), odds ratio (Exp[B]), and 95% Confidence Intervals.

Predictor	Regression Coefficient (B)	*SE*	Exp(B)	95% CI
(Intercept)	.1	.53	-	-
Prosody (base = fast)				
high	.55	.57	.65	[-.56, 1.7]
high+fast	.67	.58	.75	[-.47, 1.83]
slow	.47	.58	.57	[-.65, 1.62]
low	.4	.58	.5	[-.73, 1.56]
low+slow	.65	.58	.74	[-.48, 1.82]
Objects (base = bright/dark)				
high/low	-.16	.54	-.06	[-1.22, .89]
pointy/rounded	.15	.57	.24	[-.93, 1.23]
small/big	-.4	.53	-.31	[-1.46, .66]
thin/thick	-.42	.53	-.33	[-1.48, .61]

Note. Model specified in R as Congruency ~ Prosody + Objects
For Regression Coefficient (B), positive values indicate an increase in the probability of choosing the congruent object over the incongruent object and negative values indicate a decrease in the same probability.

4.3.3 Discussion

The results of the present experiment showed that adults did not pair objects with pseudowords based on the prosody in which the pseudowords were spoken. One possibility is that participants were not attending to the prosody of sentences but were instead matching pseudowords with objects based on the sentences' linguistic content (e.g. a pseudoword's phonetic sound). Given our efforts in Chapter 2 to identify a list of pseudowords that were neutral in terms of their sound-symbolic potential, it is unsurprising that, by including our list of pseudowords as a predictor variable in the logistic regression model for this experiment, the model's overall goodness of fit was reduced (i.e. a pseudoword's sound did not predict the pseudoword-object pairings that we observed). Therefore, if participants were matching pseudowords with objects based on their linguistic content, they were doing so in a seemingly random and inconsistent way.

A second possibility is that the stimuli employed in the present experiment were inappropriate at facilitating pseudoword-object associations. Whilst this might be the case for auditory stimuli (i.e. prosody, see below), it appears that visual stimuli (i.e. objects) were suitable at representing our antonym pairs (i.e. small/big, high/low, pointy/rounded, bright/dark, and thin/thick). That is, when adults were asked to report the feature by which objects within a pair contrasted (see section 3.2.2.1 of Chapter 3 and, for *thin/thick*, section 4.3.2.3 of this chapter), they identified the cross-sensory feature correctly (i.e. size, height, pointiness, brightness and thinness).

Perhaps, then, the prosodic cues included in the present experiment (and indeed Experiment 6) were unsuitable at representing a pseudoword's cross-sensory meaning. Recall in Experiment 5 of Chapter 3, infant-directed adult speakers manipulated their spoken pitch and amplitude to reflect an object's visuospatial height (higher-pitch and louder voice for objects positioned higher in space) but manipulated their spoken pitch, amplitude and rate of speech for objects of varying sizes (higher-pitch, quieter and faster for smaller objects). For the present experiment, variations to a speaker's amplitude were removed from our investigation (due to the observation that amplitude does not align with pitch e.g. higher objects were louder and smaller objects were quieter). By choosing to only manipulate a speaker's pitch of voice and rate of speech, it is possible that participants were unable to match pseudowords with their intended referents. In other words, like adult speakers produce, adult listeners require the full combination of prosodic features (i.e. pitch, speed and amplitude) to interpret a word's precise meaning. This explanation is supported by other research showing that adults could only interpret words when their unique prosodic signatures were fully intact (Nygaard, Herold & Namy, 2009).

4.4 Experiment 8: Prosody as a cue to meaning for 24-month-olds

To reiterate, the aim of this chapter is to identify whether infants attend to non-emotional prosody to resolve linguistic uncertainty (i.e. interpret word meaning), and whether they do so in a way that capitalises on the known cross-sensory correspondences they are sensitive to. In Experiment 6 we found that 13-month-old infants looked preferentially towards incongruent objects (in cross-sensory terms) in response to various pitches of speech. For instance, infants looked longer towards objects that were higher in space, smaller and pointier, when speech was lower in pitch. However, the results also suggested that infants were not processing the linguistic content of speech, with infants being unable to match familiar words with their referents and their looking behaviour towards each object unchanging immediate before and after a target word was presented (i.e. infants did not look proportionately more towards either object within a pair after hearing its label). Employing the same stimuli as Experiment 6, in Experiment 7, we found that adult listeners were unable to interpret the meaning of pseudowords at all. These findings were interpreted as adult listeners requiring more precise prosodic cues to word meaning than was available to them and is supported by evidence revealing the unique acoustic signatures of different words (Nygaard, Herold & Namy, 2009; see also Experiment 5 of Chapter 3). It is unknown at this point whether infants also require precise prosodic cues to word meaning or if some cues (such as pitch) are enough on their own to establish pseudoword-object associations. Based on the behaviour of 13-month-olds in Experiment 6 (i.e. infants looked more towards incongruent objects in response to various pitches of speech), we predict that infants are less sensitive to prosodic cues to meaning than adults are. Therefore, in order to advance in our investigation and continue pursuing the experimental approach adopted thus far, we predict that older infants, with more developed language abilities than 13-month-olds, would be more suitable as participants.

4.4.1 Manual search as an alternative measure to looking time

Historically, screen-based looking time studies (like Experiment 6) have been a popular choice for research related to perception and cognition in preverbal infants. However, there are concerns with using looking time measures and with using screens to present stimuli to infants. For instance, looking behaviour is subject to varying interpretations, with the decision to look towards or away from a stimulus either reflecting recognition, surprise or interest (and not always clear which). Research also shows that infants under the age of 24-months acquire new words more easily if word learning occurs during a live interaction rather than through a screen (DeLoache, et al. 2010; Krcmar, 2011; Krcmar, 2014). For instance, Kuhl, Tsao and Liu (2003) found that when English-speaking 9-month-old infants were presented with regular presentations of adults speaking Mandarin, a sensitivity to Mandarin speech sounds only occurred for those infants that had observed a live rather than virtual speaker. There are many possibilities as to why screen learning fails to promote language development. Although it might include many of the right mechanisms for learning (e.g. word repetition and gestures), television lacks interpersonal social cues and the multimodal input that is provided by live interactions. Krcmar argues that the high visual and auditory stimulation that typically accompany screen media can also be overtaxing for an infant's processing system, so much so that infants fail to learn from them through not keeping up.

With these concerns in mind and to address the artificial nature of the procedure in Experiment 6, in the present experiment we present a real-life rather than screen-based version of the same search task. Rather than rely on looking behaviour to measure pseudoword-object associations, in the present experiment infants were required to identify the object that they believed a speaker was referring to by manually selecting the object and placing it in a box.

4.4.2 Age of participants

The decision to replace a screen-based looking time study with a real-life manual search task, requires a reevaluation of the age of our participants. On a practical level, there are physical limitations that need to be considered when making this decision, such as hand size (i.e. at what age are infants' hands large enough to manipulate objects of varying sizes?), strength (i.e. when are they strong enough to lift objects of varying weights?) and range of motion (i.e. when can they sit upright unaided, lean forwards and reach for objects?). There are also other considerations related to infants' language abilities (e.g. when are infants old enough to understand the linguistic content of speech?) and, given the social nature of the task (i.e. to help a speaker locate an object), their social abilities (e.g. when do infants engage with helping behaviours?). In brief, research shows that 14-month-olds do not engage with helping behaviours as readily as older infants (Hobbs & Spelke, 2015). Instead, an understanding of and willingness to assist in other's goals appears to be in place by 24 months of age (Hepach, Vaish, Grossmann & Tomasello, 2016; Krogh-Jespersen, Liberman & Woodward, 2015).

With these considerations in mind, in the present experiment we explored our aims with 24-month-old infants. During the experiment, infants were presented with three-dimensional object pairs and were asked by a live speaker to manually locate (i.e. place in a box) one of the objects. As with Experiment 6, for some trials, the target word and objects were familiar to infants and, for others, the target word and objects[22] were novel. For novel trials, sentences containing novel pseudowords were spoken in either a higher-pitch tone of voice and faster rate of speech or in a lower-pitch tone of voice and slower rate of speech, with the expectation that infants will match pseudowords with objects based on the cross-sensory nature of a speaker's prosody (e.g. higher-pitch tone of voice and faster rate of speech with objects that are higher in space).

4.4.3 Method

4.4.3.1 Participants

Ten male and 14 female infants (M = 24.10 months; range: 24.01 – 24.20 months) completed the study. A further 8 infants were excluded from the analysis because of excessive restlessness. All participants spoke English as a first language and were raised in homes where English was spoken primarily. All experimental procedures were approved by the FST Research Ethics Committee, Lancaster University, on 12th November 2018 (FST18027).

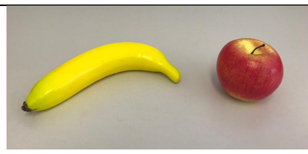
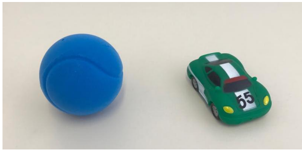
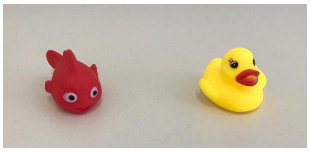
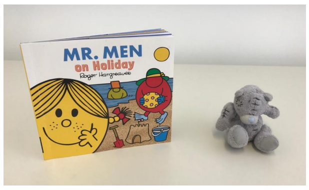
[22] Given the known cross-sensory correspondence between auditory pitch and weight (i.e. higher-pitch sounds feel lighter, Takashima, 2018), the multimodal nature of the task employed in the present experiment allowed for the opportunity to explore weight as an additional cross-sensory feature.

4.4.3.2 Materials

Auditory stimuli. The pseudowords included in this experiment were the same as those used in Experiment 5 (see section 3.2.1.2 of Chapter 3 for details), with the addition of a sixth pseudoword (again, selected as being *neutrally* sound-symbolic): *tiru* (pronounced /tɪru:/ and rhymes with *igloo*). As described in section 4.2.1.2 of Chapter 4, fourteen familiar words were also sourced from The Oxford Communicative Development Inventory database (Hamilton, Plunkett & Schafer, 2000). The familiar words were all nouns and selected from the database as being amongst the first 100 words produced by 24-month-old infants in the United Kingdom (see Table 4.6).

Table 4.6

Fourteen familiar words and their corresponding object pair representations.

Familiar Words	Object Pair
banana vs apple	
ball vs car	
fish vs duck	
sock vs shoe	
book vs teddy bear	
tree vs flower	
bed vs chair	

The novel pseudowords were all embedded within the same phrase: "(participant's name), can you find the (pseudoword) one? Which is the (pseudoword) one?" Familiar words were each embedded in one of three phrases (which word was presented in which phrase was randomly determined for each participant):

"(participant's name), look at the (familiar word). Can you find the (familiar word)?"; "(participant's name), can you find the (familiar word)? Where's the (familiar word)?"; and, "(participant's name), where is the (familiar word)? Can you help me find the (familiar word)?" During a live interaction task, a female speaker (aged 25 years) read these sentences, containing either a novel pseudoword or a familiar word, aloud to participants. The speaker was instructed to employ infant-directed speech throughout the study. Having worked with infants under the age of 36-months for 3 years, the speaker had extensive experience communicating with children of this age. For sentences containing a novel pseudoword, the speaker was instructed to manipulate her prosody across entire sentences (rather than just the pseudoword) in one of two ways: speaking with a relatively higher-pitch tone of voice and a faster rate of speech (high+fast) or speaking with a lower-pitch tone of voice and a slower rate of speech (low+slow)[23]. For sentences containing familiar words, the speaker was instructed to speak with a natural, infant-directed prosody. The whole procedure was video recorded using four Sanyo VCC-MC600P, colour pan-tilt-zoom (PTZ) cameras, each positioned in one of four different corners of the testing room and audio recorded using an Olympus DS-3500 Digital Voice Recorder.

[23] Unlike in Experiment 6 and Experiment 7 where sentences were electronically modified to be higher or lower in pitch, faster or slower in rate, or higher in pitch and faster in rate or lower in pitch and slower in rate, the naturalistic nature of the present experiment meant that this would have been difficult to achieve in a controlled way (i.e. manipulating spoken pitch without altering rate of speech, live and on cue). In preparation for the present experiment, we found that infant-directed speakers naturally increased their rate of speech when talking in a higher-pitch (as opposed to lower-pitch) tone of voice. For this reason, only two prosodic manipulations where included in the present experiment: high+fast and low+slow.

Objects. For familiar trials, participants were presented with two objects (see Table 4.6), each representing a noun (e.g. apple and banana). The footprint of all objects fit in to a 15cm (width) x 20cm (length) x 10cm (depth) box. For novel trials, the objects were printed as physical three-dimensional replicas of the computer-aided design models used in Experiment 6 (see section 4.2.1.2 of Chapter 4), with the addition of two objects designed to represent varying weights (see Table 4.7). Objects were manufactured out of Polylactic acid (PLA) using a Modela MDX-15 3D printer. For the present experiment, novel object pairs represented opposite dimensions within one of six antonym pairs: high/low, small/big, bright/dark, pointy/rounded, thin/thick and light/heavy (the details of which are outlined below).

High/low. Two identical novel objects measuring 95mm (width) x 76mm (height) x 30mm (depth) and spaced 25cm apart, vertically, represented opposite values of visuospatial height (*high* vs *low*).

Small/big. Two spheres, one with a diameter of 50mm and the other a diameter of 100mm, represented opposite values of size (*small* vs *big*, respectively). The colour and texture of the spheres were identical.

Bright/dark. Two novel objects measuring 75mm (width) x 90mm (height) x 30mm (depth) and identical in shape, weight and texture, represented opposite values of brightness (*bright* vs *dark*). The *bright* object was coloured white and the *dark* object was coloured black.

Pointy/rounded. Two novel objects, each with five 'points', represented opposite values of visual and tactile pointiness (*pointy* vs *rounded*). This was achieved by adjusting and matching the corner radiuses on all edges of the objects, which increased from 0 mm (pointy) to 8 mm (rounded). Increasing the corner radius of objects like these, whilst maintaining their volume (i.e. amount of material), results in a reduction of object size, such that the object with a larger corner radius would have a smaller footprint. To address this issue and avoid a confound between overall size and pointiness, each object was scaled so that its footprint fit snugly in to a 90 mm (width) x 90 mm (length) x 30 mm (depth) box.

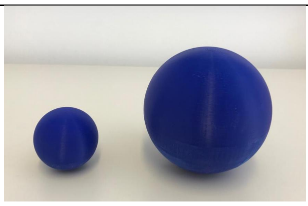
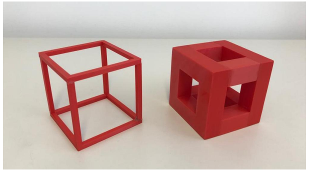
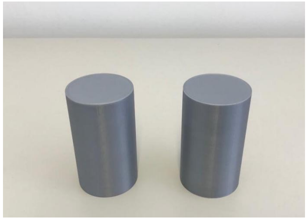
Thin/thick. Two skeletal cubes measuring 70mm (width) x 70mm (height) x 70mm (depth) and identical in colour and texture, represented opposite values of thinness (*thin* vs *thick*). For the *thin* object, the width of the skeleton frame was 5mm compared with 18mm for the *thick* object.

Light/heavy. Two visually identical hollow cylinders measuring 50mm (diameter) x 86mm (height), represented opposite values of weight (*light* vs *heavy*). The *light* cylinder remained hollow, was sealed and weighed 50g. The *heavy* cylinder was filled with solid steel, sealed and weighed 210g.

All objects were presented to participants in a three-dimension wooden display unit, measuring 55cm (width) x 60cm (height) x 30cm (depth). The unit had two wooden shelves extending the width of the unit and which were positioned 25cm apart, horizontally. When sitting and facing the unit, the higher of the two shelves was positioned above infants' line of sight, with the lower of the two shelves positioned below. The unit was backed in black, opaque fabric.

Table 4.7

Antonym pairs and their corresponding objects.

Antonym pair	Object pair
small vs big	
high vs. low	
pointy vs. rounded	
bright vs. dark	
thin vs. thick	
light vs. heavy	

4.4.3.3 Procedure

Prior to testing, caregivers of participants were issued with a complete list of familiar words and asked to indicate their child's familiarity with them. This was achieved by caregivers marking the words that they believe their child would know the meaning of. The study took place in a quiet room, with the caregiver present throughout the whole procedure. Throughout the study, infants sat on the floor, in front of and facing the display unit. The speaker sat directly behind the display unit, positioned so that they were unable to see the objects presented to participants (or their left-right positioning) but could see and interact with the participant clearly over the top of the unit. Throughout the experimental session, the speaker's hands were hidden from the participant and eye-contact maintained. A second experimenter sat to the right of the display unit and positioned the object pairs on the shelves (approximately 15cm apart) for each trial. For novel trials, it was the responsibility of the second experimenter to select which object pair (i.e. objects that differed by their size, visuospatial height, pointiness, brightness, thinness or weight) was presented to participants for a given novel pseudoword and without disclosing this information to the speaker. Unless prompted, caregivers were asked not to comment on the stimuli (i.e. objects presented or sentences spoken) but, otherwise, allowed to interact naturally with their child.

Participants were first introduced to the study and asked "Would you like to play our game? We need your help to find some objects." A trial began when two objects were in position on the display unit. These objects were either both novel or both familiar to the participant. The speaker then asked the participant to locate one of the two objects with either a novel pseudoword or familiar word and place it in a 30cm (width) x 10cm (height) x 20cm (depth) opened box, positioned at the right-hand side of the participant. For instance, "Where is the banana? Can you help me find the banana?" For novel trials, the speaker adjusted her spoken pitch and rate of speech so that sentences were either spoken with a higher-pitch tone of voice and a faster rate or a lower-pitch tone of voice and a slower rate. Once the participant had indicated their choice of object (either verbally and/or with a gesture e.g. reach or grab), the speaker asked, “Can you put it in the box?” and the location of the box gestured to by the second experimenter. After an object was selected by the participant, the speaker provided verbal feedback. For familiar trials, feedback was positive for locating the correct object (e.g. “Well done!”) and neutral for locating the incorrect object (e.g. “Good try! Let’s try another one.”). For novel trials, feedback was always positive. At the end of a trial, both objects were tidied away by the second experimenter and two new objects were positioned on the display unit.

In a fully counterbalanced within-subjects design, which novel pseudoword was used to represent a given dimensional adjective was counterbalanced, along with the prosodic manipulation of the utterance (i.e. higher-pitch and fast rate vs lower-pitch and slow rate). For instance, *temu* was used to represent size, pointiness, visuospatial height, thinness, brightness and weight for different participants and, within each cross-sensory feature listed above, sentences containing *temu* were spoken in either a higher-pitch tone and faster rate or a lower-pitch tone and slower rate, yielding twelve test trials for a given novel pseudoword. Whether an object was positioned on the left or right of the display unit was randomly generated. The whole procedure took approximately 20 minutes and participants were given an age-appropriate book for taking part in the study.

4.4.4 Results

4.4.4.1 Caregiver Reported Word Familiarity

Of the 14 familiar words included in the present experiment, caregivers reported their child to be familiar with an average of 13.5 (*SD* = .94, min = 10, max = 14).

4.4.4.2 Overview of analysis

We began our analyses by checking whether the speaker in the present experiment manipulated her prosody as intended. We then explored whether infants successfully matched familiar words with their referents and whether infants matched novel pseudowords with objects based on the cross-sensory nature of a speaker's prosody.

Using R (R Core Team, 2013), logistic regression analyses were performed to identify the variables that were predicting our dependent variable, object choice. The general strategy for creating the models reported in this experiment was to include all known variables that might be having an effect on our dependent variable as a first instance and then to eliminate those that reduced the model's overall goodness of fit. This was achieved by comparing each model's Akaike Information Criterion (AIC)[24] using the *AICcmodavg* package (Mazerolle, 2019) in R. Once the model reached its optimum (i.e. lowest) AIC value, a series of Wald tests were performed in order to identify whether independent variables were having a significant effect on the dependent variable. For all statistical tests, an alpha level of .05 was used as a significance criterion. Unless reported, visual inspections of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

4.4.4.3 Acoustic Analyses

To assess whether the speaker in the present experiment manipulated her prosody as intended (i.e. higher-pitch and faster in rate vs lower-pitch and slower in rate), each utterance was acoustically analysed as described in section 3.2.2.2 of Chapter 3 using PRAAT (Boersma & Weenink, 2017). Sentence-level mean values for fundamental frequency,

[24] AIC provides an estimate of the quality of a statistical model for a given set of data (with a lower AIC value indicating a higher quality model).

duration and amplitude are displayed in Table 4.8. Paired-samples t-tests were conducted to compare values for fundamental frequency, duration and amplitude for sentences spoken in a higher-pitch and faster rate (high+fast) vs sentences spoken in a lower-pitch and slower rate (low+slow), all of which were significant: for fundamental frequency, $t(70) = 35.34$, $p < .001$; for duration, $t(70) = -12.47$, $p < .001$; and for amplitude[25], $t(70) = 4.83$, $p < .001$. In summary, the speaker employed a higher spoken pitch, faster rate of speech and louder amplitude for prosody that intended to be higher-pitch and faster in rate compared with prosody that intended to be lower-pitch and slower in rate.

Table 4.8

Sentence-level mean values (with standard error in parenthesis) for fundamental frequency (Hz), duration measured in seconds and amplitude measured in dB according to each prosody style: higher-pitch and faster rate (high+fast) and lower-pitch and slower rate (low+slow).

	M (*SE*)		
Prosody	Fundamental frequency	Duration	Amplitude
high+fast	318.64(2.09)	4.01(.23)	68.25(.97)
low+slow	206.24(1.7)	5.01(.27)	66.2(.97)

[25] Despite being instructed not to manipulate the amplitude (i.e. loudness) of her voice, the speaker employed a relatively louder voice when speaking in a higher-pitch and at a faster rate compared with speaking in a lower-pitch and at a slower rate.

4.4.4.4 Familiar Trials

Next, we examined whether infants matched familiar words with their referents. For purpose of statistical analysis, correct word-object pairings (e.g. "banana" paired with banana) were assigned a value of 1, with incorrect word-object pairings (e.g. "banana" paired with apple) assigned a value of 0. Across all familiar words (rather than at an individual word level), the effect of familiar word ($M = .9$, $SD = .3$, min = 0, max = 1) on whether participants selected the correct or incorrect object was significant, Wald $\chi^2(1) = 71.4$, $p < .001$. That is, participants were more likely to select correct objects (and did so 90% of the time) rather than incorrect objects in response to being asked to locate familiar target words.

4.4.4.5 Novel Trials

Following on from Familiar Trials, we examined whether infants interpreted the meaning of novel pseudowords by matching the prosody in which the pseudowords were spoken with the cross-sensory features of the objects displayed. For purpose of statistical analysis, choices that were in line with the core set of correspondences in the literature (i.e. higher-pitch sounds are higher in space, smaller, pointier, brighter, thinner and lighter in weight) were coded as *congruent* and given a value of 1, with contradictory choices coded as *incongruent* and given a value of 0. Failure to make a choice was coded as missing data. The optimum model included three predictor variables: *prosody* (i.e. how sentences were spoken and included 2 levels: higher-pitch and faster rate or lower-pitch and slower rate), *objects* (i.e. which antonym pair was presented to participants on a given trial and included 6 levels: *small/big*, *high/low*, *pointy/rounded*, *bright/dark*, *thin/thick* and *light/heavy*) and an interaction term for *prosody*objects*. The overall effect of *prosody* (high+fast vs low+slow) on whether participants selected the congruent or incongruent object was not significant, Wald $\chi^2(1) = 2.4$, $p = .12$, nor was the effect of the interaction between *prosody* and *objects* (despite approaching significance), Wald $\chi^2(5) = 10.7$, $p = .06$. However, the effect of *objects* on whether participants selected the congruent or incongruent object was significant, Wald $\chi^2(5) = 12.1$, $p = .03$. The means and standard errors of the means for each object within an antonym pair are displayed in Figure 4.4.

Visual inspection of the means presented in Figure 4.4 suggested that the probability of choosing the congruent object differed within antonym pairs. For instance, for the antonym pairs *bright-dark* and *pointy-rounded*, participants appeared to select the darker and rounder object more often than the brighter and pointier object (illustrated by the lower mean congruency values for *bright* and *pointy* compared with *dark* and *rounded*, respectively). To check whether any of these differences were statistically significant, we performed Fischer's Exact tests on the data for each antonym pair and found that only for *bright-dark* was the difference marginally significant, $p = .07$.

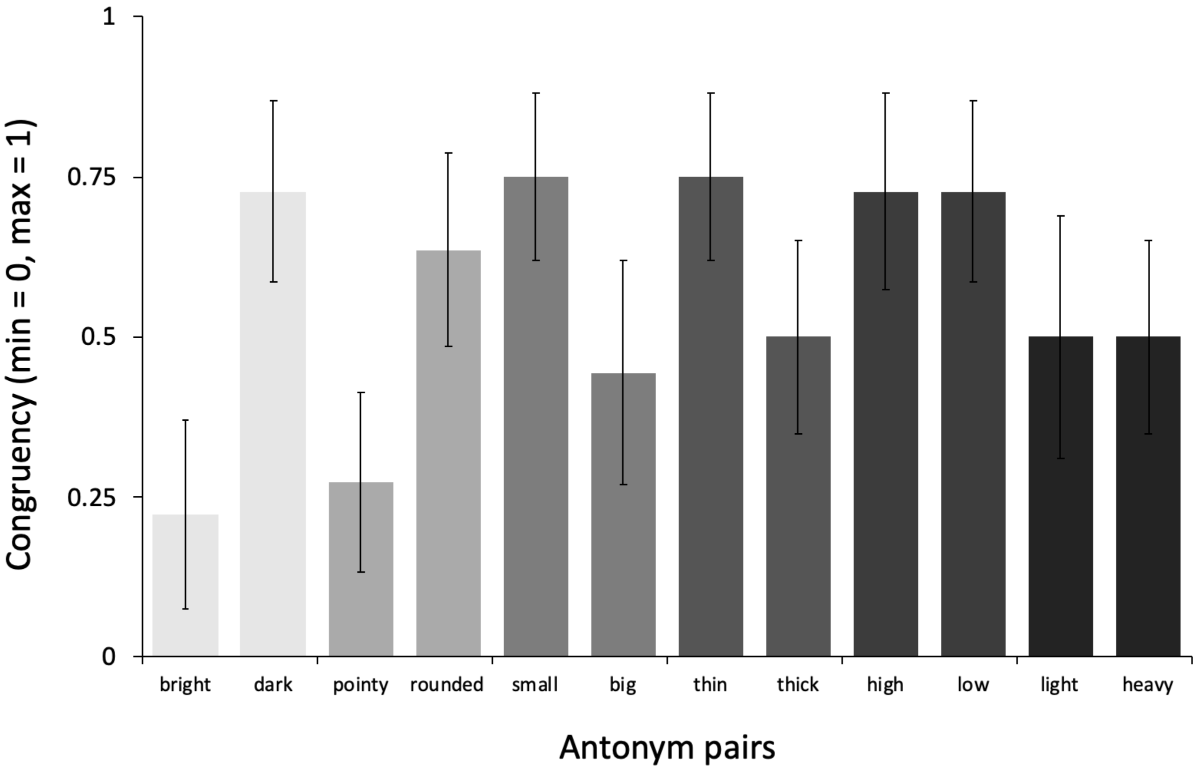


Figure 4.4. Mean values for response congruency (min = 0, max = 1, null = .5) according to antonym pairs (*bright-dark, pointy-rounded, small-big, thin-thick, high-low* and *light-heavy*). Bars represent standard errors of the means.

To identify which antonym pair/s were driving the observed effect of *objects*, the estimated marginal means[26] were obtained for the model using the emmeans package in R. The results are displayed in Table 4.9 and indicate an increased chance of choosing the congruent object for objects that differed by visuospatial height (i.e. *high-low*) only. That is, for sentences that were spoken in a higher-pitch and faster rate of speech or lower-pitch and slower rate, participants were more likely to select objects positioned higher or lower in visual space, respectively.

---

[26] Model-estimated marginal means provide a mean response for each predictor variable in a given model, adjusted for all other covariates in the model.

Table 4.9

Model-estimated marginal means, standard error (SE) and significance values (p) for levels of objects in novel trials for Experiment 8.

Antonym Pair	Estimated marginal mean	*SE*	*p*
bright-dark	-.14	.52	.80
pointy-rounded	-.21	.46	.65
small-big	.44	.47	.35
thin-thick	.55	.44	.21
high-low	.98	.48	.04*
light-heavy	>.01	.46	1.00

Note. Model specified in R as Congruency ~ Antonym Pair * Prosody
For *estimated marginal mean*, positive values indicate an increase in the probability of choosing the congruent object over the incongruent object and negative values indicate a decrease in the same probability.
* $p < .05$, otherwise non-significant

4.4.5 Discussion

The results of the present experiment demonstrate that 24-month-olds will attend to prosody in order to interpret ambiguous word meaning related to visuospatial height. For novel pseudowords spoken in a higher-pitch tone of voice and at a faster rate of speech, infants matched pseudowords with objects positioned higher in visual space and matched novel pseudowords spoken in a lower pitch tone of voice and at a slower rate of speech with objects positioned lower in visual space. These findings echo those from Experiment 5, where we found that infant-directed speakers manipulated their spoken pitch in the same direction as reported here (higher-pitch tone of voice for visually higher objects). In both cases, the cross-sensory correspondence between auditory pitch and visuospatial height is in evidence, suggesting that infant-directed adult speakers and infant listeners are both capitalising on a pre-existing sensitivity to cross-sensory correspondences in order to communicate and interpret meaning, paralinguistically (i.e. via the non-lexical elements of speech).

The observation that infant-directed adult speakers and infant listeners use spoken pitch in consistent ways raises an important question: do speakers recruit a prosody style in response to an infants' pre-existing sensitivity to pitch-height correspondences or do infants learn prosodic cues to meaning through the spoken language adults are using? The former hypothesis is supported by the emergence of pitch-height correspondences in newborns (Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018) and in infants as young as four months old (Dolscheid, Hunnius, Casasanto & Majid, 2014; Walker, Bremner, Mason, Spring, Mattock, Slater, & Johnson, 2010). However, this early sensitivity to correspondences has been demonstrated for many more object features (for a review see section 1.2 of Chapter 1), and yet participants in the present experiment did not respond to the cross-sensory nature of prosody for any features other than visuospatial height. One possibility is that, whilst a sensitivity to a range of cross-sensory correspondences is in place at birth, other factors impact an infant's ability to draw on this information to promote language comprehension. For instance, how speakers in their environment naturally manipulate their pitch of voice to communicate paralinguistically (as introduced above) or how their native language describes auditory pitch (e.g. in terms of spatial height or otherwise). Future research should address these questions with non-English speaking infants, specifically for languages that describe auditory pitch in ways other than spatial height (e.g. for Farsi or Turkish where auditory pitch is described in terms of thinness, see Shayan, Ozturk, Bowerman & Majid, 2014).

Whilst the decision to include a live speaker contributes to the authenticity of the present experiment, producing sentences afresh for each participant introduces confounds that have the potential to impact word-object associations. For instance, sentences that were produced with a higher-pitch tone of a voice and a faster rate of speech were also unintentionally louder than sentences produced with a lower-pitch tone and a slower rate. Recall in Experiment 5, infant-directed speakers employed a higher-pitch tone of voice to refer to objects that were positioned higher in space and (although only marginal at a sentence-level) smaller in size, but amplitude corresponded with visuospatial height and size in opposite directions. That is, speakers employed a louder voice to refer to *higher* and *bigger* objects and employed a quieter voice to refer to *lower* and *smaller* objects. Should infants also be sensitive to correspondences between amplitude and visuospatial height or size, incongruencies in the direction of auditory pitch and amplitude has the potential to hinder performance in the present experiment. In the case of *size*, words spoken in a higher-pitch tone of voice and a louder amplitude could be matched with either a smaller or bigger object (depending on which prosodic cue is most salient to the infant) or could result in a situation whereby neither object is judged as being appropriate (i.e. the symbolism of prosodic cues cancel each other out). Whilst evidence points towards the possibility that different object features carry their own unique acoustic profiles (see Nygaard, Herold and Namy, 2009, and Chapter 3 of this thesis), it remains unclear whether prosodic cues also carry varying degrees of symbolism and how they combine (interactively or additively) to promote language comprehension.

Behaviour in the present experiment has been interpreted as resolving linguistic uncertainty, in the sense that infants matched unfamiliar words to novel objects of varying visuospatial heights based on the prosody in which the words were spoken. This conclusion is supported by participants' performance in familiar trials (whereby infants successfully matched familiar words with their referents), indicating that participants were confident with the requirements of the task. But do infants retain the meaning of new words if they are taught in a congruent (in cross-sensory terms) style of prosody? Whilst there are only so many words (in English at least) that signify visuospatial height (e.g. high, low, tall, short, elevated, lowered etc.), the present experiment showed that by simply presenting an object in a higher or lower visual location and speaking in a higher- or lower-pitch tone of voice, respectively, word-object associations can be formed. That is, when infants heard novel pseudowords spoken in a higher-pitch tone of voice, they were attracted to higher visuospatial locations and their 'choice' of object to pair with pseudowords was determined by whichever object was occupying the space. A next step for future research is to identify whether this has lasting effects on language acquisition.

4.5 General Discussion

The aim of this chapter was to identify whether infants attend to a speaker's prosody to resolve linguistic uncertainty (i.e. interpret word meaning), and whether they do so in a way that capitalises on the known cross-sensory correspondences that they are sensitive to. The findings from the experiments reported in this chapter are summarised in more detail below but, in short, we found that 24-month-old infants will map pseudowords spoken in a higher-pitch tone of voice and a faster rate of speech with objects positioned higher in space and will map pseudowords spoken in a lower-pitch tone of voice and a slower rate of speech with objects positioned lower in space. This is the first evidence to date that infants rely on prosodic cues to interpret the cross-sensory meaning of unfamiliar words.

4.5.1 Summary of Experiments 6 - 8

When 13-month-olds were asked to visually locate a novel pseudoword spoken in a higher- or lower-pitch voice and/or a faster or slower rate of speech in Experiment 6, we found that they would look longer towards incongruent objects (in cross-sensory terms) for object pairs that differed in their visuospatial height, thinness and size. Rather than mapping pseudowords to objects, these findings were interpreted as reflecting conceptual novelty preferences for incongruent objects in response to various pitches of sound (in this case, speech). Like Friend (2001) who found that 15-month-olds were more likely to attend to the emotional prosody of speech rather than its linguistic content, our findings indicated that infants were not processing sentences' linguistic content at all. This interpretation was supported by the same infants failing to visually locate familiar objects within an otherwise identical procedure. If infants were processing linguistic content in Experiment 6, we would expect them to look preferentially towards a referent in response to hearing its label, especially when this label is known to them.

In Experiment 7, we presented adults with the same stimuli as Experiment 6 and found that they were unable to interpret the meaning of pseudowords, showing no preference for any objects in response to hearing pseudowords spoken a higher- or lower-pitch voice and/or a faster or slower rate of speech. With earlier research showing that adults are able to interpret word meaning in a similar, infant-directed context (Nygaard, Herold and Namy, 2009), we concluded that our stimuli failed to capture the precise prosodic cues to word meaning that adults required in this task. An avenue for future research would be to design stimuli that is informed by our findings in Experiment 5 of Chapter 3 and then to replicate Experiment 7. In Chapter 3, we found that speakers employed a higher-pitch and louder voice to refer to objects that were higher (as opposed to lower) in space and employed a (marginally) higher-pitch, quieter and faster rate of speech for objects that were smaller (as opposed to bigger) in size. On reflection, it is thought that these acoustic signatures of height and size must remain fully intact in order to convey meaning for adults but whether they also needed to remain intact for infants was unknown at that point.

In order to advance in our aim of establishing whether infants attend to non-emotional prosody to interpret word meaning, in Experiment 8 we ran a live and manual version of the same search task as Experiment 6 with 24-month-olds and found that infants mapped novel pseudowords spoken in a higher-pitch and faster rate of speech to objects positioned higher in space (and vice-versa). Infants' ability to correctly match familiar words with their referents confirmed that, in this experiment, infants were attending to the linguistic content of speech and understood the requirements of the task (i.e. to locate an object named by the speaker). This observation further strengthens our argument that, when a word is unknown to an infant, they will use other tools at their disposal (in this case, prosody) to make predictions about which object a speaker is referring to. An obvious next step for future research is to identify how early on in development the ability to use prosody in this way occurs. With infants demonstrating a sensitivity to emotional prosody throughout the first year of life (at 3- and 5-months, Walker-Andrews & Grolnick, 1983, at 7-months, Grossmann, Striano & Friederici, 2005, and at 9-months, Otte, Donkers, Braeken & Van den Bergh, 2015) and with newborns demonstrating a sensitivity to cross-sensory correspondences (for pitch-height, see Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018), we predict that 24-months of age does not mark the onset of this ability.

4.5.2 What is it about visuospatial height?

A recurring theme to emerge from the series of experiments reported in this thesis, is that infant-directed adult speakers and infant listeners capitalise on the correspondence between auditory pitch and visuospatial height to communicate and interpret word meaning.

Despite evidence that infants and adults are sensitive to a range of other cross-sensory correspondences, it appears that only visuospatial height is successfully communicated at a paralinguistic level. Perhaps these findings emerge because the correspondence between auditory pitch and visuospatial height is the most salient of all correspondences. The wealth of evidence demonstrating their existence across different ages (in infants, Dolscheid, Hunnius, Casasanto & Majid, 2014; Walker et al., 2010; Walker et al., 2018, in children, Nava, Grassi & Turati, 2016, and in adults, Bonetti & Costa, 2018; Chiou & Rich, 2012; Evans & Treisman, 2010) and across different languages (e.g. for Kreung, see Parkinson, Kohler, Sievers, & Wheatley, 2012) would certainly support this, but a similar volume of evidence exists for other correspondences, too (for a review, see section 1.2 and section 1.3 of Chapter 1).

A second possibility is that infants are born with an equal sensitivity to the full range of correspondences, but the saliency of a correspondence is environmentally mediated. Factors that might contribute to the prominence of a correspondence include (but not limited to), how native languages choose to describe auditory pitch (e.g. in terms of *height* like English and German, in terms of *thinness* like Farsi or Turkish, or in terms of *tightness* like Kreung), how speakers manipulate their spoken pitch when referring to different object properties (as demonstrated by Chapter 3 of this thesis, infant-directed adult speakers manipulate their spoken pitch to refer to objects of varying visuospatial heights and, although only marginal, sizes) and the extent to which correspondences co-occur naturally in the environment (e.g. birds are small in size, make a high-pitch sound and are typically found above us in the sky). Whether or not some correspondences are more salient than others, should be addressed by future research, as should the possibility that, like English-speaking infants only relying on prosody to interpret words that are related to visuospatial height, a speaker’s use of prosody to convey word meaning might also vary in response to some of these same factors. For instance, perhaps for languages that describe auditory pitch in terms of *thinness* or *tightness*, speakers are less inclined to manipulate their spoken pitch to reflect visuospatial height but instead do so to reflect thinness or tightness.

CHAPTER 5

Summary, critical reflection and further research

5.1 Introduction

The aim of this thesis was to identify the functional significance of cross-sensory correspondences in infant-directed speech. This was achieved by exploring the ways in which infant-directed speakers employ prosodic cues to meaning that are cross-sensory related and whether infants attend to and interpret prosody when language is otherwise ambiguous. To begin, a general summary of the research conducted in this thesis and their main findings are presented, including a review of the theory in light of our work. Following this, we provide a critical reflection on the work carried out in this thesis. In doing so, recommendations as to how this work could be supported and extended are provided in the context of improving our research and in the context of its broader implications.

5.2 Summary of Research

The approach to addressing the aim of the thesis was to present sentences containing novel psuedowords to participants and explore, firstly, how adults use prosody to communicate the meaning of these pseudowords (Chapter 3) and, secondly, how infants interpret meaning by attending to the prosody in which pseudowords are spoken (Chapter 4). Prior to this, a review of the literature related to sound-symbolism revealed the tendency for some speech sounds (e.g. vowel and consonant sounds) to be associated with objects of varying shapes, sizes, weights and tastes, the most notable of which being object pointiness, with research showing that infants as young as four months old associate words such as *kiki* and *kipi* with pointier shapes and *bubu* and *moma* with rounder shapes (Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada & Kita, 2015; Ozturk, Krehm & Vouloumanos, 2013;

Peña, Mehler & Nespor, 2011). In this context, sound-symbolism could be defined as an extension of cross-sensory correspondences, with the features related to one sensory modality (in this case, words encoded auditorily) being mapped to the features of another (i.e. pointiness encoded visually or tactually). Unsurprisingly, overlaps between the fields of sound-symbolism and cross-sensory correspondences are reported in the literature. For instance, when four-month-old infants were presented with two objects of varying sizes and the auditory presentation of a single vowel phoneme, infants looked preferentially towards smaller objects for vowel sounds produced nearer the front of the mouth and towards larger objects for vowel sounds produced nearer the back (Peña, Mehler & Nespor, 2011). Having analysed the intrinsic fundamental frequency of these phonemes, Peña et al. found that vowels produced nearer the front of the mouth were higher in auditory pitch than those produced nearer the back, echoing the same correspondence found between auditory pitch and size (i.e. higher-pitch sounds are judged as being smaller in size).

In Chapter 2, the importance of teasing apart sound-symbolism and cross-sensory correspondences in the context of both the present thesis and other research concerned with correspondences and language was discussed. Collectively, Experiments 1 – 4 identified a list of novel pseudowords that were judged to possess varying levels of sound-symbolism. In Experiment 1 and Experiment 2, participants rated pseudowords according to their graphemic structural appearance, and in Experiment 3 and Experiment 4, paired these same pseudowords, based on their sound, with objects representing a progression of visual pointiness. In summary, psuedowords containing a larger number of pointy-sounding phonemes compared with rounded-sounding phonemes were rated as being pointier in visuostructural appearance and were more likely to be paired with pointier objects. These findings support earlier work showing that pseudowords containing a larger number of large-sounding phonemes compared with small-sounding phonemes were more likely to be paired with larger objects (Thompson & Estes, 2011) and contribute to the understanding of how phonemic and visual features combine to predict a word's symbolic impact. By identifying a list of pseudowords that extends the full range of symbolic pointiness, we were able to isolate pseudowords that were judged as being *neutrally* sound-symbolic (i.e. they were neither pointy or rounded in visuostructural appearance, nor were they associated with pointy or rounded objects). The purpose of this was to ensure that pseudowords paired with objects later on in this thesis were non-symbolic in terms of their sound. Given that the overall aim of this thesis was to explore prosodic correlates to word meaning in infant-directed speech, without excluding sound-symbolism from our investigations, associations formed between words and objects could be the result of infants detecting sound-symbolic biases, cross-sensory correspondences related to a speaker's prosody, or to some unknown combination of these two factors.

In Chapter 3, we extended earlier research by Nygaard, Herold and Namy (2009) to explore how infant-directed speakers communicate meaning via the non-lexical elements of speech (i.e. prosody). In Experiment 5, participants were presented with object pairs, contrasting by a single cross-sensory feature (e.g. visuospatial height, size, pointiness, brightness or thinness). Adopting an infant-directed style of speech, participants were required to read aloud simple sentences containing a novel, neutrally sound-symbolic pseudoword (e.g. "Look at the *rebo* one."), each time associating the novel pseudoword with just one of the objects within a pair. In accordance with known cross-sensory correspondences between auditory pitch and visuospatial height and size, participants employed a higher spoken pitch to refer to objects positioned higher in space and (although only marginal) smaller in size. These findings led to the conclusion that for some cross-sensory related features (i.e. visuospatial height and size), speakers capitalise on correspondences to communicate meaning. They also support earlier research showing similar pitch-height (Shintel, Nusbaum & Okrent, 2006) and pitch-size (Nygaard et al., 2009) mappings in speech.

Whether infants, like children (Hupp & Junger, 2013; Herold, Nygaard, Chicos & Namy, 2011), attend to prosody to interpret a word's cross-sensory meaning remained unknown at this point in the thesis. To address this question, in Experiment 6, 13-month-old infants were presented with the same object pairs as adults in Experiment 5, alongside short, infant-directed sentences asking them to visually locate one of the two objects using a neutral, novel pseudoword (e.g. "Ooh look a *rebo* one. Where's the *rebo* one?"). Sentences were spoken in either a higher- or lower-pitch tone of voice, a faster or slower rate of speech, or a combined higher-pitch and faster rate or lower-pitch and slower rate. A surprising finding to emerge from Experiment 6 was that infants looked preferentially towards incongruent objects (in cross-sensory terms) for object pairs that contrasted by visuospatial height, thinness and (although only marginal) size. That is, in response to prosody that was lower-pitch and/or slower, infants looked more towards objects that were positioned higher in space, thinner or smaller. Chance-level performance in trials in which infants were presented with familiar objects (e.g. apple and banana) and asked to visually locate one of the objects using a caregiver-reported familiar word (e.g. "Look at the banana. Can you see the banana?"), indicated that infants might not have been processing the sentences at a linguistic-level. In this case, increased looking towards incongruent objects during novel trials (i.e. where objects and pseudowords were both novel to infants) was interpreted as reflecting a novelty preference (i.e. a preference for objects that mismatch with the more general acoustic properties of prosody), indicating that infants were capitalising on cross-sensory correspondences in this context but not to interpret a word's meaning. In this case, these findings contribute to the literature revealing the range of cross-sensory correspondences infants are sensitive to, including pitch-height (Dolscheid, Hunnius, Casasanto & Majid, 2014; Wagner, Winner, Cicchetti & Gardner, 1981; Walker et al., 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018), pitch-thinness (Dolscheid et al., 2014), and pitch-size (Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004).

Like earlier research (Nygaard, Herold and Namy, 2009), in Experiment 7 we checked whether adults could interpret word meaning in a near-identical context as Experiment 6. We found that adults were unable to match novel pseudowords with objects in either their expected or unexpected directions. For instance, in response to hearing a pseudoword spoken in a higher-pitch tone of voice and/or faster rate of speech, adults showed no preference for higher, smaller, pointier, brighter or thinner objects. By choosing to manipulate spoken pitch and rate of speech only, it was concluded that the acoustic profiles of our sentences failed to capture word meaning for adults. This conclusion was supported by our findings from Experiment 5, where we found that adults used spoken pitch, rate of speech and amplitude in unique ways to reflect different word meanings (e.g. they employed a higher-pitch and louder voice for visually *higher* objects and a higher-pitch, faster and quieter voice for visually *smaller* objects). The findings from Experiment 7 tentatively support other research showing that adults can only interpret a word's meaning when its prosodic, acoustic signature is fully intact (Nygaard et al., 2009).

Rather than replace our stimuli entirely, we chose to run a final experiment and extend the work carried out in Experiment 6 with older infants. In Experiment 8, we introduced a real-life version of the same search task, including a live speaker and tangible versions of the same objects presented visually in Experiment 6 and Experiment 7 (with the addition of two objects representing contrasting weights). As the first body of research to explore whether infants capitalise on cross-sensory correspondences to interpret the meaning of unfamiliar words, it was deemed unnecessary at this point to identify how early on in development this occurs but rather if it occurs at all. For this reason, Experiment 8 explored these aims with 24-month-old infants, for which receptive and expressive language is more established than at 13 months old and a willingness to assist in tasks of this kind is demonstrated (Hepach, Vaish, Grossmann & Tomasello, 2016; Krogh-Jespersen, Liberman & Woodward, 2015). In this experiment, infants correctly identified familiar objects in response to hearing their labels and matched novel pseudowords spoken in a higher-/lower-pitch tone of voice and faster/slower rate of speech with objects positioned higher/lower in space, respectively. These findings echo those from Experiment 5, where we observed infant-directed speakers adjusting their spoken pitch in the same direction as reported here (higher-pitch for higher objects), and provides the strongest evidence to date that, for correspondences related to visuospatial height, infants will engage with their pre-existing sensitivity to cross-sensory correspondences in order to interpret the meaning of ambiguous words. This work contributes to the growing literature revealing the extent to which prosody is used by speakers (Hanuliková & Haustein, 2016; Herold, Nygaard & Namy, 2011; Hupp & Junger, 2013; Nygaard, Herold & Namy, 2009; Nygaard & Lunders, 2002; Shintel, Nusbaum & Okrent, 2006; Tzeng, Duan, Namy & Nygaard, 2018) and listeners (Hupp & Junger, 2013; Nygaard, Herold & Namy, 2009) to communicate semantic information. Importantly, we identified that the ability to infer a word's meaning based on a speaker's prosody does not begin at 4 years of age, which, up until this point, was the earliest this had been recorded (Herold, Nygaard, Chicos & Namy, 2011; Hupp & Junger, 2013).

5.3 Critical reflection

Having summarised the main findings and identified how they contribute to the wider literature above, this section will provide a critical reflection on the work carried out with the view to discuss the challenges associated with this research and identify areas that could be improved. In doing so, we will also consider where additional work would need to be undertaken to support the findings of this thesis and extend them. To begin, we reflect on the main finding of this thesis: infants associate novel words spoken in a higher-pitch tone of voice and a faster rate of speech (relatively to a lower-pitch tone and slower rate) with objects positioned higher (rather than lower) in space. Whilst we did not find associations between novel words spoken in a higher- or lower-pitch with any other objects representing different cross-sensory features, in this section we provide a number of explanations as to why this might be the case. Ultimately, we argue that our findings are not exhaustive of an infant's ability to capitalise on cross-sensory correspondences to interpret word meaning. Next, the challenges that we faced in regard to creating our stimuli are explained, including those relevant to the objects, pseudowords and prosodic styles employed in the series of experiments reported in this thesis. We finish by discussing some alternative methods suitable for the research carried out in this thesis.

5.3.1 Visuospatial height is just the beginning

In line with the cross-sensory correspondence between auditory pitch and visuospatial height, one key finding to emerge from this thesis is that infant-directed adult speakers and infant listeners associated words spoken in a higher-pitch prosody with objects positioned higher in space. That is, speakers employed a higher-pitch tone of voice to refer to objects positioned higher (rather than lower) in space and infants associated novel words spoken in a higher-pitch tone of voice and faster rate of speech with these same objects. Surprisingly, we found no evidence to suggest that infants associate words spoken in a relatively higher-pitch voice with smaller, pointier, brighter, thinner or lighter objects, despite infants demonstrating a sensitivity to these sorts of correspondences (for *pitch-size*, Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004, *pitch-pointiness*, Walker et al., 2010; *pitch-brightness*, Haryu & Kajikawa, 2012; Mondloch & Maurer, 2004, and *pitch-thinness*, Dolscheid et al., 2014).

Given that infant-directed speakers and infant listeners used pitch in the same direction for objects that differ by visuospatial height (higher-pitch for higher object) in our research and the robust evidence illustrating that infants (Dolscheid, Hunnius, Casasanto & Majid, 2014; Wagner, Winner, Cicchetti & Gardner, 1981; Walker et al., 2010; Walker, Bremner, Lunghi, Dolscheid, Barba & Simion, 2018), children (Nava, Grassi & Turati, 2016) and adults (Bonetti & Costa, 2018; Evans & Treisman, 2010) appreciate the same correspondence between auditory pitch and visuospatial height, we are confident that our findings are not a chance occurrence. Instead, in section 4.5.2 of Chapter 4, we suggest that these findings might emerge because of the fundamental relative strength of the correspondence between pitch and height, with pitch-height being the most salient of all correspondences. This proposal is supported by the only evidence to date demonstrating cross-sensory correspondences in newborns (Walker, et al., 2018), with newborn infants preferring animations that were in line with pitch-height correspondences (i.e. higher-pitch tone associated with a ball travelling up, rather than down, in space).

An alternative explanation to pitch-height being the most salient correspondence, is that visuospatial height is the easiest cross-sensory feature to represent (this suggestion is discussed further section 5.3.2 of this chapter). That is, is it possible to contrast objects by their visuospatial height only, while preserving all other object features that they might otherwise differ by (e.g. size, shape, brightness, thinness and weight) and which have the potential to affect the detection of cross-sensory correspondences. Having said that, objects rarely contrast by a single feature in real-life, so perhaps only when objects are truly unambiguous in their interpretation can infants use prosody to interpret an unfamiliar word's meaning. Future research should check whether the ability to use a speaker's pitch of voice and rate of speech to identify word-object pairs is limited to stimuli employed by the research reported in this thesis or if infants apply the same rules (e.g. higher-pitch voice for visually higher objects) for objects that differ more considerably.

First and foremost, our findings illustrate that infants will attend to prosody to interpret word meaning. Whilst the correspondence between pitch and height might be the most salient or easiest to represent, based on our findings and the evidence revealing the range of correspondences that preverbal infants are sensitive to, we are unable to rule out the possibility that infants possess the capacity to use prosodic correlates to word meaning for other cross-sensory features as well. Instead, we believe that our findings only scratch the surface of this ability and future research should explore other ways in which correspondences can be accessed by infants through infant-directed speech. For instance, rather than applying a higher-pitch tone and faster rate of speech (relative to a lower-pitch and slower rate) across the entire length of a sentence (which is the approach adopted by Experiment 6, Experiment 7 and Experiment 8 in this thesis), prosodic cues to word meaning could be concentrated to the target word (i.e. novel pseudoword) itself. This suggestion has at least two benefits: firstly, given the relative nature of pitch and speed, increasing prosodic variation across a sentence might help infants to detect when speech is meaningful relative to when it is inconsequential (as demonstrated by Estes & Hurley, 2013). Secondly, by concentrating meaningful prosody to a target word, infants should be able to segment speech streams more easily and, in turn, identify the most prominent part of an utterance (i.e. the unfamiliar word spoken in a prosodically meaningful way).

5.3.2 Stimuli

5.3.2.1 Objects

A challenge for researchers concerned with cross-sensory correspondences is to ensure the appropriateness of stimuli used to represent contrasting cross-sensory features. For some features, such as visuospatial height and brightness, it is possible to vary objects along a single featural dimension while preserving other physical features (in contrast with perceived features) that might otherwise interfere with cross-sensory perception (e.g. size, shape and weight). For others, such as pointiness or thinness, changes to one feature can cause changes to another. For example, pointier objects (of equal mass to rounded objects) are naturally bigger in size(footprint), and thicker objects are both bigger in size and (expected to be) heavier in weight than thinner objects. Even when objects of unequal sizes are manipulated so that they are matched in terms of their mass, research shows that participants continue to perceive differences in felt heaviness, with smaller objects being judged as relatively heavier (Buckingham, Goodale, White & Westwood, 2016; Flanagan & Beltzner, 2000). In a similar way, Walker, Francis and Walker (2010) found that when objects were matched in terms of their size and mass, brighter objects were judged as being lighter in weight than darker objects.

The approach adopted by the present thesis was to design object pairs that differed (where possible) by a single cross-sensory feature. Overall, this was achieved to good effect. For example, we successfully designed objects that contrasted in pointiness without conflating pointiness with perceived size (despite variations in volume). That is, when participants rated tangible versions of these objects through touch alone, they rated them as being similar in overall size (see section 2.4.2 of Chapter 2). However, there were a number of issues with our approach to vary objects by a single feature that emerged along the way. In particular, in Experiment 5 we found that participants reported either size or shape, rather than thinness, as the main feature differentiating thin-thick objects. This led to the replacement of rods of various thicknesses (as employed in earlier research, Dolscheid, Hunnius, Casasanto & Majid, 2014) with two skeletal cubes, whose footprints were identical in size but whose frames increased or decreased in thinness, inwardly. Whilst this modification resulted in participants later identifying thinness as the distinguishing feature across thin-thick objects (see section 4.3.2.3 of Chapter 4), unavoidably, mass continued to vary with thinness, with thicker objects being (and expected to be) heavier in weight. The same applies to objects of contrasting sizes, with larger objects being heavier than smaller objects. These observations highlight the difficulty in representing single featural dimensions in cross-sensory research and illustrate the need for the findings of this and related work to be interpreted with caution. For instance, even with object size(footprint) controlled for, a cross-sensory correspondence involving *thinness* could be attributed to thinness, to weight or to the combined cross-sensory impact of thinness and weight.

Having modified the thin-thick objects part-way through this thesis, it would be useful to repeat Experiment 5 (for which rods of various thicknesses were used to probe meaningful prosody) with the newer versions of thin-thick objects (i.e. skeletal cubes). Whilst objects of contrasting thinness had no significant effect on a speaker's pitch of voice, rate of speech or amplitude, inspection of means suggest that speakers employed a relatively higher-pitch voice for thinner, as opposed to thicker, objects (at a sentence level, see Figure 3.4 of Chapter 3, and at a target-word level, see Figure 3.5 of Chapter 3). With a more accurate set of objects representing thinness, perhaps we will find that an object's thinness can be represented at a prosodic level in this context or perhaps these observed differences will be attributed to the other cross-sensory features (i.e. size or shape) participants reported these objects to differ by.

5.3.2.2 Pseudowords

In Chapter 2 we identified a list of novel pseudowords that were judged to possess varying levels of sound-symbolism. Those that were *neutral* in terms of their symbolic potential were then used in all other experiments reported in this thesis. So as not to detract too heavily from the primary aim of this thesis, we chose to grade our pseudowords in terms of their symbolic pointiness alone. This decision occurred after a review of the literature depicted symbolic pointiness as being the most widely tested and popular form of sound-symbolism (see Bottini, Barilari & Collignon, 2019; Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada & Kita, 2015; Köhler, 1947; Nielsen & Rendall, 2011; Ozturk, Krehm & Vouloumanos, 2013; Ramachandran & Hubbard, 2001), but we could have chosen to grade our pseudowords in terms of their symbolic size (Peña, Mehler & Nespor, 2011; Thompson & Estes, 2011), taste (Ngo, Velasco, Salgado, Boehm, O'Neill & Spence, 2013) or weight (Monaghan, Mattock & Walker, 2012; Walker, Barnett & Parameswaran, submitted; Walker & Parameswaran, 2019) as just a few examples.

Perhaps like cross-sensory correspondences, where higher-pitch sounds are judged as being pointier, smaller, higher in space, thinner, lighter in weight and brighter, the full range of sound-symbolic features also converge with one another, so that words that are judged to be symbolic in terms of their pointiness are also symbolic in terms of their size, taste and weight and in similar directions (e.g. *kiki* is associated with pointier, smaller, sourer and lighter objects and *bouba* is associated with rounder, bigger, sweeter and heavier objects). Or perhaps different words are symbolic in different ways, so that *kiki* is considered to be pointy but not small, sour or light or *bouba* is considered to be round, big and heavy but not sweet. A concern for our research is if pseudowords rated as being *neutral* in terms of their symbolic pointiness were in fact symbolic for other features, especially if any of these features were also represented by the objects that we paired with these same pseudowords in later experiments. In an ideal world, our *neutral* pseudowords would have been neutral for all sound-symbolic features (e.g. pointiness, size, taste and weight) but, for now, this observation opens up an avenue for future research to explore whether sound-symbolic features are aligned. As an immediate next step, it would be useful to replicate Experiment 1 – 4 with different symbolic features (e.g. size, taste or weight) and explore whether our pseudowords are judged differently when participants are asked rate them in terms of their pointiness or in terms of their size, taste or weight.

5.3.2.3 Prosodic cues to word meaning

In the series of experiments reported in Chapter 4, we chose to present participants with sentences containing pseudowords and spoken in a higher- or lower-pitch tone of voice and/or a faster or slower rate of speech. In Experiment 6, we manipulated these acoustic properties of speech electronically, which granted us greater control in our quest to identify which prosodic cues (e.g. pitch of voice, speaking rate or a combination of pitch and speaking rate) influence infants' looking behaviour towards different objects. In this experiment we found that infants showed no preference for any objects in response to hearing sentences spoken in a higher-pitch tone of voice, leading to the conclusion that our higher-pitch sentences were not *high* enough for infants to detect them as being meaningful. If we were to replicate this experiment, we would modify our stimuli so that sentences spoken in a higher-pitch tone of voice are even *higher* but, for now, we encourage future research to explore at what point infants identify higher-pitch prosody as actually being higher-pitch.

Our decision to ignore amplitude as a prosodic cue stemmed from the observation that amplitude, being magnitude-based, is distinct from cross-sensory correspondences. Unlike speed (i.e. speaking rate) which appears to align with auditory pitch, in Experiment 5 we found that infant-directed speakers employed a louder voice for higher objects and a quieter voice for smaller objects. Evidence shows that adults only interpret an unfamiliar word's meaning when its prosodic, acoustic signature is fully intact (Nygaard, Herold & Namy, 2009). The extent of this finding revealed itself in Experiment 7, where we found that adults were unable to identify the cross-sensory meaning of any of our pseudowords. On reflection, the decision to remove amplitude from our investigation might have contributed to this finding.

In Experiment 8, we chose to include a live rather than virtual speaker, which resulted in sentences spoken in a higher-pitch tone of voice and faster rate of speech being unintentionally louder than those spoken in a lower-pitch tone of voice and slower rate of speech. Perhaps it was the combination of pitch, speech rate and amplitude in Experiment 8 that facilitated an effect for visuospatial height, given that the direction of spoken pitch (e.g. higher) was congruent with the direction of amplitude (i.e. louder) for higher objects. In a similar way, this same combination of pitch, speech rate and amplitude might have extinguished an effect for size, given that the directions of pitch (e.g. higher) and speech rate (i.e. faster) were incongruent with the direction of amplitude (i.e. louder) for smaller objects. Recall in Experiment 8 that the speaker was blind to the stimuli presented to participants, but should the speaker have known which object pair was displayed for a given trial and adjusted her prosody in accordance with the findings from Experiment 5, it is possible that we would have achieved a more accurate representation of how infants (and indeed adults in Experiment 7) attend to prosody to interpret word meaning. In an attempt to address this concern, future research should replicate our work using the information obtained in Experiment 5 to better guide decisions about prosodic cues to word meaning. This suggestion is addressed further in section 5.3.3 of this chapter.

5.3.3 Alternative methods

Both looking time (see Dolscheid, Hunnius, Casasanto & Majid, 2014; Fernández-Prieto, Navarra & Pons, 2015; Haryu & Kajikawa, 2012; Lewkowicz & Minar, 2014; Mondloch & Maurer, 2004; Wagner, Winner, Cicchetti & Gardner, 1981; Walker et al., 2010; Walker et al., 2018) and manual search (see Mondloch & Maurer, 2004) have proved to be successful at measuring the existence of cross-sensory correspondences in infants. In Chapter 4, we attempted both methods and, whilst looking time in a screen-based context allowed us to explore our aims with younger participants, manual search in a naturalistic setting provided a clearer interpretation of our results. Despite being in favour of the manual search approach to addressing the aims of this thesis, this method is limited by the age of the infant when a manual response is to be made. One alternative method is to use a head-mounted eye-tracker in a situation in which infants observe a confederate taking part in the same experiment. This would allow for the opportunity to retain a naturalistic setting, whilst recording word-object expectations (e.g. looking towards an object prior to it being selected) with younger participants.

In Experiment 5 adult speakers were asked to employ infant-directed speech without an infant present. This approach is artificial for two reasons: 1. in any other context, adults would rarely be asked to consciously employ infant-directed speech, and 2. speakers would rarely (if ever) use infant-directed speech without communicating to an infant. A simple modification to Experiment 5 would be to replicate the experiment with an infant present. In this context, we would expect to achieve a truer representation of how infant-directed speakers use prosody to communicate word meaning. An extension of this would be to combine Experiment 5 and Experiment 8 and run an observational study with infant and caregiver. Like in Experiment 5, in this observational study, caregivers would be asked to communicate the meaning of unfamiliar words to their children and, similar to Experiment 8, we would observe which objects infants associate with these same words. An obvious advantage of this approach is that infants would be highly familiar with their caregiver's voice and so we would expect them to be more attuned and, in turn, responsive to any meaningful changes to their prosody.

5.3.4 Future directions for the field

In light of the findings related to spoken pitch and visuospatial height, a question that could be addressed by future research is whether this correspondence is linguistically scaffolded (given that for English-speaking infants, auditory pitch is described in terms of its height). Whilst a sensitivity to a range of cross-sensory correspondences is in place at birth (or very early on in development), how native languages describe auditory pitch (e.g. in terms of spatial height or otherwise) may be impacting infants' ability to draw on this information to promote language comprehension. Future research should replicate this work with non-English speaking infants, specifically for languages that describe auditory pitch in ways other than spatial height, such as Farsi or Turkish where auditory pitch is described in terms of its thinness. If infants are drawing on linguistic information in this context, for these alternative languages, we might expect infants to associate spoken pitch with objects of contrasting thinnesses rather than contrasting heights.

In view of the evidence that phonemic and visual properties of words combine to predict a word's sound-symbolic potential, one question that emerged in Chapter 2 is whether a word's visual appearance (i.e. its graphemic structure) enhances a sensitivity to some sound-symbolic exemplars that exist prior to language acquisition. That is, once language can be read, the availability of visual-symbolism enhances a preexisting sensitivity to sound-symbolism, making the presence of sound-symbolism even more influential. Research demonstrating the existence of sound-symbolism in prereading infants (Ozturk, Krehm & Vouloumanos, 2013; Peña, Mehler & Nespor, 2011) supports this possibility by showing that words can be symbolic based on their sound only (i.e. sound-symbolism and visual-symbolism can exist independently from one another). Our collective findings from Experiment 1 – 4 also supports this idea, with the relationship between a word's sound and its visual appearance shown to be a monotonic one: pointier sounding words were rated as visually pointier in appearance. One useful approach to addressing this question would be to explore if infants are sensitive to any sound-symbolic exemplars that literate children and adults are not sensitive to. If they are, it would be worthwhile identifying if children and adults are not sensitive to these sound-symbolic words because they fail to correspond with their visual appearance in printed form.

By establishing that infants use prosody as a cue to novel word interpretation in Chapter 4, a suitable next step would be to identify the impact of this finding in the broader context of language development. For sound-symbolism research, Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada and Kita (2015) found that 14-month-olds were better at retaining the meaning of novel words when audio-visual stimulus pairs were congruent (e.g. *kipi* with a pointy object) rather than incongruent (i.e. *kipi* with a rounded object) with expected sound-symbolic relationships during learning. In a similar way, it would be useful for future research to explore whether novel words are acquired more easily when spoken in a congruent (e.g. higher-pitch tone of voice for visually higher objects) rather than incongruent (i.e. higher-pitch tone of voice for visually lower objects) style of prosody during learning.

5.4 Concluding Remarks

For the first time, the research reported in this thesis has demonstrated that infants capitalise on their sensitivity to cross-sensory correspondences to interpret the meaning of unfamiliar words. Over and above this, this thesis has also demonstrated that infants attend to prosody in a non-emotional context to resolve linguistic uncertainty. Until now, research has tended to focus on establishing the range of cross-sensory correspondences that infants are sensitive to, rather than identifying their functional significance. We hope that this work will encourage future researchers to explore additional ways in which a sensitivity to cross-sensory correspondences might be (or is currently being) capitalised on to aid typical development in infants.

References

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

Barnett, A. M., Bremner, J. G., & Walker, P. (2019). *Higher-pitch tones feel sharper: the sensory domain generality of cross-sensory correspondences.* Manuscript submitted for publication.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258. doi:10.1073/pnas.1113380109

Boersma, Paul & Weenink, David (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.36, retrieved January 2017 from http://www.praat.org/

Bonetti, L., & Costa, M. (2018). Pitch-verticality and pitch-size cross-modal interactions. *Psychology of Music, 46*(3), 340-356. https://doi.org/10.1177/0305735617710734

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and Me. *Psychological Science, 16*(4), 298–304. doi:10.1111/j.0956-7976.2005.01531.x

Bottini, R., Barilari, M., & Collignon, O. (2019). Sound symbolism in sighted and blind. The role of vision and orthography in sound-shape correspondences. *Cognition, 185*, 62-70. https://doi.org/10.1016/j.cognition.2019.01.006

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). “Bouba” and “Kiki” in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition, 126*(2), 165-172. https://doi.org/10.1016%2Fj.cognition.2012.09.007

Broesch, T. L., & Bryant, G. A. (2014). Prosody in Infant-Directed Speech Is Similar Across Western and Traditional Cultures. *Journal of Cognition and Development, 16*(1), 31-43. http://dx.doi.org/10.1080/15248372.2013.833923

Brunetti, R., Indraccolo, A., Del Gatto, C., Spence, C., & Santangelo, V. (2018). Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification. *Attention, Perception, & Psychophysics, 80*(2), 527-534. https://doi.org/10.3758/s13414-017-1445-z

Buckingham, G., Goodale, M. A., White, J. A., & Westwood, D. A. (2016). Equal-magnitude size-weight illusions experienced within and between object categories. *Journal of Vision, 16*(3), 25. https://doi.org/10.1167/16.3.25

Chiou, R., & Rich, A. N. (2012). Cross-Modality Correspondence between Pitch and Spatial Location Modulates Attentional Orienting. *Perception, 41*(3), 339-353. https://doi.org/10.1068/p7161

Cooper, R. P., & Aslin, R. N. (1990). Preference for Infant-Directed Speech in the First Month after Birth. *Child Development, 61*(5), 1584. https://doi.org/10.2307/1130766

Cuskley, C., Simner, J., & Kirby, S. (2017). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research, 81*, 119-130. https://doi.org/10.1007/s00426-015-0709-2

DeLoache, J. S., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G. L., Strouse, G. A., O’Doherty, K. (2010). Do Babies Learn From Baby Media? *Psychological Science, 21*(11), 1570–1574. https://doi.org/10.1177/0956797610384145

Dink, J. W., & Ferguson, B. (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis. Retrieved from http://www.eyetrackingr.com.

Dolscheid, S., Hunnius, S., Casasanto, D., Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science, 25*(6), 1256-1261. https://doi.org/10.1177/0956797614528521

Dunst, C. J., Gorman, E., & Hamby, D. W. (2012). Preference for Infant-Directed Speech in Preverbal Young Children. *CELLReviews, 5*(1), 1-13.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1):6, 1–12. https://doi.org/10.1167/10.1.6

Fernández -Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development, 38*, 77 -81. https://doi.org/10.1016/j.infbeh.2014.12.008

Flanagan, J. R., & Beltzner, M. A. (2000). Independence of perceptual and sensorimotor predictions in the size–weight illusion. *Nature Neuroscience, 3*(7), 737–741. https://doi.org/10.1038/76701

Fort, M., & Alexander-Peperkamp, S. (2015). Consonants are more important than vowels in the bouba-kiki effect. *Language and Speech, 58*(2), 247-266. https://doi.org/10.1177/0023830914534951

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental Science, 21*(5), e12659. doi:10.1111/desc.12659

Friend, M. (2001). The transition from affective to linguistic meaning. *First Language, 21*(63), 219–243. doi:10.1177/014272370102106302

Fryer, L., Freeman, J., & Pring, L. (2014). Touching words is not enough: How visual experience influences haptic–auditory associations in the "Bouba–Kiki" effect. *Cognition, 132*(2), 164–173. http://dx.doi.org/10.1016/j.cognition.2014.03.015

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191-1203. https://doi.org/10.3758/bf03193720

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *NeuroReport, 16*(16), 1825–1828. doi:10.1097/01.wnr.0000185964.34336.b1

Hamilton, A., Plunkett, K., & Schafer, G., (2000). Infant vocabulary development assessed with a British Communicative Development Inventory: Lower scores in the UK than the USA. *Journal of Child Language, 27*, 689-705.

Hanuliková, A., & Haustein, J. (2016). Resolution of lexical ambiguity by emotional prosody in a non-native language. *Speech Prosody 2016*. http://dx.doi.org/10.21437/speechprosody.2016-96

Haryu, E., & Kajikawa, S. (2012). Are higher-frequency sounds brighter in color and smaller in size? Auditory-visual correspondences in 10-month-old infants. *Infants Behavior and Development, 35*(4), 727-732. https://doi.org/10.1016/j.infbeh.2012.07.015

Hepach, R., Vaish, A., Grossmann, T., & Tomasello, M. (2016). Young Children Want to See Others Get the Help They Need. *Child Development, 87*(6), 1703–1714. https://doi.org/10.1111/cdev.12633

Herold, D. S., Nygaard, L. C., & Namy, L. L. (2011). Say it like you mean it: mothers’ use of prosody to convey word meaning. *Language and Speech, 55*(3), 423-436. http://dx.doi.org/10.1177/0023830911422212

Herold, D. S., Nygaard, L. C., Chicos, K. A., & Namy, L. L. (2011). The developing role of prosody in novel word interpretation. *Journal of Experimental Child Psychology, 108*(2), 229-241. http://dx.doi.org/10.1016/j.jecp.2010.09.005

Hobbs, K., & Spelke, E. (2015). Goal attributions and instrumental helping at 14 and 24 months of age. *Cognition, 142*, 44–59. https://doi.org/10.1016/j.cognition.2015.03.014

Hupp, J. M., & Jungers, M. K. (2013). Beyond words: Comprehension and production of pragmatic prosody in adults and children. *Journal of Experimental Child Psychology, 115*(3), 536–551. http://dx.doi.org/10.1016/j.jecp.2012.12.012

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound Symbolism Facilitates Word Learning in 14-Month-Olds. *PLoS ONE, 10*(2), 1-17. http://dx.doi.org/10.1371/journal.pone.0116494

Jusczyk, P. W., & Aslin, R. N. (1995). Infants′ Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology, 29*(1), 1–23. doi:10.1006/cogp.1995.1010

Keetels, M., & Vroomen, J. (2010). No effect of synesthetic congruency on temporal ventriloquism. *Attention, Perception, & Psychophysics, 73*(1), 209–218. https://doi.org/10.3758/s13414-010-0019-0

Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., … Wang, Z. (2003). Effects of Experience on Fetal Voice Recognition. *Psychological Science, 14*(3), 220–224. http://dx.doi.org/10.1111/1467-9280.02435

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development, 24*(4), 372–392. http://dx.doi.org/10.1016/s0163-6383(02)00086-3

Köhler, W. (1947). *Gestalt psychology.* New York: Liveright Publishing.

Krcmar, M. (2011). *Can babies learn new words from Baby Einstein?* Paper presented at the International Communication Association, Boston, MA.

Krcmar, M. (2014). Can Infants and Toddlers Learn Words from Repeat Exposure to an Infant Directed DVD? *Journal of Broadcasting & Electronic Media, 58*(2), 196–214. https://doi.org/10.1080/08838151.2014.906429

Krogh-Jespersen, S., Liberman, Z., & Woodward, A. L. (2015). Think fast! The relationship between goal prediction speed and social competence in infants. *Developmental Science, 18*(5), 815–823. https://doi.org/10.1111/desc.12249

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, 100*(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

Lee, C. S., Kitamura, C., Burnham, D., & McAngus Todd, N. P. (2014). On the rhythm of infant- versus adult-directed speech in Australian English. *The Journal of the Acoustical Society of America, 136*(1), 357–365. https://doi.org/10.1121/1.4883479

Lewkowicz, D. J., & Minar, N. J. (2014). Infants are not sensitive to synesthetic cross-modality correspondences. *Psychological Science, 25*(3), 832-834. http://dx.doi.org/10.1177/0956797613516011

Ludwig, V. U., Adachi, I., Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences of the United States of America, 108*(51), 20661-20665. http://dx.doi.org/10.1073/pnas.1112605108

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development, 7*(3), 185-201. https://doi.org/10.1080/15475441.2011.579839

Marks, L.E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 384–394. https://doi.org/10.1037//0096-1523.13.3.384

Marks, L.E. (1989). On cross-modal similarity: The perceptual structure of pitch, loudness, and brightness. Journal of Experimental Psychology: *Human Perception and Performance, 15*(3), 586-602. https://doi.org/10.1037//0096-1523.15.3.586

Marx, V., & Nagy, E. (2015). Fetal Behavioural Responses to Maternal Voice and Touch. *PLOS ONE, 10*(6), e0129118. http://dx.doi.org/10.1371/journal.pone.0129118

Maurer, D. (1993). *Neonatal synesthesia: Implications for the processing of speech and faces.* In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), Developmental neurocognition: Speech and face processing in the first year of life (pp. 109–124). Dordrecht, The Netherlands: Kluwer Academic. https://doi.org/10.1007/978-94-015-8234-6_10

Maurer, D., Gibson, L. C., & Spector, F. (2013). *Synesthesia in Infants and Very Young Children.* In J. Simner & E. Hubbard (Eds.), Oxford Handbook of Synesthesia. http://dx.doi.org/10.1093/oxfordhb/9780199603329.013.0003

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143–178. http://dx.doi.org/10.1016/0010-0277(88)90035-2

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1152–1164. http://dx.doi.org/10.1037/a0027747

Mondloch, C., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective & Behavioral Neurosciences, 4*(2), 133-136. https://doi.org/10.3758/cabn.4.2.133

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paediatrica, 102*(2), 156-160. http://dx.doi.org/10.1111/apa.12098

Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., & Kojima, S. (2012). Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroIM, 59*(2), 1735–1744. https://doi.org/10.1016/j.neuroiM.2011.07.093

Nava, E., Grassi, M., & Turati, C. (2016). Audio-Visual, Visuo-Tactile and Audio-Tactile Correspondences in Preschoolers. *Multisensory Research, 29*(1-3), 93–111. https://doi.org/10.1163/22134808-00002493

Nielsen, A., & Rendall, D. (2011). The sound of round: evaluating the sound symbolic role of consonants in the classic takete-maluma phenomenon. *Canadian Journal of Experimental Psychology, 65*(2), 115-124. https://doi.org/10.1037/a0022268

Ngo, M. K., Velasco, C., Salgado, A., Boehm, E., O'Neill, D., & Spence, C. (2013). Assessing crossmodal correspondences in exotic fruit juices: The case of shape and sound symbolism. *Food Quality and Preference, 28*(1), 361-369. https://doi.org/10.1016/j.foodqual.2012.10.004

Nygaard, L., & Lunders, E. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition, 30*(4), 583-593. http://dx.doi.org/10.3758/bf03194959

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33*(1), 127-146. http://dx.doi.org/10.1111/j.1551-6709.2008.01007.x

Otte, R. A., Donkers, F. C. L., Braeken, M. A. K. A., & Van den Bergh, B. R. H. (2015). Multimodal processing of emotional information in 9-month-old infants I: Emotional faces and voices. *Brain and Cognition, 95*, 99–106. doi:10.1016/j.bandc.2014.09.007

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. https://doi.org/10.1016/j.jecp.2012.05.004

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on languages: evidence from a remote population. *Perception, 41*(7), 854-861. https://doi.org/10.1068/p7225

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.* 10.3758/s13428-018-01193-y

Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science, 22*, 1419–1421. https://doi.org/10.1177/0956797611421791

Perlman, M., Clark, N., & Falck, M. J. (2015). Iconic prosody in story reading. *Cognitive Science, 39*(6), 1348-1368. https://doi.org/10.1111/cogs.12190

Puigcerver, L., Rodríguez-Cuadrado, S., Gómez-Tapia, V., & Navarra, J. (2019). Vertical Mapping of Auditory Loudness: Loud is High, but Quiet is not Always Low. *Psicológica Journal, 40*(2), 85–104. https://doi.org/10.2478/psicolj-2019-0006

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ramachandran, V. S., & Hubbard, W. S. (2001). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Saryazdi, R., Bannon, J., Rodrigues, A., Klammer, C., & Chambers, C. G. (2018). Picture perfect: A stimulus set of 225 pairs of matched clipart and photographic images normed by Mechanical Turk and laboratory participants. *Behavior Research Methods, 50*(6), 2498–2510. https://doi.org/10.3758/s13428-018-1028-5

Sbattella, L., Colombo, L., Rinaldi, C., Tedesco, R., Matteucci, M., & Trivilini, A. (2014). Extracting Emotions and Communication Styles from Prosody. *Lecture Notes in Computer Science*, 21–42. https://doi.org/10.1007/978-3-662-45686-6_2

Shayan, S., Ozturk, O., Sicoli, M. A. (2011). The thickness of pitch: crossmodal metaphors in Farsi, Turkish, and Zapotec. *The Senses and Society, 6*(1), 96-105. http://dx.doi.org/10.2752/174589311x12893982233911

Shayan, S., Ozturk, O., Bowerman, M., & Majid, A. (2014). Spatial metaphor in language can promote the development of cross-modal mappings in children. *Developmental Science, 17*(4), 636–643. http://dx.doi.org/10.1111/desc.12157

Shintel, H., Anderson, N. L., Fenn, K. M., & Gauthier, I. (2014). Talk this way: the effect of prosodically conveyed semantic information on memory for novel words. *Journal of Experimental Psychology: General, 143*(4), 1437-1442. http://dx.doi.org/10.1037/a0036605

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech. *Cognition, 105*(3), 681–690. doi:10.1016/j.cognition.2006.11.005

Shintel, H., Nusbaum, H. C., Okrent, A. (2006). Analog acoustic expression in speech communication. *Journal of Memory and Language, 55*(2), 167-177. https://doi.org/10.1016/j.jml.2006.03.002

Slavin, S. (2003-2014). PsyScript Version 2.3.0 [Computer software]. Lancaster Univeristy. Available at: http://www.lancaster.ac.uk/psychology/research/research-software/ [accessed October 2016]

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C., & Deroy, O. (2012). Crossmodal Correspondences: Innate or Learned? *i-Perception, 3*(5), 316–318. http://dx.doi.org/10.1068/i0526ic

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects*. London, England: John Wiley.

Takashima, M. (2018). Perceived Weight Is Affected by Auditory Pitch Not Loudness. *Perception, 47*(12), 1196–1199. https://doi.org/10.1177/0301006618808937

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *The Quarterly Journal of Experimental Psychology, 64*(12), 2392-2404. https://doi.org/10.1080/17470218.2011.605898

Tzeng, C. Y., Duan, J., Namy, L. L., & Nygaard, L. C. (2018) Prosody in speech as a source of referential information. *Language, Cognition and Neuroscience, 33*(4), 512-526, DOI: 10.1080/23273798.2017.1391400

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017). Developmental change in children's sensitivity to sound symbolism. *Journal of Experimental Child Psychology, 160*, 107-118. https://doi.org/10.1016/j.jecp.2017.03.004

Wagner, S., Winner, E., Cicchetti, D., Gardner, H. (1981). "Metaphorical" mapping in human infants. *Child Development, 52*(2), 728-731. http://dx.doi.org/10.1111/j.1467-8624.1981.tb03106.x

Walker, P. (2012) Cross-sensory correspondences and cross talk between dimensions of connotative meaning: visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics, 74*(8), 1792–1809. https://doi.org/10.3758/s13414-012-0341-9

Walker, P. (2016). Cross-sensory correspondences and symbolism in spoken and written language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(9), 1339-1361. https://doi.org/10.1037/xlm0000253

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development, 6*(4), 491–498. doi:10.1016/s0163-6383(83)90331-4

Walker, P., Barnett, A. M., & Parameswaran, C. R. (2019). *Cross-Sensory Correspondences in Language: Front-Back and Close-Open Vowel Contrasts Symbolize the Felt Heaviness of Objects.* Manuscript submitted for publication.

Walker, P., Bremner, J. G., Lunghi, M., Dolscheid, S., Barba, B. D., & Simion, F. (2018). Newborns are sensitive to the correspondence between auditory pitch and visuospatial elevation. *Developmental Psychobiology, 60*(2), 216-223. https://doi.org/10.1002/dev.21603

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S.P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21-25. https://doi.org/10.1177/0956797609354734

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S.P. (2014). Preverbal infants are sensitivity to crosssensory correspondences: much ado about the null results of Lewkowicz and Minar (2014). *Psychological Science, 25*(3), 835-836. http://dx.doi.org/10.1177/0956797613520170

Walker, P., Francis, B. J., & Walker, L. (2010). The Brightness-Weight Illusion: Darker Objects Look Heavier But Feel Lighter. *Experimental Psychology, 57*(6), 462-469.

Walker, P., & Parameswaran, C. R. (2019). Cross-sensory correspondences in language: Vowel sounds can symbolize the felt heaviness of objects. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 45(2), 246–252. https://doi.org/10.1037/xlm0000583

Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception, 41*, 1186-1192. https://doi.org/10.1163/187847612x646884

Walker, P., Walker, L., & Francis, B. (2015). The size-brightness correspondence: evidence for crosstalk among aligned conceptual feature dimensions. *Attention, Perception, & Psychophysics, 77*(8), 2694–2710. https://doi.org/10.3758/s13414-015-0977-3

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language, 93*(1), 10-19. https://doi.org/10.1016/j.bandl.2004.07.006

Zangl, R., & Mills, D. L. (2007). Increased Brain Activity to Infant-Directed Speech in 6- and 13-Month-Old Infants. *Infancy, 11*(1), 31–62. https://doi.org/10.1207/s15327078in1101_2

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects. *Attention, Perception, & Psychophysics, 81*(5), 1609–1623. https://doi.org/10.3758/s13414-019-01668-w

Appendix A

Table 2.1

Pseudowords displayed according to their assumed phonological pointiness categories (extreme rounded, rounded, neutral, pointy, extreme pointy), with their corresponding International Phonetic Alphabet (IPA) transcription. In the final four columns, mean ratings for visuostructural pointiness in Experiment 1 and Experiment 2 (min = 1, median = 4, max = 7) and mean ratings for phonological pointiness in Experiment 3 and Experiment 4 (min = 1, median = 3, max = 5) are reported, with standard deviation of the mean (in parenthesis).

Category	4 letter string	Pseudoword	IPA transcription	Mean (*SD*) Experiment 1	Mean (*SD*) Experiment 2	Mean (*SD*) Experiment 3	Mean (*SD*) Experiment 4
Extreme Rounded	r-m-o-u	rumo	/rʌmoʊ/	5.72(1.02)	5.82(.8)	3.75(.92)	3.64(1.25)
		muro	/mʌroʊ/	6.11(.66)	5.77(1.25)	3.78(.91)	3.98(.97)
		romu	/rɒmu:/	5.89(.68)	5.77(.92)	3.78(.91)	3.58(1.18)
		moru	/mɒru:/	6.00(.59)	5.82(.91)	3.91(.78)	3.64(1.09)
Rounded	r-m-o-i	romi	/rɒmi:/	5.33(.97)	4.77(1.23)	3.25(1.06)	3.00(1.14)
		mori	/mɒri:/	5.17(.86)	4.59(1.01)	3.56(.73)	3.34(.97)
		rimo	/rɪmoʊ/	4.33(1.28)	4.41(.96)	3.63(1.31)	2.91(1.00)
		miro	/mɪroʊ/	3.94(1.3)	4.82(1.22)	3.25(1.06)	3.22(1.13)

	r-m-o-e	remo	/remoʊ/	5.89(1.08)	5.59(1.22)	3.75(.86)	3.53(.92)
		mero	/meroʊ/	5.83(1.1)	5.14(1.21)	3.31(.79)	3.47(1.05)
	r-m-i-u	rumi	/rʌmiː/	4.21(1.27)	4.09(1.38)	3.75(1.00)	3.44(1.05)
		muri	/mʌriː/	4.21(1.27)	4.00(1.11)	3.38(.96)	3.28(1.02)
		rimu	/rɪmuː/	3.72(1.27)	4.09(1.06)	3.81(1.05)	3.19(1.03)
		miru	/mɪruː/	3.61(1.58)	4.05(1.05)	3.44(.89)	3.09(.93)
	r-m-e-u	remu	/remuː/	5.28(1.18)	5.05(1.13)	3.75(.93)	3.31(1.06)
		meru	/meruː/	5.39(1.54)	4.95(1.46)	3.38(.81)	3.25(1.02)
	r-t-o-u	rotu	/rɒtuː/	4.44(1.42)	4.45(1.34)	2.19(1.05)	2.91(1.06)
		toru	/tɒruː/	5.22(1.06)	4.73(1.12)	3.50(1.26)	2.75(1.24)
		ruto	/rʌtoʊ/	4.44(1.42)	4.73(1.03)	2.63(1.59)	2.53(1.05)
		turo	/tʌroʊ/	5.06(1.39)	4.64(1.09)	3.50(.89)	3.22(1.04)
	r-b-o-u	robu	/rɒbuː/	5.67(1.19)	5.73(1.24)	3.63(1.02)	3.69(1.03)
		boru	/bɒruː/	5.67(.59)	5.45(.96)	3.63(.89)	3.69(.90)
		rubo	/rʌboʊ/	5.22(1.52)	5.59(1.14)	3.63(1.26)	3.38(1.13)
		buro	/bʌroʊ/	5.33(1.08)	5.64(1.00)	3.25(1.39)	3.69(1.15)

	m-t-o-u	motu	/mɒtu:/	5.06(1.11)	5.23(1.15)	3.13(1.31)	3.00(1.05)
		tomu	/tɒmu:/	5.61(1.09)	5.27(.77)	3.19(.83)	3.16(1.14)
		muto	/mʌtoʊ/	5.11(1.02)	4.68(1.25)	2.44(1.09)	3.38(1.04)
		tumo	/tʌmoʊ/	5.44(1.38)	5.55(1.1)	3.50(1.15)	3.38(1.36)
	m-b-o-u	mobu	/mɒbu:/	5.78(1.17)	5.77(1.23)	3.88(1.09)	3.88(.87)
		bomu	/bɒmu:/	6.11(1.13)	6.27(.63)	3.31(1.35)	3.81(1.03)
		mubo	/mʌboʊ/	5.84(1.21)	6.23(.69)	4.19(.75)	3.63(1.01)
		bumo	/bʌmoʊ/	6.00(1.03)	5.82(.91)	3.81(1.17)	4.19(1.03)
Neutral	r-t-o-i	roti	/rɒti:/	4.22(1.06)	3.41(1.33)	2.38(1.45)	2.28(1.11)
		tori	/tɒri:/	3.94(1.39)	3.05(1.09)	2.44(1.31)	2.25(1.02)
		rito	/rɪtoʊ/	3.15(1.46)	2.64(.9)	2.25(1.00)	2.19(1.06)
		tiro	/tɪroʊ/	3.39(1.2)	3.18(1.22)	3.00(1.32)	2.41(1.01)
	r-b-o-i	robi	/rɒbi:/	4.33(1.33)	4.32(1.17)	2.75(1.00)	2.97(1.03)
		bori	/bɒri:/	4.67(1.08)	4.45(1.22)	2.94(1.12)	2.97(1.00)
		ribo	/rɪboʊ/	4.28(1.18)	4.27(1.12)	3.06(1.00)	2.91(1.15)
		biro	/bɪroʊ/	4.33(1.46)	3.82(1.33)	3.13(1.15)	2.84(1.30)

m-t-o-i	moti	/mɒti:/	3.83(1.1)	4.05(.95)	2.38(1.02)	2.88(1.41)
	tomi	/tɒmi:/	4.78(1.11)	4.18(1.26)	3.63(.96)	2.81(1.00)
	mito	/mɪtoʊ/	4.06(1.3)	3.77(1.27)	2.19(1.05)	2.66(.87)
	timo	/tɪmoʊ/	4.21(1.03)	3.86(1.25)	2.44(1.41)	2.38(1.16)
m-b-o-i	mobi	/mɒbi:/	5.44(.7)	5.09(.97)	3.38(1.20)	3.66(.90)
	bomi	/bɒmi:/	5.06(1.00)	5.09(1.41)	3.06(1.34)	3.75(1.91)
	mibo	/mɪboʊ/	5.11(1.45)	5.00(.87)	3.38(1.02)	3.44(1.22)
	bimo	/bɪmoʊ/	5.44(.92)	4.68(1.25)	3.25(.68)	3.50(1.11)
r-t-o-e	reto	/retoʊ/	4.78(1.22)	4.18(1.26)	2.13(1.15)	2.38(1.07)
	tero	/teroʊ/	4.83(.92)	4.55(1.18)	3.19(1.05)	2.47(.95)
r-b-o-e	rebo	/reboʊ/	5.89(.83)	5.18(1.18)	3.00(.89)	3.19(1.20)
	bero	/beroʊ/	5.72(1.36)	5.41(1.00)	3.19(.83)	3.28(1.14)
m-t-o-e	meto	/metoʊ/	5.28(.89)	4.82(1.33)	2.50(1.15)	2.69(1.09)
	temo	/temoʊ/	5.5(.79)	5.05(.9)	2.88(1.20)	2.72(1.14)
m-b-o-e	mebo	/meboʊ/	6.00(1.03)	5.59(.91)	3.75(1.18)	3.94(.84)
	bemo	/bemoʊ/	6.00(1.37)	5.55(1.3)	3.75(.93)	3.47(1.02)

r-t-u-i	ruti	/rʌti:/	3.17(1.34)	3.00(1.11)	2.31(1.30)	1.75(.95)
	turi	/tʌri:/	2.94(1.11)	2.77(1.23)	2.50(1.03)	2.28(1.11)
	ritu	/rɪtu:/	2.58(1.61)	2.77(.92)	1.94(.68)	2.00(.95)
	tiru	/tɪru:/	3.00(1.03)	2.91(1.15)	2.94(1.00)	2.47(.92)
r-b-u-i	rubi	/rʌbi:/	4.06(1.43)	3.86(1.32)	3.50(1.46)	2.66(1.15)
	buri	/bʌri:/	3.79(1.18)	3.77(1.34)	3.31(1.14)	3.16(.92)
	ribu	/rɪbu:/	3.61(1.09)	3.64(1.14)	3.44(.89)	2.75(1.08)
	biru	/bɪru:/	3.28(1.07)	3.5(1.1)	2.63(1.15)	3.28(1.20)
m-t-u-i	mitu	/mɪtu:/	3.28(1.23)	3.09(.81)	2.00(.97)	2.63(1.16)
	timu	/tɪmu:/	4.00(1.11)	3.32(1.09)	3.06(.93)	3.03(1.26)
m-b-u-i	mubi	/mʌbi:/	4.56(1.25)	4.36(1.18)	3.69(.95)	3.59(1.07)
	bumi	/bʌmi:/	5.11(1.18)	4.32(1.32)	3.88(1.15)	3.75(1.19)
	mibu	/mɪbu:/	3.72(1.41)	4.36(1.26)	3.25(1.06)	3.72(.89)
	bimu	/bɪmu:/	4.44(1.34)	4.36(1.22)	3.31(1.01)	3.28(1.25)
r-t-u-e	retu	/retu:/	4.00(1.37)	3.86(1.04)	2.88(1.31)	2.63(1.10)
	teru	/teru:/	4.06(1.51)	4.23(1.02)	2.94(1.18)	2.56(1.05)

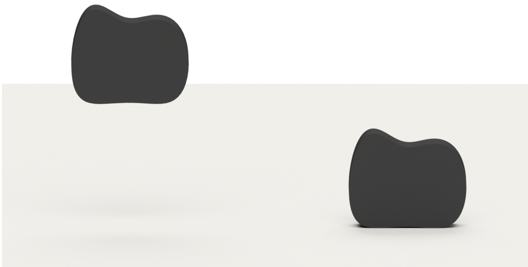
	r-b-u-e	rebu	/rebu:/	4.83(1.42)	4.86(1.32)	3.44(.89)	2.94(.98)
		beru	/beru:/	4.84(1.01)	4.77(.97)	3.31(.95)	3.19(1.33)
	m-t-u-e	metu	/metu:/	4.94(1.06)	4.55(1.14)	2.75(1.13)	2.81(1.00)
		temu	/temu:/	4.84(1.26)	4.73(1.28)	2.94(1.00)	3.13(.98)
	m-b-u-e	mebu	/mebu:/	5.58(.9)	5.18(1.05)	3.63(1.02)	3.53(1.05)
		bemu	/bemu:/	5.39(1.14)	5.18(1.1)	3.50(.97)	3.72(.81)
Pointy	t-b-o-e	tebo	/teboʊ/	5.28(1.28)	4.77(1.23)	2.56(1.26)	3.06(1.08)
		beto	/betoʊ/	4.83(1.34)	4.45(1.1)	2.56(1.21)	2.97(1.20)
	t-b-i-u	tubi	/tʌbi:/	4.00(1.28)	3.73(1.16)	3.81(1.28)	3.38(1.24)
		buti	/bʌti:/	3.21(1.13)	3.5(1.3)	2.63(1.31)	2.75(1.16)
		tibu	/tɪbu:/	3.53(1.35)	3.27(.94)	3.00(1.10)	2.63(1.10)
		bitu	/bɪtu:/	3.39(1.29)	2.68(.89)	2.13(1.15)	2.56(1.01)
	t-b-e-u	tebu	/tebu:/	4.67(1.08)	4.55(1.06)	2.69(.95)	3.16(1.11)
		betu	/betu:/	4.28(1.13)	4.55(1.22)	2.50(1.03)	2.88(.83)
	m-t-i-e	meti	/meti:/	3.94(1.3)	3.55(.91)	2.06(1.39)	2.50(1.02)
		temi	/temi:/	3.72(1.41)	3.55(1.06)	2.25(1.00)	2.72(1.35)

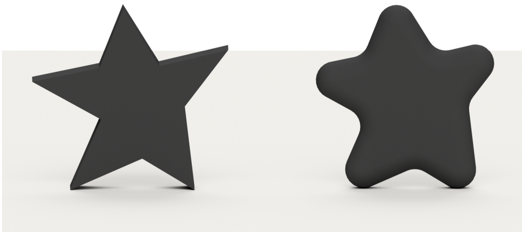
	m-b-i-e	mebi	/mebi:/	4.84(1.17)	3.86(1.36)	2.75(1.00)	2.81(1.06)
		bemi	/bemi:/	4.5(1.34)	4.32(1.36)	3.13(.96)	3.44(1.01)
Extreme Pointy	t-b-i-e	tebi	/tebi:/	3.78(1.4)	3.45(1.06)	2.78(1.34)	2.75(1.11)
		beti	/beti:/	3.39(.43)	3.00(1.02)	2.25(1.16)	2.23(1.14)

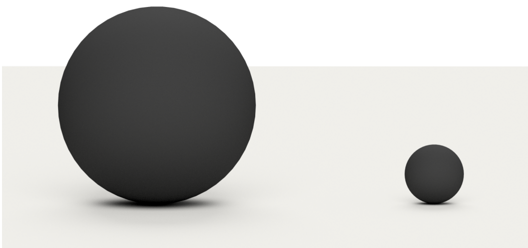
Appendix B

Questionnaire administered in Experiment 5 of Chapter 3.

What makes these two objects different?







In each of the below statements, you are told some limited information about objects you have never seen before. Based on this information, it is your task to answer questions about these objects.

Please CIRCLE your answer on each scale.

1. This object makes a high-pitched sound. What shape do you think it is?

VERY ROUNDED	ROUNDED	SLIGHTLY ROUNDED	NEITHER ROUNDED NOR POINTY	SLIGHTLY POINTY	POINTY	VERY POINTY

2. This object is very bright in colour. How thick do you think it is?

VERY THICK	THICK	SLIGHTLY THICK	NEITHER THICK NOR THIN	SLIGHTLY THIN	THIN	VERY THIN

3. This object is small in size. How bright do you think it is?

VERY DARK	DARK	SLIGHTLY DARK	NEITHER DARK NOR BRIGHT	SLIGHTLY BRIGHT	BRIGHT	VERY BRIGHT

4. This object makes a low-pitched sound. How big do you think it is?

VERY SMALL	SMALL	SLIGHTLY SMALL	NEITHER SMALL NOR BIG	SLIGHTLY BIG	BIG	VERY BIG

5. This object is very heavy. How bright in colour do you think it is?

VERY DARK	DARK	SLIGHTLY DARK	NEITHER DARK NOR BRIGHT	SLIGHTLY BRIGHT	BRIGHT	VERY BRIGHT

6. This object is dark in colour. What pitch of sound does it make?

VERY HIGH-PITCHED	HIGH-PITCHED	SLIGHTLY HIGH-PITCHED	NEITHER HIGH-PITCHED NOR LOW-PITCHED	SLIGHTLY LOW-PITCHED	LOW-PITCHED	VERY LOW-PITCHED