Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 8, 2270–2299
https://doi.org/10.1037/xge0001393

Crossmodal Transfer of Perceptual Learning: Evidence From the Recognition of First and Second Language Characters in Young Children

A. K. M. Rezaul Karim[1], Tanzia Mizan[1], and Samsad Afrin Himi[2]
[1] Department of Psychology, University of Dhaka
[2] Department of Psychology, Jagannath University

Crossmodal transfer of learning is a neurocognitive process whereby a learner's experience and knowledge acquired through one sensory mode enable him/her to perform a similar task using a different sensory mode. This study examined the transfer of (mostly) visually acquired knowledge of first- and second-language characters to the tactile modality typically not used in that acquisition process. Two experiments were conducted, one to assess letter recognition skills and one to assess digit recognition skills in both Bangla and English, in 30 sighted young children who had mastered those characters through sensory learning in natural settings. Results unequivocally demonstrated that children were able to recognize/classify the first and second language letters or digits presented not only to the (trained) visual modality but to the (untrained) tactile modality as well, and as expected, with greater recognition accuracy and shorter recognition time in the former than the latter modality. Their character recognition performance was found to be significantly influenced not by language but by character type, with digits being more accurately and more speedily recognized than letters. Moreover, language–task modality interaction was found to mediate letter recognition accuracy, digit recognition accuracy, and digit recognition time, whereas character type–task modality interaction was found to significantly mediate character recognition time only. The ecological and theoretical significance of these findings is discussed.

Public Significance Statement
Crossmodal learning is crucial to adaptive behavior in altered sensory environments, such as when somebody loses vision or hearing. Crossmodal learning is usually thought to operate through interactions of two or more input-receiving sensory modalities. For example, learning to read letters or digits during childhood involves interactions of both visual and auditory modalities, and learning to write those characters involves interactions between visual, auditory, proprioceptive, and motor systems. This study introduces a novel form of crossmodal learning that implies that learning to read a letter or digit in one sensory modality, such as visual modality, leads to recognition of the same letter or digit presented in a second sensory modality, such as tactile modality, that was not exposed to that character during the course of learning. The findings presented here are ecologically significant as crossmodal transfer of the complex linguistic shapes can be shaped by semantic quality, orthographic structure, and other characteristic features of linguistic characters.

Keywords: perceptual learning, crossmodal transfer, recognition, letters and digits, orthography

The last few years have seen a marked growth of interest in the topic of crossmodal transfer of sensory experience or learning. The term crossmodal (aka cross-sensory) refers to interactions between two or more sensory modalities. The term crossmodal learning refers to a form of learning that involves information from multiple sensory modalities as they interact (Skocaj et al., 2012; Zhang et al., 2022). From this definition, it appears that crossmodal interactions exclusively occur between two or more input-receiving sensory modalities. However, this is not always the case. In reality, the interactions can occur not only

This article was published Online First May 1, 2023.
A. K. M. Rezaul Karim https://orcid.org/0000-0002-8762-7126
We do not have any conflict of interest to disclose.
We did not present the data and ideas appearing in this manuscript at any conference, symposium, or academic meeting. All data and study analysis codes (but not the tactile versions of linguistic characters) are available in the repository of the OSF and can be accessed via https://osf.io/zn7yu/.
A. K. M. Rezaul Karim served as lead for conceptualization, data curation, formal analysis, methodology, project administration, supervision, writing–original draft, writing–review and editing, and writing final draft; contributed equally to resources; and served in a supporting role for investigation. Tanzia Mizan served as lead for investigation and resources and served in a supporting role for writing–original draft. Samsad Afrin Himi served in a supporting role for formal analysis, methodology, and writing–review and editing. A.K.M. Rezaul Karim and Samsad Afrin Himi contributed to software equally.
Correspondence concerning this article should be addressed to A. K. M. Rezaul Karim, Department of Psychology, University of Dhaka, Dhaka 1000, Bangladesh. Email: karim.akmr.monscho06@gmail.com; akm.rezaul.karim@du.ac.bd

between two input-receiving sensory modalities, but between an input-receiving sensory modality and a new sensory modality as well. The latter form of crossmodal interactions enables us to use a sensory modality to perform a perceptual task learned through a different sensory modality (Gottfried et al., 1977; Meltzoff & Borton, 1979; Picard, 2007; Rose & Orlian, 1991; Rubinstein & Gruenberg, 1971; Spence et al., 2009). For example, we can recognize by touch an object that we previously explored by vision only, and similarly, we can recognize visually an object that we previously felt by touch only. These capabilities allow for adaptive behaviors in altered sensory environments, such as when somebody loses vision or hearing. Though some earliest researchers used the term crossmodal transfer interchangeably with the term crossmodal integration (e.g., J. P. Jones, 1972; Silverston & Deichmann, 1975), there is a sharp distinction between the two concepts. Crossmodal integration is a process of combining and integrating information from different sensory modalities (Alais & Burr, 2004; Ernst & Banks, 2002; Ghazanfar & Schroeder, 2006; Raij et al., 2000; Stein & Stanford, 2008), whereas crossmodal transfer is a process of conveying information from the input-receiving sensory modality to a new sensory modality (Wismeijer et al., 2012). Both processes inevitably involve interactions at the neural level; however, unlike the concept of crossmodal transfer, the concept of crossmodal integration (aka multisensory integration) is generally used to refer to a process of sensory interactions when stimuli are coincidentally presented in multiple sensory modalities (Picard, 2007; Stein & Meredith, 1993). Thus, crossmodal integration might play a crucial role in the development of the crossmodal association, a necessary step to convey sensory information from one sense mode to another (see Birch & Belmont, 1965; Birch & Lefford, 1963; Meltzoff & Borton, 1979). However, because there might be intrinsic connections between sensory modalities (Bresciani et al., 2006; Ettlinger & Wilson, 1990; Held et al., 2011; Ludwig et al., 2011; Shepherd, 2012), crossmodal transfer does not necessarily require the benefit of simultaneous stimulation in multiple sensory modalities (Meltzoff & Borton, 1979). This suggests that crossmodal association may develop not only between trained sensory modalities but between a trained sensory modality and an untrained sensory modality as well (see Karim et al., 2021a). For example, when children learn to read linguistic characters, such as letters or digits, they typically do this job through vision and audition (as auditory dictation is needed about how the characters pronounce), and when they learn to write those symbols they use motor programs and action plans in addition to vision and/or audition (see Fridland, 2021; Levy-Tzedek et al., 2012; Prinz, 1997). However, the association may develop not only between the used (experienced) sensory modalities but between the used and unused modalities as well, such as between vision and touch. The former form of crossmodal association has been studied and well documented in a plethora of past studies (for a comprehensive review, see Karim et al., 2021a). The present study intends to offer new insight into crossmodality by examining the latter form of crossmodal association to which little attention has been paid thus far.

A review of the relevant literature shows that crossmodal transfer has been investigated in plenty of studies on recognition memory, perception, and categorical (or discrimination) learning in both normal adults and children. Memory studies in normal adults showed evidence of robust crossmodal priming in visual and tactile modalities when using nonverbal (e.g., Easton, Greene, & Srinivas, 1997; Reales & Ballesteros, 1999) and verbal (e.g., Easton, Srinivas, & Greene, 1997) materials. In the nonverbal domain, one study assessed implicit and explicit memory performance within and between vision and haptics, using both two-dimensional (2D) and three-dimensional (3D) stimuli (Easton, Greene, & Srinivas, 1997). This study revealed robust crossmodal priming in implicit memory for both 2D patterns and 3D objects and modality specificity in explicit memory for 3D objects only. A second study evaluated implicit memory for real objects by a speeded naming task in young adults and showed equivalent perceptual facilitation for visually and haptically studied objects in both intramodal and crossmodal conditions (Reales & Ballesteros, 1999). This study found a total transfer between sensory modalities when participants performed a physical or shallow encoding task and a larger priming effect when participants were presented with line drawings than when presented with visual or haptic real objects. In the verbal domain, another study assessed implicit and explicit memory performance for (English) "letters forming a word" within and between vision and haptics in young adults and revealed no effect of modality change on implicit or explicit memory performance (Easton, Srinivas, & Greene, 1997). The study suggested that at an early stage of processing, the letters and words might be coded as geometric shapes before they are lexically recognized. In an earlier study, this proposition has been suggested for the tactile encoding of letters: letters presented tactilely are first analyzed in spatial (geometric) codes and then translated into linguistic codes (Witelson, 1974). This further suggests that there is no direct link between input and linguistic analysis in the tactile modality, and the translation into linguistic codes in a later stage is only possible if there is an association between the nonlinguistic tactile modality and a linguistic modality like visual or auditory modality. Interestingly, there is evidence that in cases of blindness: cortical territories that are normally involved in visual information processing can also be activated by verbal stimuli encoded through touch in congenitally blind individuals, which suggests that there are shared neural machineries between visual and tactile letter perception (e.g., Amedi et al., 2003).

From a developmental perspective, studies in normal children have shown evidence for an innate nature of crossmodality, and how this capacity further develops with age (see Karim et al., 2021a). In support of the innate nature of crossmodality, some prior studies have shown that transfer of shape from (manual or oral) touch to vision occurs in human newborns (Streri & Gentaz, 2004), and in 1-month-old (Meltzoff & Borton, 1979) and 2-month-old (Streri, 1987) infants. Inconsistently, other studies failed to demonstrate touch-to-vision transfer of shape during the first 1 month of life (Maurer et al., 1999; Pêcheux et al., 1988), and vision-to-touch transfer of shape at 2 months of age (Streri, 1987). However, a more recent study showed that vision-to-touch transfer of shape is possible even at 2 months of age under certain conditions (Streri & Molina, 1993). Despite some discrepancies between the findings for infants in the first 2 months of age, touch-to-vision transfer of shape was consistently observed in older infants, such as in 6-month-old (Rose et al., 1981b; Ruff & Kohler, 1978) and 1-year-old (Gottfried et al., 1977; Rose et al., 1981a, 1983; Rose & Orlian, 1991) infants. Studies in 1-year-old infants further demonstrated that overall intramodal transfer performance was superior to crossmodal transfer performance (Rose et al., 1981a, 1983; Rose & Orlian, 1991). Specifically, tactile intramodal recognition of a shape was superior to visual-tactile transfer (Rose et al., 1981a;

Rose & Orlian, 1991), and visual intramodal recognition of a shape was superior to tactile–visual transfer (Rose et al., 1983; Rose & Orlian, 1991). However, there is a common trend in intramodal and crossmodal ability development. Research has shown that both visual and tactile processing speeds increase between 5 and 9 years of age (Enns & Girgus, 1985; Hatwell et al., 1990). In regard to crossmodal ability development, an early study examined recognition capacity in children of 5- to 11-year-old using paired comparison tasks with geometric form (Birch & Lefford, 1963). This study demonstrated that while individual variation in crossmodal function does exist, the ability to make various crossmodal (visual-haptic, visual-kinesthetic, haptic–kinesthetic) judgments improves with age, with a sufficient improvement occurring by the age of 5. Using crossmodal shape matching tasks, Stoltz-Loike & Bornstein (1987) observed that children performed better on vision-to-touch than touch-to-vision transfer task, and that their crossmodal performance significantly improved between the ages of 5 and 7 years for both kinds of transfer tasks. Highly consistently, a study on the crossmodal transfer of texture in 5- and 8-year-old children found that transfer performance from vision to touch improved between 5 and 8 years of age, whereas transfer performance from touch to vision did not vary with age (Picard, 2007). Taking these findings together, it can be concluded that crossmodal transfer can occur without the benefit of months of experience in simultaneous visual–tactual exploration (e.g., Meltzoff & Borton, 1979), and that such experiences with increasing age can potentially foster the performance. On the contrary, using visual and haptic spatial information one study has shown that there is a lack of multisensory/crossmodal integration before 8 years of age (Gori et al., 2008). This indicates that crossmodal transfer and crossmodal integration might be two independent processes. One principle of capacity development in children is that simple abilities tend to develop prior to more complex ones. Thus, crossmodal transfer might be easier and operate earlier than crossmodal integration. This suggests that young children who have just acquired letter or digit recognition skills mostly through vision should be able to recognize those symbols by touch alone. The present study is designed to test this possibility in 5- to 6-year-old children who have mastered those symbols, in both the first and second language, in Bangladesh.

Studies on categorical or discrimination learning have shown that learning that operates through the use of only one sensory modality can transfer to a new sensory modality not used in that learning. For example, a person trained to categorize a set of objects visually is able to categorize novel objects from the same categories when the objects are grasped but not seen, and vice versa (Wallraven et al., 2014; Yildirim & Jacobs, 2013). A person trained to recognize or categorize sequences of spatial locations auditorily is able to recognize or categorize the same sequences visually and vice versa (Yildirim & Jacobs, 2015). However, another study has shown that the learned improvements in spatial task transfer from the visual to the auditory modality, but not vice versa (McGovern et al., 2016). Conversely, the same study has shown that the benefits derived from training on temporal task transfer from the auditory to the visual modality, but not the other way round. These findings together suggest that the transfer of unisensory learning can be robust across sensory modalities, but occurs in a task-specific fashion. On the contrary, learning that occurs through the use of two sensory modalities transfers unidirectionally rather than bidirectionally. For example, one study has shown that audiovisual learning of temporal order transfers to vision but not to audition (Alais et al., 2010). Conversely, a few studies have shown that audiovisual training with speech stimuli can promote auditory-only perceptual learning in normal-hearing adults (Bernstein et al., 2013, 2014; Eberhardt et al., 2014; Pilling & Thomas, 2011). Thus, it appears that if two sensory modalities are used to acquire something, the learned improvements transfer to one of the sensory modalities used in learning but not to the other one. All these findings inform theoretical accounts of crossmodal learning, particularly regarding the extent to which a sensory modality guides sensory discrimination and perceptual improvement in another sensory modality. Despite this advancement in the understanding of crossmodal transfer, there is a paucity of systematic research investigating the transfer of linguistic symbol learning, such as learning of letters and digits that humans use in verbal (written or spoken) messages while communicating with each other. Second, prior studies investigated the transfer of unisensory learning to a new sensory modality, and the transfer of multimodal learning to a new sensory modality still remains unexplored. Third, prior studies generally trained adult individuals and investigated crossmodal transfer of immediately preceding sensory experience or knowledge to a new or used sensory modality (see above), but the transfer of learning has rarely been examined during early childhood, the crucial time for mastering and transferring letter or digit recognition skills across sensory modalities in natural settings. Fourth, prior studies did not give any attention to the crossmodal transfer of linguistic knowledge in individuals who learn two languages concurrently. The first language characters (e.g., Bangla letters and digits) and the second language characters (e.g., English letters and digits) do largely vary in shape and pronunciation complexity. It can be assumed that these differences might influence not only sensory performance but cross-sensory performance as well. However, the current literature does not tell anything about such a linguistic impact on sensory or cross-sensory knowledge development. Thus, the crossmodal nature of acquired knowledge remains poorly understood. Therefore, the present study intends to understand crossmodal transfer of knowledge by examining post-learning sensory performance in young children who cultivate two languages simultaneously. To this end, linguistic character recognition skills were assessed in normal, sighted school children who, while in home and school settings, have multimodally mastered letter and digit recognition skills in both Bangla (first language) and English (second language).

It follows from the above discussion that the main purpose of this study is to see how multimodally learned letters and digits are recognized by touch alone. Letter or digit recognition skills are acquired through perceptual learning and categorization. In principle, perceptual categories form through sensory convergence by using all relevant information about a concept (K. I. Taylor et al., 2006). For example, learning a letter or digit depends on multisensory information, including visual information about its shape as well as auditory information about how it pronounces (Fraga González et al., 2017; Karipidis et al., 2018; Xu et al., 2018, 2020). This example indicates that the acquisition of a letter or digit depends on successful integration between visual and auditory information (for a review, see Blomert & Froyen, 2010; Raij et al., 2000). As mentioned before, letters or digits are multimodally experienced also during learning to write as it involves motor programs and action plans in addition to vision and/or audition (see Fridland, 2021; Levy-Tzedek et al., 2012; Prinz, 1997). However, one outstanding question that still remains to answer is: does this kind of acquired knowledge transfer to a new sensory modality, such as tactile modality that is typically not used to encode linguistic inputs while learning? The present study aims to answer this question by examining whether children's performance in visual and tactile modalities is greater than the baseline performance which is assumed to be null before linguistic character learning took place. This will reveal not only the extent of knowledge transfer from vision to touch but the extent of performance in visual modality as well. Letters and digits are mainly visual stimuli, and young children are also learning to read which primarily proceeds visually despite having some multimodal experiences of those shapes. Therefore, in addition to tactile transfer, it is important to understand how well they have visually mastered the letters and digits so far. This will further meet a second purpose, by allowing a comparison of recognition performance between vision and touch, the two sensory modalities that are inherently capable of representing reverse versions of the same stimuli or objects (see Amedi et al., 2007; Auvray et al., 2007; Kim & Zatorre, 2008, 2010). Letters and digits are spatially complex shapes, and the capacity to recognize these shapes depends on macrospatial analysis for which vision is more suited than touch (see Bushnell & Boudreau, 1991; Klatzky et al., 1989; Lederman et al., 1986; Mahar et al., 1994; Tokita & Ishiguchi, 2016). Moreover, because vision is directly used to encode linguistic inputs this sensory mode is at advantage in processing linguistic symbols as compared with touch. *Taken together, these two ideas lead to the formulation of a hypothesis that children can recognize a linguistic character (letter or digit) more efficiently in the visual than in the tactile modality.*

A third crucial purpose of this study is to compare and contrast Bangla and English character recognition skills irrespective of character type (digit, letter) in both vision and touch. English is a second language in Bangladesh and is taught concurrently with Bangla, the first language, from the very beginning of schooling, but is not typically used in everyday conversation. Because letters and digits are culturally dependent symbols the acquisition of first- and second-language letters and digits operates through different learning strategies. The first-language letters and digits are usually acquired through implicit rather than explicit learning (see Choo et al., 2012; Ellis, 2016), and are therefore likely to be encoded, retained, and recalled more accurately and more fluently than the second language letters and digits. However, the spatial complexity of those linguistic symbols appearing from their orthographic features might play a crucial role in processing efficiency. An inspection of the first and second language symbols in Bangladesh indicates that Bangla letters and digits typically have more critical orthographic features, such as complex strokes, lines, angles, and curves, which may put them at a disadvantage in processing, identifying, naming, or recalling as compared with English letters and digits that have much simpler and parsimonious orthographic features (see Figure 1 for examples). Prior studies on Arabic letter naming and Hindi or Urdu word reading have shown that orthographic differences and visual complexity of language might influence letter naming and word reading performance (Asaad & Eviatar, 2013; Rao et al., 2011). In particular, this influence was found to supersede the influence of language nativeness in the study that compared Hindi and Urdu word reading capacities in biliterate readers (Rao et al., 2011). *These findings together lead to the formulation of a second hypothesis that irrespective of character type (letter or digit), children would be better at recognizing an English character than a Bangla character.*

A fourth crucial purpose of this study is to compare and contrast letter and digit recognition skills irrespective of the language in both visual and tactile modalities. Within each linguistic system, digits have similar visual characteristics to letters; however, they are different from letters in potentially significant ways. Letters are elements of an alphabetic writing system, whereas digits are logographs (Besner & Coltheart, 1979; Ingles & Eskes, 2008; McCloskey & Schubert, 2014). Thus, digits represent entire words and convey conceptual meanings, whereas letters typically convey meanings only when combined meaningfully (except single-letter words; Deloche & Seron, 1987; Holender & Peereman, 1987; McCloskey & Schubert, 2014). Additionally, irrespective of language, letters appear to have more complex shapes as compared with digits in terms of orthographic features (see Figures 1 and 2). All these characteristic differences together likely put a letter at a disadvantage in identifying, naming, or recalling as compared with a digit. *This leads to the formulation of a third hypothesis that irrespective of language, children would be better at recognizing a digit than a letter.*

Perceptual learning and categorization of linguistic characters depend on all relevant information available to the learner (K. I. Taylor et al., 2006), and despite having distinct properties, many of those characters are similar to each other in some fashion, such as in terms of shape and pronunciation. Therefore, one more pertinent question is: how is recognition/classification performance for individual letters or digits influenced by those characters each other? The present study aims to answer this final interesting question by generating "confusion matrices" for both first- and second-language letters or digits, showing how they are mistaken for each other in each sensory modality, and whether the recognition/classification performance for individual letters or digits is above the chance level. This will further give insight into the nature of the errors children make in each sensory modality, and the patterns of linguistic character learning and crossmodal transfer of that learning in young children.

To address the aforementioned questions or test the hypotheses this study is designed to assess letter or digit recognition (or classification) capacity in 5 to 6-year-old children who have multimodally mastered those symbols, in both first and second language, in such natural settings as home and school. A group of children of this age range would be assessed for recognition/classification skills using Bangla and English characters (letters and digits) in both vision and touch without further giving them any artificial learning experience during experimentation. The investigation of linguistic character recognition skills as a function of sensory modality, language, and character type using such a unique approach can provide new insight into the development of crossmodal capacity in humans.

Materials and Method

Transparency and Openness

This study was designed following a rigorous methodology in accordance with the Declaration of Helsinki guidelines (World Medical Association, 2013), and in accordance with the legal requirements of Bangladesh and the institutional protocols for research at the Department of Psychology, University of Dhaka. To adopt transparency and avoid any potential vagueness the study design, variables, sample, treatment, stimuli, materials, data acquisition procedure, and analysis software, packages, and techniques are explicitly described below so that the results reported here can be generalized or reproduced on a new sample not only from Dhaka/Bangladesh but from other regions/cultures as well. To maximize transparency all prior studies relevant to this work are appropriately cited in the text and listed in the Reference section. Although the present study and its analysis plan were not preregistered in a repository, they were at least designed and specified prior to implementation. However, the reviewer-suggested some additional analyses were also done to improve the quality of this work. The data sets and study analysis codes (but not the tactile versions of linguistic characters) are available at https://osf.io/zn7yu/ on the Open Science Framework (OSF). These data can be reanalyzed and synthesized in future replication studies that likely reproduce the present findings.

Study Design

Two experiments were conducted, one with letters and one with digits. Each experiment was conducted using a within-subjects experimental design that involved two factors, namely (sensory) task modality and language, each with two levels. Task modality comprised the levels of visual and tactile, and language comprised the levels of Bangla and English.

Experiment 1

Participants

A total of 30 5- to 6-year-old sighted school children, 12 boys and 18 girls, 24 Muslims and 6 Hindus, with normal or corrected to normal vision, who were naïve to the study purpose, were recruited. In line with the study purpose, children of this small age range were recruited because they had just mastered the letters and digits in Bangla and English through multimodal training in such natural settings as home and school (Foulin, 2005; Hayashi et al., 2013), but did not have any prior experience of playing with 3D letter blocks. They were children of Grade 1 from three different schools selected pseudorandomly from the city of Dhaka. Children who attended a class on particular dates and showed interest, and met these criteria were all included in the study. Thus, 13 of the recruited children were from Begum Sufia Bidya Niketon, 9 from Deepshikha Kindergarten, and 8 from Holy Heart International School.

The aforementioned sample size was justified based on a review of the relevant literature. Most prior behavioral studies on letter and digit/number processing used various sample sizes, ranging from 3 to 20 individuals (Fischler, 1975; Fiset et al., 2009; Hayashi et al., 2013; James et al., 2005; Legge et al., 2001; Madec et al., 2016; Myers & Szücs, 2015; Perea et al., 2015; Rączy, Czarnecka, Zaremba, et al., 2020; Santee & Egeth, 1980, 1982b; Staller & Lappin, 1979; Wood, 1977), with the exception that only one study used a large sample, comprising 118 individuals (Mueller & Weidemannb, 2012). Prior neuroimaging studies on similar topics used sample sizes, ranging from 7 to 40 individuals (Cantlon et al., 2006; Carreiras et al., 2015; Lochya et al., 2018; Merkley et al., 2019; Park et al., 2012; Peters et al., 2015; Raij et al., 2000; Rączy, Czarnecka, Paplińska, et al., 2020; Santens et al., 2010; Shum et al., 2013; Tang et al., 2006; van Atteveldt et al., 2004).

Training of Participants

In this study, training refers to a multimodal activity (e.g., audiovisual, visuomotor, and/or audiomotor) through which the selected children learned to read and write both first (Bangla) and second (English) language characters (letters, digits) from their parents and teachers in home and school settings. Thus, it was the training in natural settings, and no additional training was introduced in the experimental setting. In Bangladesh, children usually start getting coached on both first and second language characters simultaneously at the same age. However, learning of the spoken (first) language starts through audiomotor activity from birth. This might enhance children's speed of recognition, but not the recognition of the Bangla letters/digits per se. Letter and digit learning starts operating from very early ages in the multisensory environment, and by the age of 5/6 every normal child likely completes learning of both first and second language letters/digits, in addition to mastering some other skills by using those linguistic symbols, in home and school settings. Because primary education is compulsory in Bangladesh a control group formed by children who do not know the letters/digits does not actually exist, and thus it was not realistic to recruit a control group. However, because a within-subjects design was used here, participants served as their own control by providing baseline scores across different conditions, and before receiving multimodal linguistic training from parents and teachers their baseline character recognition scores were assumed to be zero in each condition.

Stimuli and Materials

Ten 3D Bangla consonants and 10 3D English capital letters, all made of plastic, were chosen to test each child's letter recognition capacity in tactile and visual modalities. The modern English alphabet comprises 26 characters, including 5 vowels and 21 consonants, each one having an upper- and lower-case form (aka capital and small letters), whereas the Bangla alphabet comprises 50 basic characters, including 11 vowels and 39 consonants (Pal et al., 2009; Rabby et al., 2018; Rahman et al., 2019; Tabassum et al., 2020). The concept of upper/lower case is absent in the Bangla alphabet script (Pal et al., 2009). The two sets of alphabets were collected from a commercial store, and the characters in each set appeared to be of the same, but unknown, font style. The stimulus set in each language included the first three letters, the last three letters and four letters arbitrarily chosen from the rest of the Bangla consonant series or English capital letter series (Figure 1). This approach to stimulus selection was used in order to control or minimize the potential serial position effect or Von Restorff effect on letter or digit recognition (see Howe et al., 2000; T. C. Jones & Roediger III, 1995; Von Restorff, 1933). All letters were of the same color, the Bangla letters having dimensions of 3.7 cm × 1.5 cm × 0.4 cm to 3.8 cm × 3.1 cm × 0.4 cm, and the English letters having dimensions of 3.8 cm × 2.9 cm × 0.6 cm to 4.5 cm × 3.8 cm × 0.6 cm. A wooden board having dimensions of 28.5 cm × 17.5 cm × 1.6 cm was used to present the letters, and a digital stop watch was used to measure the participants' response time.

Data Acquisition

Data were collected following a standard procedure. First, the second author, a trained experimenter, went to the selected schools and

Figure 1
Stimuli Used in Experiment 1

Note. (a) 3D Bangla consonants, (b) 3D English alphabets.

received informed consent from the respective school authorities (heads) who assigned her a specific, quiet room to conduct the experiment. The experiment was conducted on the selected children individually, one or two a day. Prior to conduction, each child's demographic and tactual experiential information was collected. To this end, the child was asked to tell its name and the grade it was reading in, with its responses recorded in free response boxes. The child was further asked to tell whether it was normally sighted or was using any artificial lens/spectacles, and whether it had any prior experience of playing with 3D letter blocks in Bangla or English. For the latter question, the child was given two options to choose from: (a) yes and (b) no. The other demographic information, such as the child's sex, date of birth, and ethnicity/religion were collected from the school administrative records.

The experiment was conducted in two sensory phases: a tactile phase and a visual phase. Each phase included two conditions, one with Bangla and one with English letters. At the outset of each sensory phase, each child was warmly greeted and briefly explained what to do and how to do (the procedure) in the experiment. Then, a few practice trials were given with letters that were different from those used in the actual experiment in both the linguistic conditions under each sensory phase.

In the actual experimental setting, each child was presented with a specific set of Bangla or English letters in a (previously specified) random order and one at a time on the wooden board, the letters being the same in both sensory phases. The letters were presented facing the child the correct way (from the child's view). Prior to presentation in each phase, each child was informed of whether she/he was going to be presented with the Bangla or English set of letters. Each child was required to perform a naming task—to tell the name of each letter (as soon as possible) tactilely (i.e., by touching with two hands) while blindfolded in the tactile phase, and visually in the visual phase, but with no time limit. In the tactile phase, she/he was allowed to use two hands so as to eliminate the potential effect of handedness/laterality or the hand used in tactile exploration (Birch & Lefford, 1963; Rose, 1984; Rose et al., 1998; Stoycheva et al., 2021; Streri & Gentaz, 2004). In this phase, each child was also allowed to lift the letter from the board but instructed not to rotate or alter its orientation while touching and exploring the letter. This allowed her/him to receive a tactile experience of the whole letter shape by holding it in the correct orientation. Each child's response accuracy for each letter was measured on a correct–incorrect scale, in addition to recording the incorrect response (if any), with the response time measured using a digital stopwatch. The stopwatch was started at the onset of a stimulus and stopped at the onset of a response made by the child.

Here, of note is that each child was first tested in the tactile phase, followed by the visual phase, and this was justified on two grounds. First, there is evidence that transfer of form discrimination, without relevant verbalization, occurs from vision to touch, but not from touch to vision (Blank et al., 1968), and that the 3D object recognition performance does not depend on which sensory modality experiences the objects first (Suzuki & Gyoba, 2008). Second, because children participating in this study already visually well mastered the letters of their first and second language presenting these symbols first visually might enhance (if it really does) the subsequent recognition performance in the tactile phase, but not the other way around. This indicates how counterbalancing between the two sensory phases would be inappropriate and problematic in this particular study. If counterbalanced, the tactile performance of half (50%) of the study participants in the second phase would be influenced by the visual experience they were likely to have in the first phase, but the visual performance of the remaining participants in the second phase would not be influenced by the tactile experience they were likely to have in the first phase. Thus, counterbalancing between the two sensory phases would not be able to neutralize the practice effects if any; it would instead create unequal practice effects, the effects transferable from the visual to the tactile phase but not the other way round. However, this kind of problem was unlikely to occur for the two linguistic conditions within a sensory phase. Therefore, in each sensory phase, the order of the Bangla and English conditions was counterbalanced across children. That is, 15 children were first presented with the set of Bangla letters followed by the set of English letters (one at a time), and the remaining 15 children were first presented with the set of English letters followed by the set of Bangla letters in each sensory phase. Thus by presenting letters tactilely in the first phase and visually in the second phase, and by counterbalancing the order of Bangla and English letters within each sensory phase the potential practice effects were excluded or neutralized.

Experiment 2

Participants

All the children who voluntarily participated in Experiment 1 also participated in this experiment.

Stimuli and Materials

In this experiment, 10 3D Bangla digits (made of wood; plastic digits were unavailable) and 10 3D English digits (made of plastic), each digit set being of the same but unknown font style, were used to assess each child's digit recognition capacity in tactile and visual modalities (Figure 2). All digits were of the same color, the Bangla digits having dimensions of 6.4 cm × 2.9 cm × 0.4 cm to 6.2 cm × 4.4 cm × 0.4 cm, and the English digits having dimensions of 4.4 cm × 1.6 cm × 0.6 cm to 3.9 cm × 3.3 cm × 0.6 cm. The same wooden board and the digital stopwatch that were used in Experiment 1 were also used here to present the digits and measure the response time respectively.

Data Acquisition

Data were collected here following the same procedure as in Experiment 1.

Scoring and Organizing Data

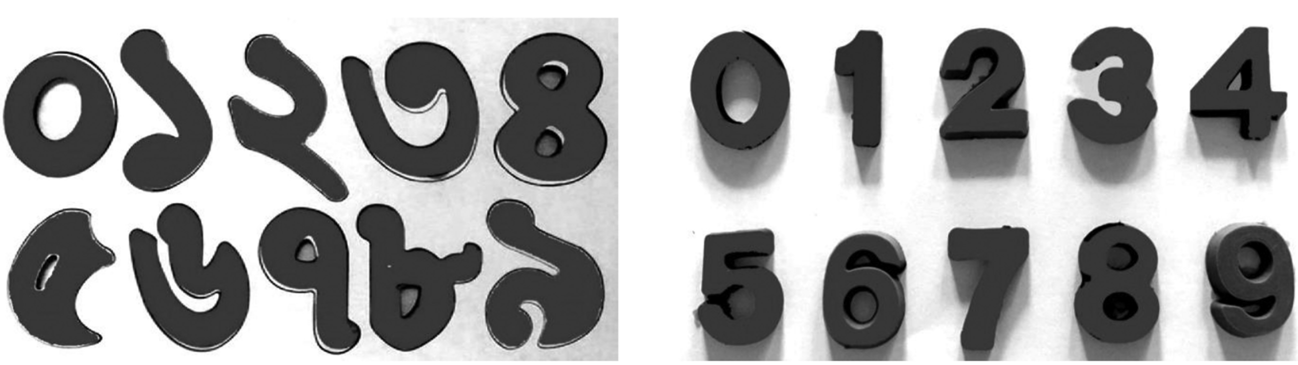
A set of character recognition data was produced for individual participants in each sensory phase of each experiment. To this end, each child's responses were coded following a binary coding system. That is, each correct response was assigned a score of "1" and each incorrect response was assigned a score of "0." It can be noted that if the digit 6 is rotated it is upside down, and we get a 9, and vice versa. Although the child was not allowed to physically rotate any characters she/he might mentally rotate them while responding. So, in the blindfolded tactile phase in Experiment 2, when a child's response to either 6 or 9 was six in a prior trial and nine in a subsequent trial or vice versa, both the responses were considered as correct and assigned a score of "1" each. If someone's responses to 6 and 9 were six and six or nine and nine in both trials the first response was considered as correct and the second response as incorrect, and they were assigned a score of "1" and "0," respectively. Following this scoring, each child's correct response proportion, labeled as "*Recognition Score*" was calculated for a letter or digit by adding all the "1s" and "0s" divided by 10, the number of items, in each linguistic condition of the two sensory phases. Similarly, each child's mean "*Recognition Time*" for a letter or digit was calculated by adding the time (second) taken for all the recognized (correctly answered) letters or digits, divided by the total number of recognized items in each linguistic condition of the two sensory phases. Because there were two levels of task modality and two levels of the linguistic variable each experiment comprised four stimulus (letter or digit) recognition measures and four recognition time measures, with 30 independent cases in each.

Data Trimming and Transformations

Here, data trimming and appropriate transformations of the trimmed data were done to deal with skewed data. For trimming, the character recognition data were fed onto SPSSv26 for windows, and observation(s) falling beyond ± 3 *SDs* from the mean of each measure were replaced by a value equal to the $M \pm 3$ *SDs*. In Experiment 1, one out of 120 letter "*Recognition Scores*" and one out of 120 letter "*Recognition Times*" were found to lie beyond $M \pm 3$ *SDs*. Thus, only 0.83% of data points were trimmed for both letter recognition measures and recognition time measures. In Experiment 2, two out of 120 digit "*Recognition Scores*" and two out of 120 digit "*Recognition Times*" were found to lie beyond $M \pm 3$ *SDs*. Thus, only 1.67% of data points were trimmed for both digit recognition measures and recognition time measures. This replacement technique aimed to preserve each participant's rank ordering, while preventing extreme outliers from unduly influencing parameter estimation (for a similar trimming procedure, see Friedman et al., 2016; Himi et al., 2019).

Because letter or digit recognition data were in proportion and recognition time data were in second (i.e., original continuous data) for the former datasets arcsine transformations and for the latter datasets log transformations were done (see Cohen et al., 2003; Fernández &

Figure 2
Stimuli Used in Experiment 2



(a) (b)

Note. (a) 3D Bangla digits, (b) 3D English digits.

Vadillo, 2020). The mean (*M*), standard deviation (*SD*), skewness, and kurtosis of arcsine "*Recognition Scores*" or log "*Recognition Times*" computed for each measure are presented in Tables A1 and A2 (see Appendix). These tables show that after arcsine or log transformations, the data for each behavioral measure were approximately normally distributed, with a skewness value <3 and a kurtosis value much <10 (see Kline, 2005).

Data Analysis

To meet the first four purposes of the study, various inferential statistical analyses were carried out on normalized recognition datasets for 30 participants (see above). In this stage, initially, a series of *t*-tests were done on arcsine or log-transformed data to see whether there were gender differences in letter/digit "*Recognition Score*" and letter/digit "*Recognition Time.*" No significant gender differences were found in both of these dependent variables. Thus the data were collapsed across genders. Then, arcsines of letter or digit "*Recognition Scores"* obtained in each linguistic condition of the visual and tactile phases were analyzed in a series of one-sample *t*-tests to see whether the participant's letter or digit recognition performance was significantly different from the baseline performance which was assumed to be zero before learning of the linguistic symbols took place. Because there was no control group these analyses were crucial for understanding the extent to which letter or digit recognition skills were mastered by the visual modality and transferred to the tactile modality. Finally, the arcsine "*Recognition Scores*" and log "*Recognition Times*" were analyzed in a series of 2×2 repeated measures analysis of variances (ANOVAs), taking task modality and language as within-subjects factors in each experiment (for similar analyses of proportion data, see Butler et al., 2007; Juola et al., 1974; MacDonald et al., 2002; Mate & Baqués, 2009; Van der Heijden et al., 1992). SPSSv26 for windows directly produces such outputs as effect sizes for ANOVAs but not for *t*-tests. Therefore, the effect size for one sample *t*-test was ascertained in terms of Cohen's *d* (Cohen, 1988) using "powerAnalysis" package (Fan, 2017) with "ES.t.one" function in the open-source statistical software R (R Development Core Team, 2015). The obtained values of Cohen's *d* were then transformed to η_p^2 (partial eta-squared)[1] to be consistent with the units of effect sizes obtained for repeated measures ANOVAs.

Compiling Confusion Matrices

Because children performed a character naming task their responses became classified, and allowed the researchers to compile confusion matrices that met the last purpose of the study. A total of four stimulus/response "confusion matrices" were compiled, capturing the performance of the visual and tactile classifiers on 30 participants' pooled responses (sets of raw data) obtained in Bangla and English letter naming task and Bangla and English digit naming task. A confusion matrix (aka error matrix) presents a table layout of the number or proportion of correct and incorrect predictions for each class by comparing an observation's predicted class and its true class. The confusion matrices were generated here on the basis of the proportion of True Positives, the proportion of True Negatives, the proportion of False Positives, and the proportion of False Negatives. A True Positive is an outcome where the sensory classifier correctly predicts the positive class, and a True Negative is an outcome where the classifier correctly predicts the negative class. A False Positive is an outcome where the sensory classifier incorrectly predicts the positive class, and similarly, a False Negative is an outcome where the sensory classifier incorrectly predicts the negative class. For example, labeling "A" as "A" is a True Positive, labeling "A" as "not A or something else" is a False Negative, labeling a letter which is "not A" as "A" is a False Positive, and labeling a letter which is "not A" as "not A" is a True Negative.

There are three common metrics of classification performance, such as "Recall," "Precision," and "F1-score." "Recall" (aka Sensitivity) is the proportion of relevant cases that were retrieved whereas "Precision" (aka Positive Predictive Value) is the proportion of relevant cases among the retrieved cases. Thus "Recall" tells us how many times the classifier was able to identify a specific category whereas "Precision" tells us how good the classifier is at predicting a specific category. "F1-score" is the harmonic mean of "Recall" and "Precision," and tells us how accurate the classifier is in prediction on a dataset. The three classification metrics were calculated here for each character stimulus by using the following formulas.

$$\text{Recall} = \text{ProportionTruePositives} \, / \, (\text{ProportionTruePositives} + \text{ProportionFalseNegatives})$$

$$\text{Precision} = \text{ProportionTruePositives} / (\text{ProportionTruePositives} + \text{ProportionFalsePositives})$$

$$\text{F1} - \text{score} = 2(\text{Recall} \times \text{Precision}) \, / \, (\text{Recall} + \text{Precision})$$

Statistical Power

Power analysis was carried out to examine whether the current study had adequate power to confirm statistical inferences. Through a review of the relevant literature, two behavioral studies (Myers & Szücs, 2015; Perea et al., 2015) and one neuroimaging study (Carreiras et al., 2015) were identified to report effect sizes. However, the experimental designs and the nature of the manipulating variables used in those works do not correspond to the designs and the nature of the manipulating variables used in the present study. Thus, it was not appropriate to use the effect sizes reported in prior studies to conduct power analyses in the present study. Therefore, the minimum effect size that can be detected with the expected sample size ($n = 30$) was estimated here with $\alpha = 0.05$ and $1 - \beta = 0.80$ (e.g., Hilbert et al., 2017; Himi, 2018; Norman & Thaler, 2021) in each experiment.

A series of a posteriori power analysis were performed with $n = 30$ for one sample *t*-test and repeated measures ANOVAs using two different packages in the open-source statistical software R. First, the minimum detectable effect sizes in terms of Cohen's *d* were computed for one sample *t*-test using "pwr" package (Champely et al., 2020) with "pwr.t.test" function, and the obtained values of Cohen's *d* were transformed to η_p^2 to be consistent with the units of effect sizes for repeated measures ANOVAs. Second, the minimum detectable effect sizes in terms of Cohen's *f* were computed for repeated measures ANOVAs using "WebPower" package (Zhang & Yuan, 2018) with

[1] The partial eta-squared represents the proportion of the variance in a criterion (dependent variable) accounted for by a given predictor, in which the effects of other predictor variables are partialled out.

"wp.rmanova" function. Then, the obtained values of Cohen's *f* were transformed to η_p^2. As described above, a 2 (language: English vs. Bangla) × 2 (task modality: tactile vs. visual) experimental design was used in both Experiments 1 and 2. The parameters used for power analysis in this type of design were: "interaction between within-subject factors," "number of groups = 1," and "number of repeated measures = 4." Secondly, a 2 (character type: letter vs. digit)× 2 (language: English vs. Bangla) × 2 (task modality: tactile vs. visual) experimental design was followed for aggregate data, that is, when data were collapsed across experiments for further analysis. The parameters used for power analysis in this type of design were: "interaction between within-subject factors," "number of groups = 1," and "number of repeated measures = 8."

Results showed that corresponding to $\alpha = 0.05$, $n = 30$, and $1 - \beta = 0.80$, the minimum detectable effect size as computed in terms of Cohen's *d* for one sample *t*-test against the null hypothesis of a sample mean = 0 was 0.529 which is equivalent to $\eta_p^2 = 0.0654$. The minimum detectable effect size as computed for each of the two main effects ("task modality" and "language") and one interaction effect (Task Modality × Language) in ANOVAs with the same value of α, *n*, and 1 – β was $\eta_p^2 = 0.276$ (corresponding Cohen's $f = 0.617$) in both Experiments 1 and 2. Finally, the minimum detectable effect size for each of the three main effects ("character type," "language," and "task modality") and each of the four interaction effects in ANOVAs (in the analysis of aggregate data) computed with the same value of α, *n*, and 1 –β was $\eta_p^2 = 0.331$ (corresponding Cohen's $f = 0.704$). Here, of note is that the minimum detectable effect sizes for the main effects and those for the interaction effects in repeated measures ANOVAs were all equal as these power analyses were done for the within-effect test about the mean differences among measurements.

All data and study analysis codes (but not the tactile versions of linguistic characters) are available in the repository of the OSF, and can be accessed via https://osf.io/zn7yu/.

Results

Character Recognition Performance

The results presented here include outputs of inferential statistical analyses of arcsine or log data, descriptive statistics (e.g., mean, standard error, confidence interval, etc.) calculated from (trimmed) raw data, and bar charts generated on the basis of those descriptive statistics. The inferential statistics obtained from arcsine or log data help understand the strength of the effects of different factors on letter or digit recognition skills, whereas the descriptive statistics and the bar charts obtained thereof from (trimmed) raw data help understand those effects meaningfully in terms of original units of data, such as seconds.

Experiment 1

Performance in Letter Recognition

The average "*Recognition Scores*" and average "*Recognition Times*" (second) calculated over 30 participants' (trimmed) raw data for Bangla or English letters in both visual and tactile modalities are plotted in Figure 3. This figure shows that the average "*Recognition Score*" for Bangla letters was 0.920 ($SE = 0.023$, 95% CI = [0.874–0.966]) in the visual, and 0.533 ($SE = 0.036$, 95% CI = [0.460–0.607]) in the tactile modality (panel a), with the corresponding "*Recognition Times*" being 3.06 s ($SE = 0.259$, 95% CI = [2.530–3.589]) and 26.319 s ($SE = 2.273$, 95% CI =[21.670–30.967]), respectively (panel b). The average "*Recognition Score*" for English letters was 0.977 ($SE = 0.010$, 95% CI = [0.955–0.998]) in the visual, and 0.530 ($SE = 0.039$, 95% CI = [0.450–0.610]) in the tactile modality (panel a), with the corresponding "*Recognition Times*" being 2.568 s ($SE = 0.290$, 95% CI = [1.975–3.161]) and 24.41 s ($SE = 2.691$, 95% CI = [18.905–29.915]), respectively (panel b).

When the trimmed "*Recognition Scores*" were transformed to arcsine (see above) and analyzed in a series of one sample *t*-tests results demonstrated that the arcsine "*Recognition Scores*" in the tactile modality were significantly different from zero, the baseline performance, for both Bangla ($t = 21.243$, $p < .001$, Cohen's $d = 3.945$ equivalent to $\eta_p^2 = 0.796$) and English ($t = 19.370$, $p < .001$, Cohen's $d = 3.597$ equivalent to $\eta_p^2 = 0.7639$) letters. Similarly, the arcsine "*Recognition Scores*" in the visual modality were significantly different from zero, the baseline performance, for both Bangla ($t = 31.167$, $p < .001$, Cohen's $d = 5.788$ equivalent to $\eta_p^2 = 0.8933$) and English ($t = 56.428$, $p < .001$, Cohen's $d = 10.478$ equivalent to $\eta_p^2 = 0.965$) letters.

To see the effects of task modality and language on letter recognition arcsine data were analyzed in a two-way repeated measures ANOVA with task modality and language as within-subjects factors. Because there were only two levels of each factor the sphericity assumption was not violated, and therefore no correction was applied. Results demonstrated that the main effect of task modality on letter recognition was significant, $F(1, 29) = 186.613$, $p < .001$, $\eta_p^2 = 0.865$, with an average (trimmed) letter "*Recognition Score*" being greater in the visual ($M = 0.948$, $SE = 0.015$, 95% CI = [0.917–0.980]) than the tactile ($M = 0.532$, $SE = 0.033$, 95% CI = [0.464–0.599]) modality (Figure 3, panel c). Although the main effect of language was nonsignificant, $F(1, 29) = 3.223$, $p = .083$, $\eta_p^2 = 0.100$, the effect of its interaction with task modality was significant, $F(1, 29) = 6.369$, $p = .017$, $\eta_p^2 = 0.180$.

Similar to the analysis of arcsine "*Recognition Score*" data log "*Recognition Time*" data were also analyzed in a two-way repeated measures ANOVA, using task modality and language as within-subjects factors. Results demonstrated that the main effects of both task modality and language on letter recognition time were significant, $F(1, 29) = 290.988$, $p < .001$, $\eta_p^2 = 0.909$ for task modality; $F(1, 29) = 5.183$, $p = .030$, $\eta_p^2 = 0.152$ for language, but not the effect of interaction between these two factors, $F(1, 29) = 0.334$, $p = .568$, $\eta_p^2 = 0.011$. On average, the (trimmed) "*Recognition Time*" was shorter in the visual ($M = 2.814$ s, $SE = 0.242$, 95% CI = [2.320–3.308]) than the tactile ($M = 25.364$ s, $SE = 2.124$, 95% CI = [21.020–29.709]) modality (Figure 3, panel d), and shorter for an English ($M = 13.489$, $SE = 1.394$, 95% CI = [10.638–16.339]) than for a Bangla ($M = 14.689$, $SE = 1.414$, 95% CI = [12.356–17.023]) letter (Figure 3, panel f; the difference in trimmed data in this panel appears to be negligible as contradictory to the significant difference detected in ANOVA run on the log-transformed data).

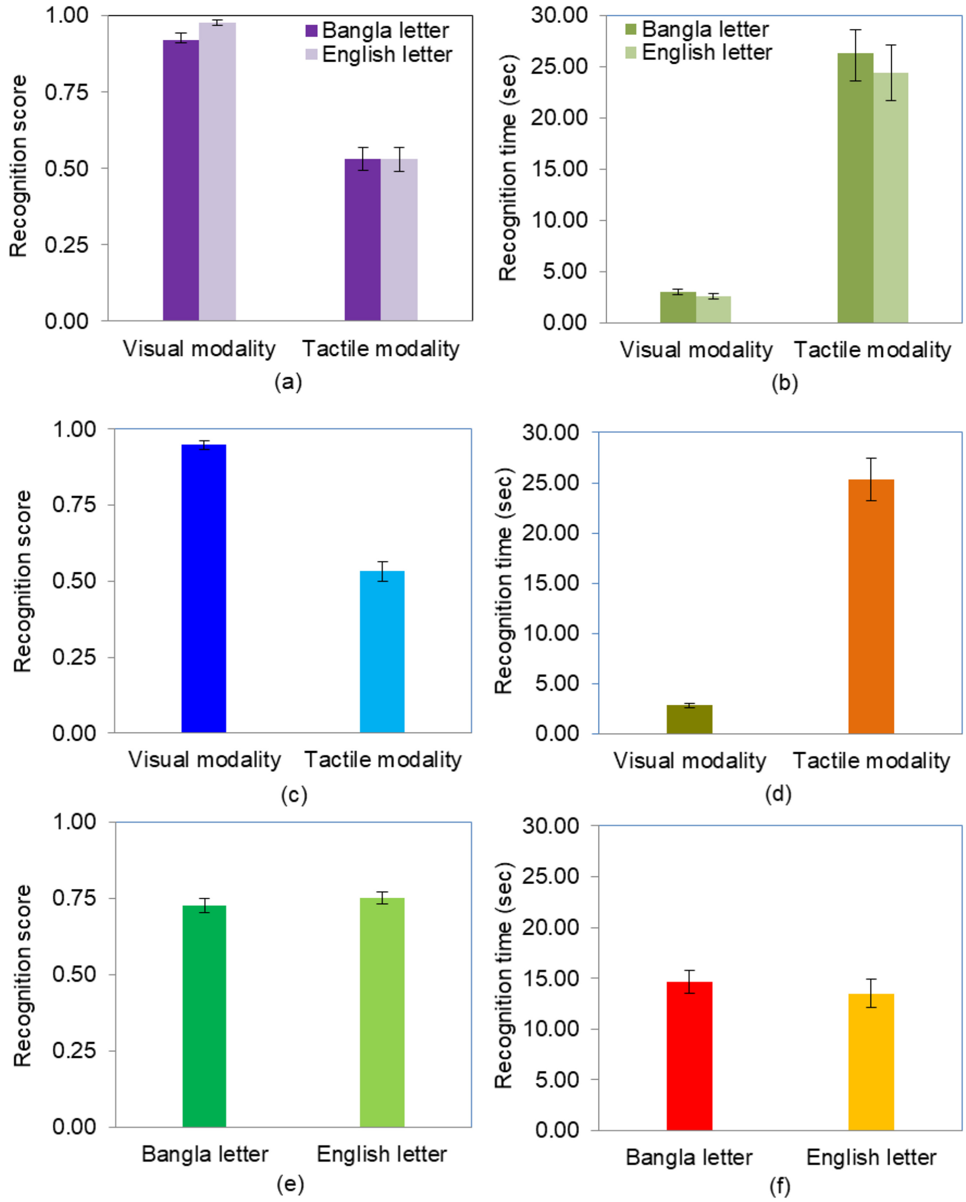
Experiment 2

Performance in Digit Recognition

The average "*Recognition Score*" and average "*Recognition Times*" (s) calculated over 30 participants' (trimmed) raw data for

Figure 3

Letter Recognition Scores and Letter Recognition Times Taken in Visual and Tactile Modalities

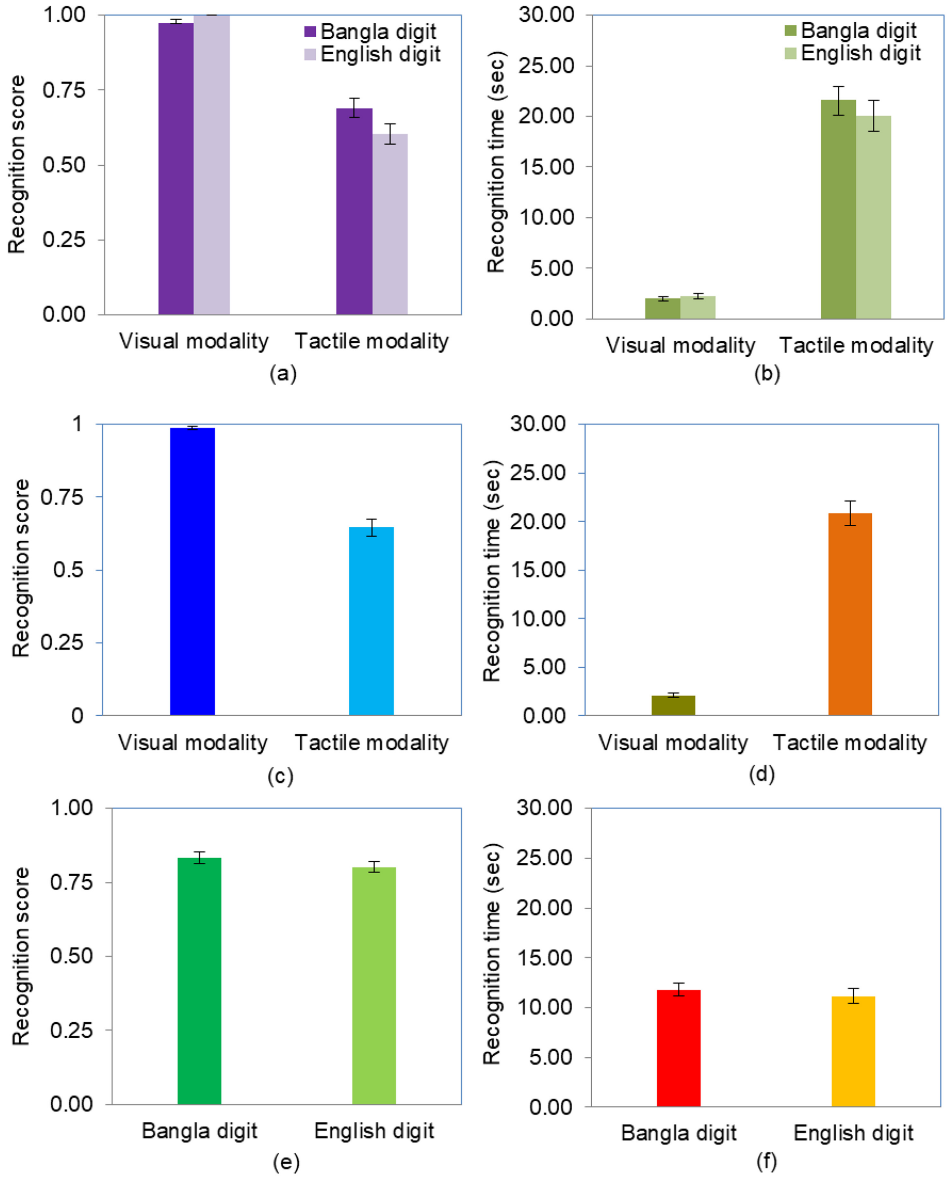


Note. (Panel a) Bangla and English letter recognition scores in visual and tactile modalities, (panel b) Bangla and English letter recognition times in visual and tactile modalities, (panel c) letter recognition scores in visual and tactile modalities irrespective of language, (panel d) letter recognition times in visual and tactile modalities irrespective of language, (panel e) Bangla and English letter recognition scores irrespective of task modality, and (panel f) Bangla and English letter recognition times irrespective of task modality. Error bars represent *standard errors* of the means. See the online article for the color version of this figure.

Bangla or English digits in both visual and tactile modalities are plotted in Figure 4. This figure shows that the average "*Recognition Score*" for a Bangla digit was 0.973 ($SE = 0.012$, 95% CI = [0.949–0.997]) in the visual, and 0.690 ($SE = 0.034$, 95% CI = [0.621–0.759]) in the tactile modality (panel a), with the corresponding "*Recognition Times*" being 2.003 s ($SE = 0.227$, 95% CI = [1.540–2.466]) and 21.620 s ($SE = 1.282$, 95% CI = [18.998–24.242]), respectively (panel b). The average "*Recognition Score*" for an English digit was 1.000 ($SE = 0.000$, 95% CI = [1.000–1.000]) in the visual and 0.603 ($SE = 0.033$, 95% CI = [0.535–0.672]) in the tactile modality (panel a), with the corresponding "*Recognition Times*" being 2.219 s ($SE = 0.231$, 95% CI = [1.746–2.691]) and 20.061 s ($SE = 1.524$, 95% CI = [16.944–23.178]), respectively (panel b).

Figure 4

Digit Recognition Scores and Digit Recognition Times Taken in Visual and Tactile Modalities



Note. (Panel a) Bangla and English digit recognition scores in visual and tactile modalities, (panel b) Bangla and English digit recognition times in visual and tactile modalities, (panel c) digit recognition scores in visual and tactile modalities irrespective of language, (panel d) digit recognition times in visual and tactile modalities irrespective of language, (panel e) Bangla and English digit recognition scores irrespective of task modality, and (panel f) Bangla and English digit recognition times irrespective of task modality. Error bars represent *standard errors* of the means. See the online article for the color version of this figure.

When the trimmed "*Recognition Scores*" were transformed to arcsine (see above) and analyzed in a series of one-sample *t*-tests results demonstrated that the arcsine "*Recognition Scores*' in the tactile modality were significantly different from zero, the baseline performance, for both Bangla ($t = 26.483$, $p < .001$, Cohen's $d = 4.918$ equivalent to $\eta_p^2 = 0.858$) and English ($t = 20.985$ $p = .001$, Cohen's $d = 3.897$ equivalent to $\eta_p^2 = 0.791$) digits. Similarly, the arcsine "*Recognition Scores*" in the visual modality were significantly different from zero, the baseline performance, for both Bangla ($t = 52.465$, $p < .001$, Cohen's $d = 9.743$ equivalent to $\eta_p^2 = 0.959$) and English ($t = 1.528^{-16}$, $p < .001$, Cohen's d was not computed here as the standard deviation was 0 after data trimming) digits.

To see the effects of task modality and language on digit recognition arcsine data were analyzed in a two-way repeated measures ANOVA with task modality and language as within-subjects factors.

Because there were only two levels of each factor the sphericity assumption was not violated, and therefore no correction was applied. Results demonstrated that the main effect of task modality on digit recognition was significant, $F(1, 29) = 280.258$, $p < .001$, $\eta_p^2 = 0.906$, with an average (trimmed) digit "*Recognition Score*" being greater in the visual ($M = 0.987$, $SE = 0.006$, 95% CI = [0.975–0.999]) than the tactile ($M = 0.647$, $SE = 0.029$, 95% CI = [0.586–0.707]) modality (Figure 4, panel c). The main effect of language was nonsignificant, $F(1, 29) = 0.097$, $p = .757$, $\eta_p^2 = 0.003$; however, the effect of its interaction with task modality was significant, $F(1, 29) = 14.629$, $p = .001$, $\eta_p^2 = 0.335$.

Similar to the analysis of arcsine "*Recognition Score*" data log "*Recognition Time*" data were also analyzed in a two-way repeated measures ANOVA, using task modality and language as within-subjects factors. Results demonstrated that the effect of task modality on digit recognition time was significant, $F(1, 29) = 518.346$, $p < .001$, $\eta_p^2 = 0.947$, with an average (trimmed) "*Recognition Time*" being shorter in the visual ($M = 2.111$ s, $SE = 0.210$, 95% CI = [1.681–2.541]) than the tactile ($M = 20.841$ s, $SE = 1.272$, 95% CI = [18.239–23.442]) modality (Figure 4, panel d). The main effect of language was nonsignificant, $F(1, 29) = 0.022$, $p = .884$, $\eta_p^2 = 0.001$; however, the effect of its interaction with task modality was significant, $F(1, 29) = 7.619$, $p = .010$, $\eta_p^2 = .208$.

Aggregate Data for Experiments 1 and 2

Effect of Character Type on Character Recognition

Because letter recognition and digit recognition were assessed in the same group of children successively in two different experiments, it was reasonable to collapse the data across experiments, and further analyze them in three-way repeated measures ANOVAs, taking task modality, language, and character type (digit, letter) as independent factors. Because the effects of task modality and language were tested for the data of individual experiments the main purpose of the analysis of aggregate data was to see the effect of character type on character recognition, and whether character type interacts with the two other factors as language does interact with task modality (see Results in Experiments 1 and 2).

The analysis of arcsine recognition data revealed that the main effects of character type, $F(1, 29) = 25.153$, $p < .001$, $\eta_p^2 = .464$ and task modality, $F(1, 29) = 287.642$, $p < .001$, $\eta_p^2 = .908$, but not the main effect of language, $F(1, 29) = 1.024$, $p = .320$, $\eta_p^2 = .034$, on character recognition were significant. Data plotted in Figure 5 show that the average (trimmed) "*Recognition Score*" for a digit was 0.987 ($SE = 0.006$, 95% CI = [0.975–0.999]) in the visual, and 0.647 ($SE = 0.029$, 95% CI = [0.586–0.707]) in the tactile modality (panel a). On the contrary, the average (trimmed) "*Recognition Score*" for a letter was 0.948 ($SE = 0.015$, 95% CI = [0.917–0.980]) in the visual, and 0.532 ($SE = 0.033$, 95% CI = [0.464–0.599]) in the tactile modality (panel a). Irrespective of task modality and language, the average (trimmed) "*Recognition Score*" was greater for a digit ($M = 0.817$, $SE = 0.016$, 95% CI = [0.784–0.849]) than for a letter ($M = 0.740$, $SE = 0.019$, 95% CI = [0.702–0.778]; panel c), and irrespective of character type and language, the average (trimmed) "*Recognition Score*" was greater in the visual ($M = 0.968$, $SE = 0.010$, 95% CI = [0.948–0.987]) than the tactile ($M = 0.589$, $SE = 0.028$, 95% CI = [0.532–0.646]) modality (panel e). The effect of interaction between task modality and language was significant, $F(1, 29) = 16.020$, $p < .001$, $\eta_p^2 = .356$. However, the effects of interaction between task modality and character type, between language and character type, and between these three factors were all nonsignificant, $F(1, 29) = 1.323$, $p = .259$, $\eta_p^2 = .044$, for Task Modality × Character Type; $F(1, 29) = 2.983$, $p = .095$, $\eta_p^2 = .093$, for Language × Character Type; $F(1, 29) = 0.821$, $p = .372$, $\eta_p^2 = .028$, for Task Modality × Language × Character Type.

The analysis of log "*Recognition Time*" data revealed that the main effects of character type, $F(1, 29) = 14.399$, $p = .001$, $\eta_p^2 = .332$, and task modality, $F(1, 29) = 404.426$, $p < .001$, $\eta_p^2 = .933$, but not the main effect of language, $F(1, 29) = 3.420$, $p = .075$, $\eta_p^2 = .105$, on recognition time were significant. Data plotted in Figure 5 show that the average (trimmed) "*Recognition Time*" for a digit was 2.111 s ($SE = 0.210$, 95% CI = [1.681–2.541]) in the visual, and 20.841 s ($SE = 1.272$, 95% CI = [18.239–23.442]) in the tactile modality (panel b). On the contrary, the average (trimmed) "*Recognition Time*" for a letter was 2.814 s ($SE = 0.242$, 95% CI = [2.320–3.308]) in the visual, and 25.364 s ($SE = 2.124$, 95% CI = [21.020–29.709]) in the tactile modality (panel b). Irrespective of task modality and language, the average (trimmed) "*Recognition Time*" was shorter for a digit ($M = 11.476$ s, $SE = 0.637$, 95% CI = [10.173–12.778]) than for a letter ($M = 14.089$ s, $SE = 1.085$, 95% CI = [11.870–16.309]; panel d), and irrespective of character type and language, the average (trimmed) "*Recognition Time*" was shorter in the visual ($M = 2.462$ s, $SE = 0.219$, 95% CI = [2.014–2.911]) than the tactile ($M = 23.103$ s, $SE = 1.538$, 95% CI = [19.956–26.249]) modality (panel f). The effects of interaction between task modality and character type, and between task modality, language and character type were significant, $F(1, 29) = 10.193$, $p = .003$, $\eta_p^2 = .260$, for Task Modality × Character Type; $F(1, 29) = 4.439$, $p = .044$, $\eta_p^2 = .133$, for Task Modality × Language × Character Type. However, the other two two-factor interaction effects were nonsignificant, $F(1, 29) = 1.010$, $p = .323$, $\eta_p^2 = .034$, for Task Modality × Language; $F(1, 29) = 3.451$, $p = .073$, $\eta_p^2 = .106$, for Language × Character Type.

Character Classification Performance

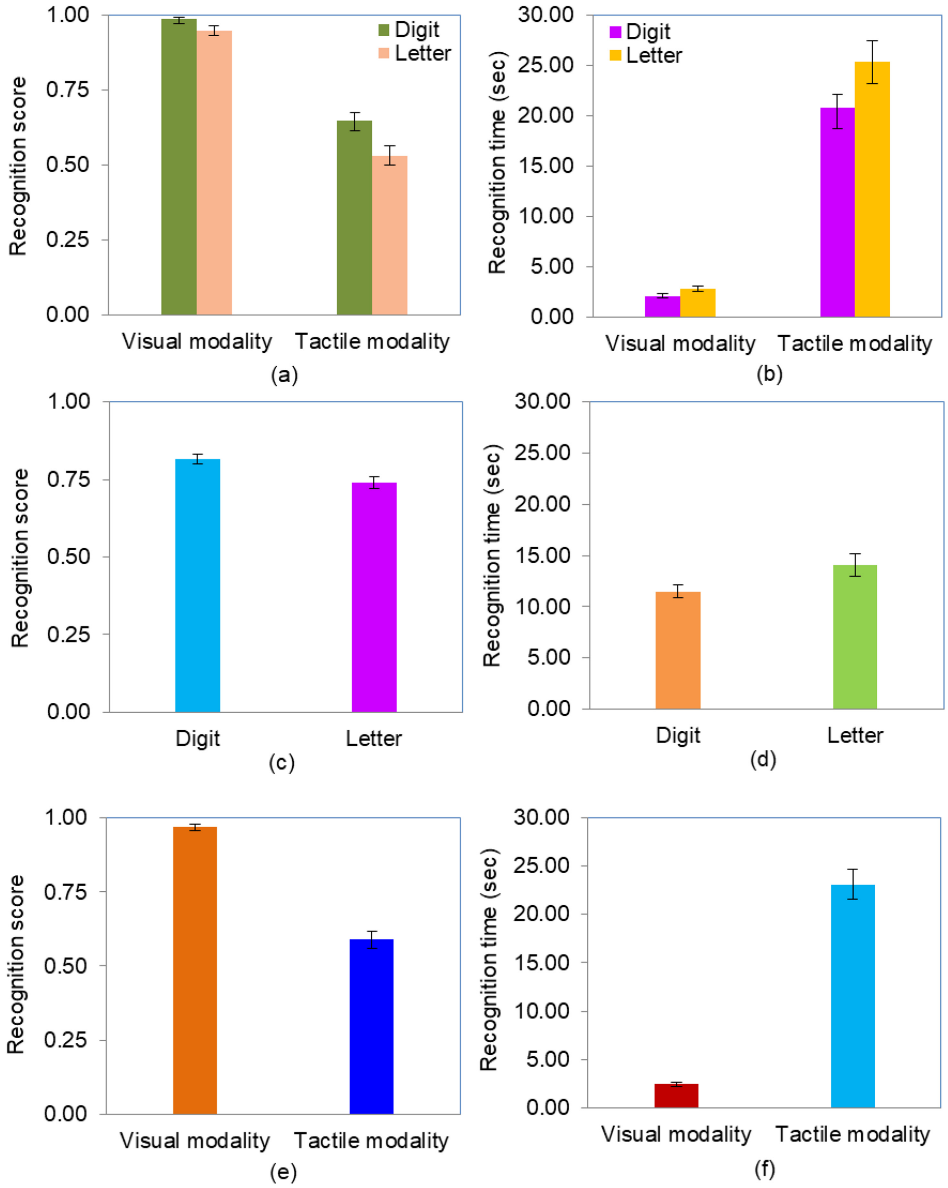
Experiment 1

Performance in Letter Classification

The visual and tactile confusion matrices computed for Bangla letters are compositely presented in Table 1, and those generated for English letters are compositely presented in Table 2. All these stimulus/response matrices look asymmetrical and rectangular as they expanded to the right to show participants' omissions and substitution responses that were not part of the stimulus set used in the experiment (see Danhauer & Lucks, 1987). Each cell entry in the matrices was calculated as the proportion of 30 participants' pooled responses to a letter on which the row stimulus resulted in the column response. The diagonal cell values represent the proportions of correct responses or correctly classified outcomes and off-diagonal cell values represent the proportions of confusions or misclassified outcomes. The former values indicate the degree to which children's average performance for individual letters was greater than the chance level, and the latter values indicate how the letters in each language were mistaken for each other.

Figure 5

Recognition Scores and Recognition Times Taken in Visual and Tactile Modalities



Note. (Panel a) Digit and letter recognition scores in visual and tactile modalities irrespective of language, (panel b) digit and letter recognition times in visual and tactile modalities irrespective of language, (panel c) letter and digit recognition scores irrespective of task modality and language, (panel d) digit and letter recognition times irrespective of task modality and language, (panel e) character recognition scores in visual and tactile modalities irrespective of language and character type, and (panel f) character recognition times in visual and tactile modalities irrespective of character type and language. Only data for significant effects are plotted, with error bars representing standard errors of the means. See the online article for the color version of this figure.

The classification metrics for each of the 10 classes depicted at the bottom of Table 1 further give a comparative picture of the visual and tactile performance in the Bangla letter classification task (for a visual picture of this comparison see Figure A1, panels a–c in Appendix). Specifically, the values of "Recall" demonstrated that the classification accuracies for individual letters highly varied within the tactile, and between the tactile and visual modalities. The classification accuracies for various letters ranged from 80% (ং and ঃ) to 100% (ক, খ, গ, and ঁ) in the visual, and from 20% (ধ) to 83% (ক) in the tactile modality. The classification accuracy for all 10 letters, on average, along the main diagonal was 94% in the visual and 52% in the tactile modality. Both these sensory performances are extremely greater than the chance level, 2% (as it was a letter naming task the child could respond with any of the

Table 1

Confusion Matrices and Classification Performance Metrics for Two-Sensory Datasets Obtained in Bangla Letter Classification Task

(a)

Stimulus letters	Response letters: ক	খ	গ	ধ	ন	প	ফ	ং	ঃ	৬	*ণ	*ব	*ৎ	SUM
ক	**1**	0	0	0	0	0	0	0	0	0	0	0	0	1
খ	0	**1**	0	0	0	0	0	0	0	0	0	0	0	1
গ	0	0	**1**	0	0	0	0	0	0	0	0	0	0	1
ধ	0	0	0	**.90**	0	0	0	0	0	0	0	.10	0	1
ন	0	0	0	0	**.97**	0	0	0	0	0	.03	0	0	1
প	0	0	.03	0	0	**.97**	0	0	0	0	0	0	0	1
ফ	.03	0	0	0	0	0	**.97**	0	0	0	0	0	0	1
ং	0	0	0	0	0	0	0	**.80**	.03	0	0	0	.17	1
ঃ	0	0	0	0	0	0	0	.20	**.80**	0	0	0	0	1
৬	0	0	0	0	0	0	0	0	0	**1**	0	0	0	1
SUM	1.03	1.0	1.03	.90	.97	.97	.97	1.0	.83	1.0				
Recall:	1.0	1.0	1.0	.90	.97	.97	.97	.80	.80	1.0	**(*M* = .94)**			
Precision:	.97	1.0	.97	1.0	1.0	1.0	1.0	.80	.96	1.0	**(*M* = .97)**			
F1-score:	.98	1.0	.98	.95	.98	.98	.98	.80	.87	1.0	**(*M* = .95)**			

(b)

Stimulus letters	Response letters: ক	খ	গ	ধ	ন	প	ফ	ং	ঃ	৬	*ঘ	*ণ	*দ	*ব	*স	*ৎ	SUM
ক	**.83**	0	0	0	0	0	.17	0	0	0	0	0	0	0	0	0	1
খ	0	**.47**	.47	0	0	0	0	0	0	0	.03	0	0	0	.03	0	1
গ	0	0	**.47**	0	0	0	0	0	0	0	.03	0	0	0	.50	0	1
ধ	0	0	0	**.20**	0	0	0	0	0	0	0	0	.03	.77	0	0	1
ন	0	0	0	0	**.63**	.23	0	0	0	0	0	.13	0	0	0	0	1
প	0	0	.20	0	0	**.70**	0	0	0	0	0	0	0	.10	0	0	1
ফ	.73	0	0	0	0	0	**.23**	0	0	0	0	0	0	.03	0	0	1
ং	0	0	0	0	0	0	0	**.40**	.07	0	0	0	0	0	0	.53	1
ঃ	0	0	0	0	0	0	0	.37	**.50**	0	0	0	0	0	0	.13	1
৬	0	0	0	0	0	0	0	.03	.07	**.80**	0	0	0	0	0	.10	1
SUM	1.56	.47	1.14	.20	.63	.93	.40	.80	.64	.80							
Recall:	.83	.47	.47	.20	.63	.70	.23	.40	.50	.80	**(*M* = .52)**						
Precision:	.53	1.0	.41	1.0	1.0	.75	.58	.50	.78	1.0	**(*M* = .69)**						
F1-score:	.65	.64	.43	.33	.78	.72	.33	.44	.61	.89	**(*M* = .59)**						

Note. (a) Visual classification of Bangla letters and (b) Tactile classification of Bangla letters. *Response letter outside the stimulus set. Each cell entry was calculated on the basis of 30 participants' pooled responses. The diagonal cell values represent the proportions of correct responses or correctly classified outcomes and off-diagonal cell values represent the proportions of confusions or misclassified outcomes. The values are shown with two digits after the decimal point following conventional mathematical rules. See the online article for the color version of this table.

50 letters available in the Bangla alphabet set). The "Precision" achieved by various letters ranged from 80% (ং) to 100% (খ, ধ, ন, প, ফ, and ৬) with an average "Precision" of 97% in the visual, and from 41% (গ) to 100% (ক, ধ, ন, and ৬) with an average "Precision" of 69% in the tactile modality. As demonstrated by "F1-scores," the testing accuracy on the letter set ranged from 80% (ং) to 100% (খ and ৬) with an average accuracy of 95% in the visual, and from 33% (ধ and ফ) to 89% (৬) with an average accuracy of 59% in the tactile modality. All these classification metrics together indicate that the visual modality outperformed the tactile modality in the Bangla letter classification task.

Table 1 further shows that the distribution of confusions (misclassifications) in the Bangla letter classification task was highly asymmetric both within and between sensory modalities (for a visual picture of this asymmetry see Figure A1, panel d in Appendix). A look at the off-diagonal elements of the two matrices indicates that confusion occurred for six letters (ধ, ন, প, ফ, ং, and ঃ) in the visual modality and for each and every letter in the tactile modality. In the visual modality, the observed confusion rates for the mentioned six letters varied within a range of 3% to 20%, with the lowest rate recorded for "ন," "প, and "ফ" and the highest rate recorded for "ং" and " ঃ." Specifically, "ন" was reported as "ণ" (a letter outside the original stimulus set) by 3% of children, "ধ" was reported as "ব" by 10% of children, "প" was reported as "গ" by 3% of children, "ফ" was reported as "ক" by 3% of children, "ঃ " was reported as "ং" by 20% of children, and "ং" was reported as "ৎ" by 17% of children and as "ঃ" by 3% of children. In the tactile modality, the observed confusion rates varied within a range of 17% to 80%, with the lowest rate recorded for "ক" and the highest rate recorded for "ধ." Specifically, "ক" was reported as "ফ" by 17% of children, and "ধ" was reported as "ব" by 77% of children and as "দ" by 3% of children. The remaining eight letters (খ, গ, ন, প, ফ, ং, ঃ, and ৬) in this sensory modality were found to be confused with various letters both inside and outside the original stimulus set, yielding a confusion rate of 20% (৬) to 77% (ফ). On average, the observed rate of confusion errors occurring in the Bangla letter classification task was much smaller in the visual (6%) than the tactile (48%) modality.

The classification metrics for each of the 10 classes depicted at the bottom of Table 2 further give a comparative picture of the visual and tactile performance in the English letter classification task (for a visual picture of this comparison see Figure A2, panels a–c in Appendix). Specifically, the values of "Recall" demonstrated that the classification accuracies for individual letters highly varied within the tactile modality, and between the tactile and visual modalities. The classification accuracies for various letters ranged from 87% (Q) to 100% (A, B, C, N, O, X, and Z) in the visual, and from 13% (Q) to 87% (A) in the tactile modality. The classification accuracy for all 10 letters, on average, along the main diagonal was 98% in the visual and 54% in the tactile modality. Both these sensory performances are extremely greater than the chance level, 3.85% (as it was a letter naming task the child could respond with any of the 26 letters available in the English alphabet set). The "Precision" achieved by various letters ranged from 88% (O) to 100% (A, B, C, L, N, Q, Y, and Z) with an average "Precision" of 98% in the visual, and from 39% (O) to 100% (A and B) with an average "Precision" of 65% in the tactile modality. As demonstrated by "F1-scores," the testing accuracy on the letter set ranged from 92% (Q) to 100% (A, B, C, N, and Z) with an average accuracy of 98% in the visual, and from 22% (Q) to 92% (A) with an average accuracy of 59% in the tactile modality.

Table 2
Confusion Matrices and Classification Performance Metrics for Two-Sensory Datasets Obtained in English Letter Classification Task

(a)

Stimulus letters	Response letters: A	B	C	L	N	O	Q	X	Y	Z	*I	SUM
A	**1**	0	0	0	0	0	0	0	0	0	0	1
B	0	**1**	0	0	0	0	0	0	0	0	0	1
C	0	0	**1**	0	0	0	0	0	0	0	0	1
L	0	0	0	**.97**	0	0	0	0	0	0	.03	1
N	0	0	0	0	**1**	0	0	0	0	0	0	1
O	0	0	0	0	0	**1**	0	0	0	0	0	1
Q	0	0	0	0	0	.13	**.87**	0	0	0	0	1
X	0	0	0	0	0	0	0	**1**	0	0	0	1
Y	0	0	0	0	0	0	0	.03	**.97**	0	0	1
Z	0	0	0	0	0	0	0	0	0	**1**	0	1
SUM	1.0	1.0	1.0	.97	1.0	1.13	.87	1.03	.97	1		
Recall:	1.0	1.0	1.0	.97	1.0	1.0	.87	1.0	.97	1.0	**(*M* = .98)**	
Precision:	1.0	1.0	1.0	1.0	1.0	.88	1.0	.97	1.0	1.0	**(*M* = .98)**	
F1- score:	1.0	1.0	1.0	.98	1.0	.93	.92	.98	.98	1.0	**(*M* = .98)**	

(b)

Stimulus letters	Response letters: A	B	C	L	N	O	Q	X	Y	Z	*D	*I	*M	*P	SUM
A	**.87**	0	0	0	.03	0	0	.03	.07	0	0	0	0	0	1
B	0	**.60**	0	0	0	0	0	0	0	0	.13	0	0	.27	1
C	0	0	**.57**	0	0	.40	.03	0	0	0	0	0	0	0	1
L	0	0	0	**.67**	0	0	0	0	0	0	0	.33	0	0	1
N	0	0	0	0	**.37**	0	0	0	0	.10	0	0	.53	0	1
O	0	0	.07	0	0	**.80**	0	0	0	0	.13	0	0	0	1
Q	0	0	.03	0	0	.83	**.13**	0	0	0	0	0	0	0	1
X	0	0	0	0	0	0	0	**.53**	.47	0	0	0	0	0	1
Y	0	0	0	0	0	0	0	.57	**.43**	0	0	0	0	0	1
Z	0	0	0	.03	.30	0	0	0	0	**.40**	0	0	.27	0	1
SUM	.87	.60	.67	.70	.70	2.03	.16	1.13	.97	.50					
Recall:	.87	.60	.57	.67	.37	.80	.13	.53	.43	.40	**(*M* = .54)**				
Precision:	1.0	1.0	.85	.96	.53	.39	.81	.47	.44	.80	**(*M* = .65)**				
F1- score:	.92	.75	.68	.79	.44	.52	.22	.50	.44	.53	**(*M* = .59)**				

Note. (a) Visual classification of English letters and (b) Tactile classification of English letters. *Response letter outside the stimulus set. Each cell entry was calculated on the basis of 30 participants' pooled responses. The diagonal cell values represent the proportions of correct responses or correctly classified outcomes and off-diagonal cell values represent the proportions of confusions or misclassified outcomes. The values are shown with two digits after the decimal point following conventional mathematical rules. See the online article for the color version of this table.

All these classification metrics together indicate that the visual modality outperformed the tactile modality also in the English letter classification task similar to the Bangla letter recognition task as outlined above.

Table 2 further shows that similar to the distribution of confusion in the Bangla letter classification task, the distribution of confusions in the English letter classification task was also highly asymmetric within the tactile, and between the tactile and visual modalities (for a visual picture of this asymmetry, see Figure A2, panel d in Appendix). A look at the off-diagonal elements of the two matrices indicates that confusion occurred only for three letters (L, Q, and Y) in the visual modality, and for each and every letter in the tactile modality. In the visual modality, the observed confusion rates for the mentioned three letters varied within a range of 3% to 13%, with the lowest rate recorded for L and Y and the highest rate recorded for Q that was reported as "O." Both "L" and "Y" yielded a confusion rate of 3%, with the former one being reported as "I" (a letter outside the original stimulus set) and the latter one as "X." In the tactile modality, the observed confusion rates varied within a range of 13% to 87%, with the lowest rate recorded for "A" and the highest rate recorded for "Q." Here, "A" was reported as "N" by 3% of children, as "X" by 3% of children and as "Y" by 7% of children, and "Q" was reported as "C" by 3% of children and as "O" by 83% of children. The remaining eight letters (B, C, L, N, O, X, Y, and Z) in this sensory modality were found to be confused with various letters both inside and outside the original stimulus set, yielding a confusion rate of 20% (O) to 63% (N). On average, the rate of confusion errors occurring in the English letter classification task was much smaller in the visual (2%) than the tactile (46%) modality, similar to the rate of confusion errors occurring in the Bangla letter classification task as outlined above.

Experiment 2

Performance in Digit Classification

The visual and tactile confusion matrices computed for Bangla digits are compositely presented in Table 3, and those computed for English digits are compositely presented in Table 4. Unlike the matrices for letters, each of these matrices is a symmetrical 10 × 10 "square" matrix that contains responses exclusively from within the stimulus set used in the experiment (see Danhauer & Lucks, 1987). Each diagonal and off-diagonal cell entry in the matrices was calculated following the same procedure stated above in Experiment 1. The diagonal values indicate the degree to which children's average performance for individual digits was greater than the chance level, and the off-diagonal values indicate how the digits in each language were mistaken for each other.

The classification metrics for each of the 10 classes depicted at the bottom of Table 3 further give a comparative picture of the visual and tactile performance in the Bangla digit classification task (for a visual picture of this comparison see Figure A3, panels a–c in Appendix). Specifically, the values of "Recall" demonstrated that the classification accuracies for individual digits highly varied within the tactile, and between the tactile and visual modalities. The classification accuracies for various digits ranged from 93% (৫ and ৯) to 100% (০, ১, ২, ৩, and ৭) in the visual, and from 47% (৯) to 97% (১) in the tactile modality. The classification

Table 3

Confusion Matrices and Classification Performance Metrics for Two-Sensory Datasets Obtained in Bangla Digit Classification Task

(a)

Stimulus digits	Response digits ০	১	২	৩	৪	৫	৬	৭	৮	৯	SUM
০	**1.0**	0	0	0	0	0	0	0	0	0	1
১	0	**1.0**	0	0	0	0	0	0	0	0	1
২	0	0	**1.0**	0	0	0	0	0	0	0	1
৩	0	0	0	**1.0**	0	0	0	0	0	0	1
৪	0	0	0	0	**.93**	.03	0	0	0.3	0	1
৫	0	0	0	0	0	**.97**	.03	0	0	0	1
৬	0	0	0	0	0	.03	**.97**	0	0	0	1
৭	0	0	0	0	0	0	0	**1.0**	0	0	1
৮	0	0	0	0	0	0	0	.03	**.97**	0	1
৯	0	.07	0	0	0	0	0	0	0	**.93**	1
SUM	1.0	1.07	1.0	1.0	.93	1.03	1.0	1.03	1.0	.93	
Recall :	1.0	1.0	1.0	1.0	.93	.97	.97	1.0	.97	.93	**(*M* = .98)**
Precision:	1.0	.94	1.0	1.0	1.0	.94	.97	.97	.97	1.0	**(*M* = .98)**
F1 -score:	1.0	.97	1.0	1.0	.97	.95	.97	.98	.97	.97	**(*M* = .98)**

(b)

Stimulus digits	Response digits ০	১	২	৩	৪	৫	৬	৭	৮	৯	SUM
০	**.90**	0	0	.10	0	0	0	0	0	0	1
১	0	**.97**	0	0	0	0	0	0	0	.03	1
২	0	0	**.67**	0	0	0	0	.03	0	.30	1
৩	.13	0	0	**.87**	0	0	0	0	0	0	1
৪	0	0	0	0	**.73**	.03	0	.10	.13	0	1
৫	0	0	0	0	0	**.50**	.47	0	.03	0	1
৬	0	0	0	.03	0	.30	**.67**	0	0	0	1
৭	0	0	0	0	0	0	0	**.53**	.47	0	1
৮	0	0	0	0	.03	0	0	.30	**.67**	0	1
৯	0	.53	0	0	0	0	0	0	0	**.47**	1
SUM	1.03	1.50	.67	1.0	.76	.83	1.14	.96	1.30	.80	
Recall :	.90	.97	.67	.87	.73	.50	.67	.53	.67	.47	**(*M* = .70)**
Precision:	.87	.65	1.0	.87	.96	.60	.59	.55	.52	.59	**(*M* = .70)**
F1 -score:	.89	.78	.80	.87	.83	.55	.63	.54	.59	.52	**(*M* = .70)**

Note. (a) Visual classification of Bangla digits and (b) Tactile classification of Bangla digits. Each cell entry was calculated on the basis of 30 participants' pooled responses. The diagonal cell values represent the proportions of correct responses or correctly classified outcomes and off-diagonal cell values represent the proportions of confusions or misclassified outcomes. The values are shown with two digits after the decimal point following conventional mathematical rules. See the online article for the color version of this table. See the online article for the color version of this table.

accuracy for all 10 digits, on average, along the main diagonal was 98% in the visual and 70% in the tactile modality. Both these sensory performances are extremely greater than the chance level, 10% (as it was a digit naming task the child could respond with any of the 10 digits in Bangla). The "Precision" achieved by various digits ranged from 94% (১ and ৫) to 100% (০, ২, ৩, ৪, and ৯) with an average "Precision" of 98% in the visual, and from 52% (৮) to 100% (২) with an average "Precision" of 70% in the tactile modality. As demonstrated by "F1-scores," the testing accuracy on the digit set ranged from 95% (৫) to 100% (০, ২, and ৩) with an average accuracy of 98% in the visual, and from 52% (৯) to 89% (০) with an average accuracy of 70% in the tactile modality. All these classification metrics together indicate that the visual modality outperformed the tactile modality in the Bangla digit classification task.

Table 3 further shows that the distribution of confusions (misclassifications) in the Bangla digit classification task was highly asymmetric both within and between sensory modalities (for a visual picture of this asymmetry see Figure A3, panel d in Appendix). A look at the off-diagonal elements of the two matrices indicates that confusion occurred for five digits (৪, ৫, ৬, ৮, and ৯) in the visual modality and for each and every digit in the tactile modality. In the visual modality, the observed confusion rates for the mentioned digits varied within a range of 3% to 7%, with the lowest rate recorded for ৫, ৬, ৮ and the highest rate recorded for ৪ and ৯. Specifically, ৫, ৬, and ৮ were reported as ৬, ৫, and ৭ respectively, each one yielding a confusion rate of 3%. "৪" was reported as "৫" by 3% of children and as "৮" by 3% of children, and "৯" was reported as "১" by 7% of children. In the tactile modality, the observed confusion rates varied within a range of 3% to 53%, with the lowest rate recorded for "১" and the highest rate recorded for "৯." Here, "১" was reported as "৯" by 3% of children and "৯" was reported as "১" by 53% of children. The remaining eight digits (০, ২, ৩, ৪, ৫, ৬, ৭, and ৮) in this sensory modality were found to be confused with various digits, yielding a confusion rate of 10% (০) to 50% (৫). On average, the observed rate of confusion errors occurring in the Bangla digit classification task was much smaller in the visual (2%) than the tactile (30%) modality.

The classification metrics for each of the 10 classes depicted at the bottom of Table 4 further give a comparative picture of the visual and tactile performance in the English digit classification task (for a visual picture of this comparison see Figure A4, panels a–c in Appendix). The values of "Recall" demonstrated that the classification accuracies for individual digits varied highly within the tactile, and between the visual and tactile modalities. The classification accuracies for various digits ranged from 97% (5) to 100% (all other digits) in the visual modality, and from 23% (7) to 93% (0) in the tactile modality. The classification accuracy for all 10 digits, on average, along the main diagonal was approximately 100% in the visual and 60% in the tactile modality. Both these sensory performances are extremely greater than the chance level, 10% (as it was a digit naming task the child could respond with any of the 10 digits in English). The "Precision" achieved by various digits ranged from 97% (3) to 100% (all other digits) with an average "Precision" of approximately 100% in the visual modality, and from 39% (5) to 100% (0) with an average "Precision" of 60% in the tactile modality. As demonstrated by "F1-scores," the testing accuracy achieved on the digit set ranged from 98% (3 and 5) to

Table 4

Confusion Matrices and Classification Performance Metrics for Two-Sensory Datasets Obtained in English Digit Classification Task

(a)

Stimulus digits \ Response digits	0	1	2	3	4	5	6	7	8	9	SUM
0	**1.0**	0	0	0	0	0	0	0	0	0	1
1	0	**1.0**	0	0	0	0	0	0	0	0	1
2	0	0	**1.0**	0	0	0	0	0	0	0	1
3	0	0	0	**1.0**	0	0	0	0	0	0	1
4	0	0	0	0	**1.0**	0	0	0	0	0	1
5	0	0	0	.03	0	**.97**	0	0	0	0	1
6	0	0	0	0	0	0	**1.0**	0	0	0	1
7	0	0	0	0	0	0	0	**1.0**	0	0	1
8	0	0	0	0	0	0	0	0	**1.0**	0	1
9	0	0	0	0	0	0	0	0	0	**1.0**	1
SUM	1.0	1.0	1.0	1.03	1.0	.97	1.0	1.0	1.0	1.0	
Recall:	1.0	1.0	1.0	1.0	1.0	.97	1.0	1.0	1.0	1.0	**(*M*≈ 1.0)**
Precision:	1.0	1.0	1.0	.97	1.0	1.0	1.0	1.0	1.0	1.0	**(*M*≈ 1.0)**
F1-score:	1.0	1.0	1.0	.98	1.0	.98	1.0	1.0	1.0	1.0	**(*M*≈ 1.0)**

(b)

Stimulus digits \ Response digits	0	1	2	3	4	5	6	7	8	9	SUM
0	**.93**	0	0	0	0	0	0	0	.03	.03	1
1	0	**.80**	0	0	0	0	0	.20	0	0	1
2	0	0	**.77**	0	.17	0	0	.03	0	.03	1
3	0	0	0	**.77**	0	.10	0	0	.13	0	1
4	0	.07	.43	0	**.37**	0	0	.13	0	0	1
5	0	0	0	.03	0	**.63**	.33	0	0	0	1
6	0	0	0	0	0	.10	**.53**	0	0	.37	1
7	0	.70	0	0	.07	0	0	**.23**	0	0	1
8	0	0	0	.33	.07	0	0	0	**.53**	.07	1
9	0	0	0	.03	.03	0	.50	0	.03	**.40**	1
SUM	.93	1.57	1.20	1.16	.71	.83	1.36	.59	.72	.90	
Recall:	.93	.80	.77	.77	.37	.63	.53	.23	.53	.40	**(*M* = .60)**
Precision:	1.0	.51	.64	.66	.52	.76	.39	.39	.74	.44	**(*M* = .60)**
F1-score:	.97	.62	.70	.70	.43	.70	.45	.29	.62	.42	**(*M* = .60)**

Note. (a) Visual classification of English digits and (b) Tactile classification of English digits. Each cell entry was calculated on the basis of 30 participants' pooled responses. The diagonal cell values represent the proportions of correct responses or correctly classified outcomes and off-diagonal cell values represent the proportions of confusions or misclassified outcomes. The values are shown with two digits after the decimal point following conventional mathematical rules. See the online article for the color version of this table.

100% (all other digits) with an average accuracy of approximately 100% in the visual modality, and from 29% (7) to 97% (0) with an average accuracy of 60% in the tactile modality. All these classification metrics together indicate that the visual modality outperformed the tactile modality in the English digit classification task similar to the Bangla digit classification task as outlined above.

Table 4 further shows that the confusion in the English digit classification task occurred only for one digit (5) in the visual, and for each and every digit in the tactile modality with the distribution of confusions being widely scattered across various digits (for a visual picture of this see Figure A4, panel d in Appendix). A look at the off-diagonal elements in the tactile modality indicates that the observed confusion rates varied within a range of 7% to 77%, with the lowest rate recorded for "0" and the highest rate recorded for "7." "0" was reported as "8" by 3% of children, as "9" by 3% of children, and 7" was reported as "1" by 70% of children and as "4" by 7% of children. The remaining eight digits (1, 2, 3, 4, 5, 6, 8, and 9) in this sensory modality were found to be confused with various digits, yielding a confusion rate of 20% (1) to 63% (4). On average, the observed rate of confusion errors occurring in the English digit classification task was much smaller in the visual (≍0%) than the tactile (40%) modality, similar to the rate of confusion errors occurring in the Bangla digit classification task as outlined above.

General Discussion

This study investigated the crossmodal transfer of linguistic knowledge by examining children's post-learning character recognition skills in the visual modality involved in the acquisition of those characters, and in the tactile modality not directly involved in that acquisition process. The study assessed the capacity of 30 5 to 6-year-old sighted, normal children to recognize visually or tactilely 10 3D letters and 10 3D digits in Bangla or English that they multimodally mastered in home or school settings. Analysis of arcsine recognition data in one sample *t*-tests demonstrated that children's letter or digit recognition scores were significantly different from zero, the baseline performance, for both Bangla and English in both visual and tactile modalities, indicating that the acquisition of letters and digits occurred in the visual modality, and that the acquired recognition skills transferred to the untrained tactile modality as well. The classification performance data obtained from various confusion matrices were subjected to three common classification metrics (Recall, Precision, and F1-score) which revealed that most letters or digits in the test data were identified with high precision and accuracy (that was tremendously greater than the chance level performance), with some letters or digits being more or less confusing in both visual and tactile modalities. This was evident for both Bangla and English letters or digits with some variations across linguistic characters and across task modalities (Tables 1–4; Figures A1–A4 in Appendix). This is a novel and highly significant finding as the participants were quite young children, they were given a character naming (classification) task using 10 shapes, and the tested shapes were intricate linguistic symbols in both experiments. Prior studies in children examined crossmodal transfer using nonverbal stimuli and demonstrated that crossmodal transfer occurs before the age of 1 year (Gottfried et al., 1977; Meltzoff & Borton, 1979; Rose et al., 1981a, 1981b, 1983; Rose & Orlian, 1991; Ruff & Kohler, 1978;

Streri, 1987; Streri & Gentaz, 2004), and sufficiently or significantly improves between the ages of 5 and 8 years (Birch & Lefford, 1963; Picard, 2007; Stoltz-Loike & Bornstein, 1987). Studies in newborns have particularly shown that humans can recognize crossmodal matches without the benefit of months of experience in simultaneous visual–tactual exploration (e.g., Meltzoff & Borton, 1979). On the contrary, one study reported a lack of multisensory integration before the age of 8 years (Gori et al., 2008). This indicates that crossmodal transfer and crossmodal integration might be two independent processes. The former one might be easier and operate earlier than the latter one. Thus, the present findings together with the past ones suggest that crossmodal transferability develops, in the first few years of life, much earlier than the development of crossmodal integration capacity.

How Crossmodal Transfer of Character Learning Might Operate?

The ability of an individual to recognize letters and digits learned multimodally or mostly visually by using a new sensory modality, such as the tactile modality in this study, indicates shared representations of linguistic characters. This fact is popularly known as modality-invariant or amodal representations of sensory information (Erdogan et al., 2015; Yildirim & Jacobs, 2013, 2015). Three principles have been put forward to account for these amodal representations (Karim et al., 2021a). One principle is that the detection of shape invariants across different sensory modalities is a fundamental innate characteristic of a human's perceptual–cognitive system, available without the need for learned correlations (e.g., Meltzoff & Borton, 1979). A second principle posits that humans begin life with independent sensory modalities, and simultaneous tactual and visual exploration of shapes are needed to learn to correlate the separate tactual and visual sense impressions of them (e.g., Meltzoff & Borton, 1979; Speed et al., 2021). Thus crossmodal capacity develops with experience and learning how information from different sensory modalities are combined and integrated for a given stimulus (e.g., Karim et al., 2021a; Speed et al., 2021). A third and related principle states that multisensory experience or learning influences not only the sensory modalities recruited in multisensory processing but the other sensory modalities through neuromodular reorganization as well (Karim et al., 2021a). Because children typically do not use tactile modality to encode linguistic characters the second principle is not good at explaining the findings of the present study. The first principle in combination with the third one better explains these findings. According to the first principle, the brain has an innate capacity to extract modality-independent intrinsic properties of letters and digits like other objects, and use these properties to process and understand the same and similar stimuli when they are sensed by a novel sensory modality (see Erdogan et al., 2015; J. J. Gibson, 1966; Yildirim & Jacobs, 2013, 2015). Thus like the representations of other objects, the representations of a letter or digit are identical when it is viewed, actively touched, or both viewed and touched (see Erdogan et al., 2015). However, the role of mental imagery in such an amodal representation cannot be denied. Someone can argue that through mental imagery the participant might form an instant association between the currently available tactual input which was a letter or digit in this study and the previously stored information about that input in the visual memory (see Berger & Ehrsson, 2018; Nanay, 2018; Schmidt & Blanenburg, 2019; Spence & Deroy, 2013). Such an instant association might be formed on the basis of crossmodal correspondence between information derived from the visual and tactile systems that requires a "transducer" to translate the visual (or tactile) information into a tactile (or visual) code (B. Jones & Connolly, 1970). Because the tactile modality is not directly involved in the acquisition of letters or digits it can be suggested that such a translation is only plausible if there is an intrinsic connection between the two sensory modalities. Thus, the role of mental imagery, if any, does not contradict the first principle. Finally, according to the third principle, the role of neuromodular reorganization operating through multimodal training (without the involvement of touch) on the formation of visual-tactile association cannot be denied. There is evidence that training in a particular task can exert significant changes on the functional architecture of the brain (Carmel & Carrasco, 2008; Karim et al., 2021a; Kelly & Garavan, 2005; Poldrack, 2000; Sasaki et al., 2010). This suggests that the acquisition of letters or digits through sensory training in natural settings can influence brain activity through various mechanisms, including a reorganization or redistribution of neural activity within and across neural networks, within and across sensory modalities (see Karim et al., 2021a). In support of this, prior neuroimaging studies on artificial grapheme–phoneme training with novel letters have demonstrated that several cortical regions become active during the formation of crossmodal associations (Hashimoto & Sakai, 2004; J. S. H. Taylor et al., 2014; Quinn et al., 2017). For example, one early fMRI study in Japanese adults has revealed that the left posterior inferior temporal gyrus and the left parieto-occipital cortex show neuroplasticity in forming new connections between orthography and phonology of Hangul letters that were unfamiliar to them (Hashimoto & Sakai, 2004). Two other studies conducted in native English-speaking healthy adults using monosyllabic consonant–vowel–consonant pseudowords have demonstrated that the parietal brain area is also involved in audio-visual mappings during the literacy acquisition of novel words (J. S. H. Taylor et al., 2014; Quinn et al., 2017). Thus, the impact of multimodal training (audiovisual, audiomotor, visuomotor, and audio-visuo-motor) on linguistic characters might expand or move forward to the somatosensory (tactile) territories, activate them, and make them functional for subsequent processing of the same inputs of a tactile version (see Macaluso et al., 2000; Murray et al., 2005; Naumer et al., 2009; Paraskevopoulos & Herholz, 2013; Shams et al., 2011; Xerri & Zennou-Azogui, 2021).

What Factors Might Determine Character Recognition Skills?

The results of the present study demonstrated that two categories of factors, such as sensory (task) modality and linguistic factors are responsible for determining character recognition skills in children.

Sensory (Task) Modality

Analyses of arcsine recognition data and log recognition time data in repeated measures ANOVAs demonstrated, in support of the first hypothesis, that children were able to recognize strikingly more letters and more digits, and took, on average, exceptionally shorter time for recognizing them in the visual than the tactile modality. Raw data (trimmed) showed that irrespective of language, the mean recognition score (correct response proportion) for a letter was greater in the visual (0.948) than the tactile (0.532) modality by 41.6% (Figure 3, panel c), and similarly, the mean recognition score for a digit was greater in the visual (0.987) than the tactile (0.647) modality by 34.0% (Figure 4, panel c). Moreover, the time taken, on average, for letter recognition in the tactile modality (25.364 s) was 9.01 times the time taken, on average, in the visual modality (2.814 s; Figure 3, panel d). Similarly, the time taken, on average, for digit recognition in the tactile modality (20.841 s) was 9.87 times the time taken in the visual modality (2.111 s; Figure 4, panel d). These differential effects of task modality were further corroborated by the analysis of aggregate data which revealed that irrespective of character type and language, the average recognition score was greater in the visual (0.968) than the tactile (0.589) modality by 37.9% (Figure 5, panel e), and the average time taken to recognize a character, irrespective of character type and language, was 9.38 times shorter in the visual (2.462 s) than the tactile (23.103 s) modality (Figure 5, panel f).

Consistent with the above, classification metrics calculated from various confusion matrices revealed visual superiority over the tactile modality (Tables 1–4). The classification metrics, such as "Recall," "Precision," and "F1-score" all varied across the two sensory modalities. In the visual modality, most letters or digits in the test data were recognized very well with some letters being slightly confusing. Conversely, in the tactile modality, the performance was relatively worse, with many letters or digits being rather more confusing. An inspection of Tables 1–4 (and Figures A1–A4, panels a, in the Appendix) indicates that in the tactile modality, none of the Bangla or English letters or digits achieved 100% classification accuracy, whereas in the visual modality, four Bangla letters (ক, খ, গ, and ঙ), seven English letters (A, B, C, N, O, X, and Z), five Bangla digits (০, ১, ২, ৩, and ৭), and nine English digits (except 5) achieved 100% classification accuracy. A comparison of misclassifications presented in Tables 1–4 (and plotted in Figures A1–A4, panels d, in the Appendix) indicates that the average confusion or misclassification rate was much higher in the tactile than the visual modality. The most highly misclassified Bangla letter was "ধ" (reported as "দ" or "ব") found in the tactile modality (Table 1b), and the least misclassified Bangla letters were "ন," "প," and "ফ" (reported as "ণ," "গ," and "ক," respectively) found in the visual modality (Table 1a). Similarly, the most highly misclassified English letter was "Q" (reported as "O" or "C") found in the tactile modality (Table 2a), and the least misclassified English letters were "L" and "Y" (reported as "I" and "X," respectively) found in the visual modality (Table 2b). Such a difference in misclassification across sensory modalities was also consistently observed in the digit classification task (Tables 3 and 4).

The aforementioned findings together unequivocally bring evidence that vision is much more efficient in the recognition or classification of linguistic characters as compared with touch. In line with this, prior studies have shown that vision has a higher information processing capacity and gives a more accurate perception of the world compared with touch (e.g., Björkman, 1967; Bliss & Hämäläinen, 2005; Cho et al., 2016; Rock & Harris, 1967). Why does vision outperform touch in letter or digit recognition? It is commonly believed that letter or digit recognition is a kind of shape recognition that depends on global feature or macrospatial analysis. Prior studies have reported that vision is more suited than touch for macrospatial analysis (see Bushnell & Boudreau, 1991; Cho et al., 2016; Klatzky et al., 1989; Lederman et al., 1986; Picard, 2007; Tokita & Ishiguchi, 2016), and spatial learning (see Wismeijer et al., 2012). Some studies have further shown that infants have greater speed in encoding visual over tactile information (Rose et al., 1983; Streri & Pecheux, 1986). Thus, one plausible reason for the visual superiority over the tactile modality is that the former modality might have inherently better capacity for shape analysis that is thought to operate through global-to-local or local-to-global processing (Karim et al., 2021b). In global-to-local processing, attention is directed to the whole and encoding spatial relationships between discrete local elements to form a coherent global frame of a stimulus (e.g., Kimchi, 1992; Kovács, 1996; Lewis et al., 2004; Neiworth et al., 2006), whereas in local-to-global processing, attention is gradually extended, following sequential allocation, to the individual local elements that make up the stimulus (Kimchi, 1992; Navon, 1977; Nayar et al., 2015; VanRullen et al., 2007). It is conceived that in order to understand and recognize a letter or a digit the brain needs to construct a coherent global frame of the stimulus. Vision can easily sense and understand the whole shape at once using either of these two strategies. The tactile territory in the brain might not be as efficient as the visual territory in doing this job irrespective of whether it uses a global-to-local or a local-to-global processing strategy. A letter or digit comprises such features as horizontal, vertical, and oblique strokes, curved strokes, dots, and gaps or openings (see E. J. Gibson, 1969; Geyer & DeWald, 1973; Laughery, 1970). These surface-level spatial features are important for identifying and differentiating the letters or digits from one another. This spatial information when received through the visual modality might be well coded and internalized but not when received through the tactile modality (see Blank et al., 1968). Moreover, visual data are maintained in their original code, whereas tactual data are systematically recoded or modified (Friedes, 1974; Pick, 1974). As a result, the memory of tactual information may fade over time (see Dopjans et al., 2009) or may not be retained in a form that can be readily used in the recognition of a letter or digit (see Blank et al., 1968). Thus when we detect some spatial features of a letter through touch and move on to other spatial features the features that we sense first may be lost, making it difficult for us to organize and integrate all those features into a whole. For example, while recognizing Bangla characters, "ধ" can be confused with "ব" due to missing the upper left stroke in "ধ" and "ফ" can be confused with "ক" due to ignoring the upper opening in "ফ," "৬" can be confused with "৫" due to ignoring the upper opening in "৬," and "৯" can be confused with "১" due to ignoring the lower opening in "৯." Similarly, in the classification of English characters, "Q" and "C" can be confused with "O" due to missing the lower stroke in "Q" and ignoring the right opening in "C" respectively, "7" can be confused with "1" due to ignoring or lessening the left upper stroke in "7," and "5" can be confused with "6" due to ignoring the lower left opening in "5" (see Tounsi et al., 2018). This kind of missing or ignorance is unlikely or less likely to occur in a vision that can see and encode the whole shape at a glance (Picard, 2007). This suggests that while organizing individual tactile elements into a whole following Gestalt grouping principles humans may miss or ignore important spatial features that can cause confusion/misclassification more in the tactile than the visual modality. Thus, the recognition/classification of a letter or digit in the tactile modality might be more erroneous and more time-consuming than in the visual modality (see Bliss & Hämäläinen, 2005).

A second and complementary reason that might underlie visual superiority is a differential experience of the two sensory modalities. Children typically use vision (in addition to audition) but not touch to encode letters and digits while learning those characters. The information received through vision during learning might be better coded and internalized than the information instantly received through touch (see Blank et al., 1968). This knowledge is then further consolidated and may put the visual modality at an advantage for later recognition and recall (see Ashton et al., 2018; Brawn et al., 2008; Cuppone et al., 2018; Ruffino et al., 2021). Thus, letters or digits presented visually might be recognized more accurately, more speedily, and fluently, but not when presented tactilely due to the lack of prior experience or familiarity and cognitive mismatching (see Karim et al., 2021b).

Linguistic Factors

Language

Analysis of arcsine recognition data in repeated measures ANOVAs demonstrated that language did not have a significant effect on letter or digit recognition. This finding receives the support from the classification accuracies and confusion rates observed in different confusion matrices (Tables 1–2; Figures A1–A2 in Appendix). Analysis of log recognition time data in repeated measures ANOVAs showed that language did not have any significant effect on log recognition time for a digit but for a letter. Raw data (trimmed) showed that children, on average, took longer time to recognize a Bangla letter (14.689 s) than an English letter (13.489 s). The effect of language–task modality interaction was significant on both letter and digit recognition. Such an interaction effect was also significant on digit recognition time, but not on letter recognition time. Raw data (trimmed) showed that the mean recognition score was greater for an English letter (0.977) than for a Bangla letter (0.92) by 5.7% in the visual modality as opposed to the tactile modality in which the mean recognition score for an English letter (0.53) was equal to that for a Bangla letter (0.533; Figure 3, panel a). Similarly, the mean recognition score was greater for an English digit (1.0) than for a Bangla digit (0.973) by only 2.7% in the visual modality as opposed to the tactile modality in which the mean recognition score was greater for a Bangla digit (0.690) than for an English digit (0.603) by 8.7% (Figure 4, panel a). The mean recognition time for an English digit (2.219 s) and that for a Bangla digit (2.003 s) were almost the same in the visual modality, whereas the mean recognition time was slightly longer for a Bangla digit (21.620 s) than for an English digit (20.061 s) in the tactile modality (Figure 4, panel b). Thus, it is clearly apparent that language alone does not affect letter or digit recognition per se, but is likely to interact with task modality to mediate character recognition capacity in children.

It follows from the aforementioned interaction effects that the visual modality performed better in the recognition of both second-language letters and second-language digits, whereas the tactile modality performed better in the recognition of first-language digits but not first-language letters. The greater recognition score accompanied by longer recognition time within the same sensory modality further supports the notion that these two are different and dissociable measures of character recognition (see Prinzmetal et al., 2005; Santee & Egeth, 1982a; van Ede et al., 2012). However, an outstanding question is: why did the visual modality excel in the recognition of second language letters as opposed to the tactile modality that did not show any difference in the recognition of first and second language letters? One plausible cause is the visual complexity arising from the orthographic structure of letters that likely differs between the two languages. As mentioned earlier in this work, there are 50 characters in the Bangla alphabet set, and among this huge amount of complex characters some are very similar in shape (Rahman et al., 2019; Tabassum et al., 2020). On the contrary, there are only 26 characters in the English alphabet set that are structurally much easier as compared with Bangla letters (see Figure 2 for examples). Moreover, Bangla letters are typically composed of strokes, lines, angles, and curves that possess a higher level of nonlinear visual complexity as compared with English letters (see Figure 1), which might put the English letters at an advantage for encoding, processing, and retention, and subsequently for recognition and recall. Thus, due to the increased shape complexity of Bangla letters children might find them more difficult to recognize as compared with English letters. This proposition receives the support from prior studies that investigated the influence of orthographic features and visual complexity on letter naming and word reading tasks (Asaad & Eviatar, 2013; Rao et al., 2011). For example, one study on native Arabic speakers demonstrated that orthographic complexity influenced letter naming performance and automaticity in Arabic (Asaad & Eviatar, 2013). A second study examined the accuracy and speed of single-word reading in Urdu and Hindi in Urdu-Hindi adult bilinguals, and revealed that responses to single Urdu words were more error-prone and consistently slower than to single Hindi words despite the fact that Urdu was the participants' native language and the language in which most of their schooling took place (Rao et al., 2011). The authors suggested that this difference was due to the fact that Urdu is orthographically deeper and visually more complex than Hindi. Taking these prior findings together with the present, it can be concluded that letter recognition is influenced more by shape complexity than by differential experience individuals have with the first and second language. However, the present study failed to show a similar difference in the tactile modality because the tactile modality might not be as efficient as the visual modality to discern structural differences and complexity between Bangla and English letters. As outlined above and earlier in this work, letter recognition depends on the shape or macrospatial analysis for which vision is more efficient than touch (see Bushnell & Boudreau, 1991; Klatzky et al., 1989; Lederman et al., 1986; Mahar et al., 1994; Tokita & Ishiguchi, 2016). Thus, despite the fact that Bangla is the mother tongue of the children participating in this study, the effect of orthographic structure and visual complexity on letter recognition superseded the effect of nativeness in Bangla.

Character Type

As demonstrated by the analysis of aggregate data in repeated measures ANOVAs, irrespective of task modality and language, digits were recognized, in support of the third hypothesis, more accurately as well as more speedily as compared with letters (Figure 5, panels b and f). This finding is in agreement with the classification accuracies observed in various confusion matrices (Tables 1–4; Figures A1–A4, panels d, in Appendix). This finding affirms the existing proposition that, in general, digits are processed differently from letters (James et al., 2005; Polk et al., 2002), and suggests a new proposition that digits are recognized/classified more accurately and more fluently than letters. In support of the former proposition, studies in healthy adults have suggested that there is a dissociation between letter and digit (or number) processing in the visual modality (e.g., Carreiras et al., 2015; Park et al., 2012). This notion of processing independence is also highly consistent with the character type–task modality interaction effect that was significant on character recognition time but not on character recognition per se. As demonstrated for the former (Figure 5, panel c), children took, on average, much longer time to recognize a letter (25.364 s) than a digit (20.841 s) in the tactile modality, whereas the average time taken to recognize them in the same modality was slightly longer for a letter (2.814 s) than for a digit (2.111 s).

Now, an outstanding question is: why are digits recognized more accurately and more fluently than letters? One plausible cause is the semantic advantage digits likely have over letters in both Bangla and English. Letters and digits, though similar in many respects, also differ in potentially significant ways. Individual digits (e.g., 8, 9) are logographic symbols, while individual letters (e.g., L, N) are elements of an alphabetic writing system and roughly correspond to phonemes (Besner & Coltheart, 1979; McCloskey & Schubert, 2014). Thus, individual digits represent entire words and convey conceptual meanings, while individual letters typically convey meanings only when combined meaningfully (except single-letter words; Deloche & Seron, 1987; Holender & Peereman, 1987; McCloskey & Schubert, 2014). For example, “9” in English or “৯” in Bangla indicates a semantic quantity, while “N” in English or “ন” in Bangla does not have any semantic value. A second cause, irrespective of language, might be the orthographic structures of these symbols that differ largely, with digits having much shallower and simpler physical features of nonlinearity than letters (Ingles & Eskes, 2008; see Figures 1 and 2 for such letters and digits). Thus, as compared with a letter, a digit is less likely to produce confusion and suspense about its spatial configuration. A third and related cause might be the smaller number of relatively dissimilar digit characters (10 in both Bangla and English) as compared with the larger number of letter characters (50 in Bangla and 26 in English) with a greater intercharacter similarity (see Islalm et al., 2019; Venkataramanan et al., 2021). For example, in Bangla alphabets, “ধ” is someway similar to “ব,” “গ” is similar to “স,” “ন” is similar to “ণ,” and “ং” is similar to “ৎ.” Similarly, in English alphabets, there is more or less similarity between “B” and “P,” between “L” and “I,” and between “N” and “M.” More interestingly, “N” is similar to “Z” in way that they are 90° rotated versions of each other, and similarly, “6” and “9” are 90° rotated versions of each other. These examples give further insight into the interclass confusions that are more likely to occur by shape similarity than by random guessing. Therefore, a larger portion of confusions occurred with letters outside the stimulus set than those inside the stimulus set (Tables 1 and 2; Figures A1 and A2, panels d, in Appendix). A fourth cause might be the acoustic confusions arising from the phonological similarity of letters (Morgan, 1973), for instance, “N” that sounds “en” might be confused with “M” that sounds “em,” “দ” that sounds “doa” might be confused with “ধ” that sounds “dhoa,” and “গ” that sounds “goa” might be confused with “ঘ” that sounds “ghoa.” Because such a phonological similarity does not exist between digits in Bangla or English digits are less likely to produce confusion and uncertainty. Taken together, it can be argued that the semantic quality, orthographic simplicity, structural/featural dissimilarity, and phonological dissimilarity of digits together might facilitate digit discriminability and recognition accuracy, and speed up the process of digit discrimination or recognition.

Within-Modal Performance Variation: An Insight Into the Role of Character Shape

An inspection of various confusion matrices generated in this study shows that the classification accuracy varied not only across sensory modalities but across linguistic characters within each sensory modality as well (Tables 1–4). This variation might be due to the fact that some letter/digit pairs seldom got confused and other letter/digit pairs got confused quite frequently. Irrespective of task modality, the most frequently confused Bangla letter pairs within the stimulus set were “ক” versus “ফ,” “ং ” versus “ ঃ ” and “খ” versus “গ” (Table 1), and the most frequently confused Bangla digit pairs were “৫” versus “৬,” “৭” versus “৮,” “১” versus “৯”(Table 3). The most frequently confused English letters within the stimulus set were “Q” versus “O,” “X” versus “Y” and “C” versus “O” (Table 2), and the most frequently confused English digit pairs were “6” versus “9,” “1” versus “7,” “3” versus “8,” “2” versus “4,” and “5” versus “6” (Table 4). Most of these confusions are likely to be associated with perceptual similarity (with the exception of “৭” versus “৮” and “2” versus “4” confusions in the tactile modality), and might not simply be the result of random guessing (Liu et al., 2009, Min & Bennett, 2020). Researchers believe that perceptual similarity and confusion between stimuli are determined by their physical properties (e.g., Liu et al., 2009). Here, it can be theorized that the linguistic characters that are spatially similar to other linguistic characters undergo Gestalt organizational process such that individuals may miss or ignore, as explained above, some surface-level important features like strokes, dots, and gaps or openings, and tend to report “Q” as “O,” “7” as “1,” “৯” as “১,” and “ফ” as “ক.” This missing or ignorance is likely to operate, as outlined above, more in the tactile than the visual modality. Prior studies demonstrated the influence of Gestalt grouping on participant’s ability to ignore distracting/irrelevant tactile information while sensing and perceiving various stimuli through touch (Chang et al., 2007; Frings & Spence, 2013; Komura et al., 2021; Overvliet et al., 2012; Prieto et al., 2014). Based on the current findings it can be suggested that following the Gestalt principles of perceptual grouping individuals may ignore not only the irrelevant information but the relevant information as well. This ignorance leads to the increased perceptual similarity between various stimuli and results in confusion or misclassification as exemplified above. However, this proposition is not universal and does not necessarily apply to other exceptional forms of confusions or misclassifications, such as “৭” versus “৮” and “2” versus “4” that occurred as outlined above, for unknown reasons, in the tactile modality.

Finally, the relative confusability might also vary between the members of a character pair. For example, in Bangla letter classification using visual modality, “ফ” was reported as “ক” by 3% of children while “ক” was reported as “ফ” by none, and in doing the same task using tactile modality, “ফ” was reported as “ক” by 73% of children, while “ক” was reported as “ফ” by 17% of children (Table 1). In English letter classification using visual modality, “Q” was reported as “O” by 13% of children while “O” was reported as “Q” by none, and in doing the same task using tactile modality, “Q” was reported as “O” by 83% of children while “O” was reported as “Q” by none (Table 2). Such pair-wise asymmetries can be exemplified also for Bangla or English digit classification in both sensory modalities (see Tables 3 and 4). One plausible cause of these pair-wise asymmetries might be, as explained above, the tendency to ignore some relevant features of a certain character in favor of the other one. A second cause might be response bias, a tendency to report one character more often than the other one, due to the unequal legibility of the tested characters (Loomis, 1982; Townsend, 1971).

To summarize, taking all the above findings together, it can be concluded that the complex linguistic characters, such as letters or digits, acquired mostly through vision can be recognized or classified by touch alone, with the recognition or classification performance being greater in the (trained) visual than the (untrained) tactile modality. The character recognition or classification performance can further be modulated by character type, with digits being more accurately and more speedily recognized than letters, but not by language per se. However, language can interact with task modality to mediate character recognition or classification performance to a certain degree. The errors in linguistic character recognition or classification are likely caused by the characters each other. It is suggested that the semantic quality, orthographic structure, and other characteristic features of linguistic characters, as well as experience can play crucial roles in shaping both sensory and cross-sensory performance.

Conclusion

This study was the first to investigate sensory and cross-sensory performance in shape recognition, using ecologically valid novel sets of stimuli—the letters and digits in Bangla and English, in a unique population of young children who, after the acquisition of those intricate characters in natural settings, are now learning to read and write in both languages that have alphabets (and digits) of distinct orthographic structures. Based on the outputs of inferential statistical analyses and confusion matrices a few novel and interesting findings are reported. One important finding is that children were able to recognize or classify the (mostly) visually acquired letters and digits, in both Bangla and English, not only by vision but by touch as well, indicating an early transfer of learned skills from vision to touch. This transfer is explained by two theoretical principles (see above) which state that the human brain has an innate or acquired capacity to encode abstract representations amodally (e.g., Erdogan et al., 2015; Karim et al., 2021a; Yildirim & Jacobs, 2013, 2015). A second important finding is that the letter or digit recognition performance was strikingly greater in the visual than the tactile modality, with a completely reversed pattern of recognition time. This can be explained by the inherently increased efficiency of the visual modality over the tactile modality in shape analysis/macrospatial skills (see Bushnell & Boudreau, 1991; Klatzky et al., 1989; Lederman et al., 1986; Wismeijer et al., 2012), or by experience the visual modality receives through multimodal training on letters and digits during childhood. A third important finding is that language alone did not affect letter or digit recognition per se; however, it was found to interact with task modality to mediate both letter and digit recognition capacities. Specifically, the visual modality performed better in the recognition of both second-language letters and second-language digits, whereas the tactile modality performed better in the recognition of first-language digits but not first-language letters. The superior performance of the visual modality in the recognition of second-language letters as compared with first-language letters is explained by the orthographic simplicity of the former over the latter (see Asaad & Eviatar, 2013; Rao et al., 2011). The visual superiority in the second-language digit processing and tactile superiority in the first-language digit processing is explained by the orthographic simplicity of digits over letters albeit there was a lack of visual superiority in the first-language digit processing and a lack of tactile superiority in the second-language digit processing. A fourth important finding is that irrespective of task modality and language, digits were recognized more accurately and more speedily as compared with letters. This finding is explained by the orthographic simplicity and better semantic quality of digits over letters (see Ingles & Eskes, 2008; McCloskey & Schubert, 2014). Finally, the errors in recognition or classification of those linguistic characters are likely caused by the aforementioned physical attributes of the characters each other. Altogether, these findings advance understanding of the crossmodal nature of perceptual and cognitive development, particularly language and numerical development, in children. Thus, this study presents findings that are ecologically highly significant, and also represents an exciting avenue for future research with the practical implications of crossmodality and neuroplasticity for psycholinguistic development in children.

Constraints on Generality

Crossmodal learning transfer is not culture or individual-specific. It is instead a general neurocognitive phenomenon that likely operates in every typically developed human in every culture. Therefore, the transfer of visual learning to tactile modality reported in this study should be reproducible with children of the same age sampled from regions outside Dhaka/Bangladesh, as well as with children of the same age sampled from any other cultures. This generalization may not be confined to linguistic characters but can expand to other stimuli as well. However, the other findings are generalizable upon the satisfaction of certain conditions. The visual superiority observed in this study may decrease in children if they are exposed to the same linguistic characters both visually and tactilely during the course of character learning. The second language advantage in character recognition/classification as compared with the first language can be generalized to children of other cultures as long as they have second language characters of simpler physical (e.g., orthographic) features as compared with first language characters. Similarly, better performance in digit recognition/classification as compared with letter recognition/classification can be generalized to children of other cultures as long as they have digits of simpler physical (e.g., orthographic) features compared with letters. The confusion or misclassification rate might increase or decrease depending on the physical (e.g., orthographic) similarity or dissimilarity of the characters. The linguistic characters used in this study can be used in future replication attempts on children of other cultures only if those children have mastered or can be taught to master the characters. All aspects of the study method and analysis codes except the order in which the tactile and visual phases were presented should be preserved in a direct replication attempt. This study did not counterbalance between the two sensory phases because in this particular study, exposure to the visual modality was likely to carry a practice effect onto the subsequent performance in the tactile modality but not vice versa. However, this is not directly generalizable to a situation where an experience in the tactile modality is likely to affect the subsequent performance in the visual modality and vice versa. Thus the results of this study can be expanded to broader populations under certain conditions. The constraints on generality discussed here would encourage future replication attempts to test the boundary conditions of the findings reported in this study.

Context of the Study

Past studies demonstrated that crossmodal transfer, a process of conveying information from one sensory mode to another, occurs before 1 year of age (Gottfried et al., 1977; Meltzoff & Borton, 1979; Rose et al., 1981a, 1981b, 1983; Rose & Orlian, 1991; Ruff & Kohler, 1978; Streri, 1987; Streri & Gentaz, 2004), and sufficiently improves between 5 and 8 years of age (Birch & Lefford, 1963; Picard, 2007; Stoltz-Loike & Bornstein, 1987). Moreover, studies in newborns showed that crossmodal transferability develops without the benefit of months of experience in simultaneous visual–tactual exploration (e.g., Meltzoff & Borton, 1979). On the contrary, one study demonstrated that multisensory integration, a process of combining and integrating information from simultaneously stimulated multiple sensory modes, does not typically occur before 8 years of age (Gori et al., 2008). This indicates that crossmodal transfer and crossmodal integration might be two independent processes. The former one might be easier and operate earlier than the latter one. To gain further insight into this idea the present study examined whether the acquired shape recognition/classification skills transfer from a trained sensory mode to an untrained sensory mode. Using a letter or digit recognition/classification task this study showed that 5–6-year-old children who learned those shapes mostly visually were able to recognize/classify them, in both Bangla and English, by touch. This finding makes an important contribution to the recent theory of crossmodal plasticity (a theory put forward by the first author of this work and other colleagues; Karim et al., 2021a), suggesting that crossmodal transferability of ecologically valid shapes develops, in the first few years of life, much earlier than the development of crossmodal integration capacity.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alais, D., Cass, J., & Chapouthier, G. (2010). Multisensory perceptual learning of temporal order: Audiovisual learning transfers to vision but not audition. *PLoS ONE*, *5*(6), Article e11283. https://doi.org/10.1371/journal.pone.0011283

Amedi, A., Raz, N., Pianka, P., Malach, R., & Zohary, E. (2003). Early "visual" cortex activation correlates with superior verbal memory performance in the blind. *Nature Neuroscience*, *6*(7), 758–766. https://doi.org/10.1038/nn1072

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*(6), 687–689. https://doi.org/10.1038/nn1912

Asaad, H., & Eviatar, Z. (2013). The effects of orthographic complexity and diglossia on letter naming in Arabic: A developmental study. *Writing Systems Research*, *5*(2), 156–168. https://doi.org/10.1080/17586801.2013.862163

Ashton, J. E., Jefferies, E., & Gaskell, M. G. (2018). A role for consolidation in cross-modal category learning. *Neuropsychologia*, *108*, 50–60. https://doi.org/10.1016/j.neuropsychologia.2017.11.010

Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with 'the vOICe'. *Perception*, *36*(3), 416–430. https://doi.org/10.1068/p5631

Berger, C. C., & Ehrsson, H. H. (2018). Mental imagery induces cross-modal sensory plasticity and changes future auditory perception. *Psychological Science*, *29*(6), 926–935. https://doi.org/10.1177/0956797617748959

Bernstein, L. E., Auer, E. T., Eberhardt, S. P., & Jiang, J. (2013). Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Frontiers in Neuroscience*, *7*, Article 34. https://doi.org/10.3389/fnins.2013.00034

Bernstein, L. E., Eberhardt, S. P., & Auer, E. T. (2014). Audiovisual spoken word training can promote or impede auditory-only perceptual learning: Results from prelingually deafened adults with late-acquired cochlear implants versus normal-hearing adults. *Frontiers in Psychology*, *5*, Article 934. https://doi.org/10.3389/fpsyg.2014.00934

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, *17*(5), 467–472. https://doi.org/10.1016/0028-3932(79)90053-8

Birch, H. G., & Belmont, L. (1965). Auditory-visual integration, intelligence and reading ability in school children. *Perceptual and Motor Skills*, *20*(1), 295–305. https://doi.org/10.2466/pms.1965.20.1.295

Birch, H. G., & Lefford, A. (1963). Intersensory development in children. *Monographs of the Society for Research in Child Development*, *28*(5), 1–48. https://doi.org/10.2307/1165681

Björkman, M. (1967). Relations between intra-modal and cross-modal matching. *Scandinavian Journal of Psychology*, *8*(1), 65–76. https://doi.org/10.1111/j.1467-9450.1967.tb01375.x

Blank, M., Altman, L. D., & Bridger, W. H. (1968). Crossmodal transfer of form discrimination in preschool children. *Psychonomic Science*, *10*(2), 51–52. https://doi.org/10.3758/BF03331402

Bliss, I., & Hämäläinen, H. (2005). Different working memory capacity in normal young adults for visual and tactile letter recognition task. *Scandinavian Journal of Psychology*, *46*(3), 247–251. https://doi.org/10.1111/j.1467-9450.2005.00454.x

Blomert, L., & Froyen, D. (2010). Multi-sensory learning and learning to read. *International Journal of Psychophysiology*, *77*(3), 195–204. https://doi.org/10.1016/j.ijpsycho.2010.06.025

Brawn, T. P., Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2008). Consolidation of sensorimotor learning during sleep. *Learning and Memory*, *15*(11), 815–819. https://doi.org/10.1101/lm.1180908

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5), 554–564. https://doi.org/10.1167/6.5.2

Bushnell, E. W., & Boudreau, J. P. (1991). The development of haptic perception during infancy. In M. A. Heller & W. Schiff (Eds.), *The psychology of touch* (pp. 139–161). Lawrence Erlbaum.

Butler, A. C., Karpicke, J. D., & Roediger IIIH. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273–281. https://doi.org/10.1037/1076-898X.13.4.273

Cantlon, J. F., Brannon, E. M., Carter, E. J., Pelphrey, K. A., & Dehaene, S. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*(5), Article e125. https://doi.org/10.1371/journal.pbio.0040125

Carmel, D., & Carrasco, M. (2008). Perceptual learning and dynamic changes in primary visual cortex. *Neuron*, *57*(6), 799–801. https://doi.org/10.1016/j.neuron.2008.03.009

Carreiras, M., Quiñones, I., Hernández-Cabrera, J. A., & Duñabeitia, J. A. (2015). Orthographic coding: Brain activation for letters, symbols, and digits. *Cerebral Cortex*, *25*(12), 4748–4760. https://doi.org/10.1093/cercor/bhu163

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & Rosario, H. D. (2020). *pwr:*

Basic functions for power analysis. R package version 1.3-0 [Computer software]. https://cran.r-project.org/web/packages/pwr/pwr.pdf

Chang, D., Nesbitt, K. V., & Wilkins, K. (2007). The Gestalt principles of similarity and proximity apply to both the haptic and visual grouping of elements. In *AUIC '07: Proceedings of the eighth Australasian conference on user interface* (Vol. *64*, pp. 79–86).

Cho, Y., Craig, J. C., Hsiao, S. S., & Bensmaia, S. (2016). Vision is superior to touch in shape perception even with equivalent peripheral input. *Journal of Neurophysiology*, *115*(1), 92–99. https://doi.org/10.1152/jn.00654.2015

Choo, L. B., Lin, D. T. A., & Pandian, A. (2012). Language learning approaches: A review of research on explicit and implicit learning in vocabulary acquisition. *Procedia Social and Behavioral Sciences*, *55*, 852–860. https://doi.org/10.1016/j.sbspro.2012.09.572

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum.

Cuppone, A. V., Semprini, M., & Konczak, J. (2018). Consolidation of human somatosensory memory during motor learning. *Behavioural Brain Research*, *347*, 184–192. https://doi.org/10.1016/j.bbr.2018.03.013

Danhauer, J. L., & Lucks, L. E. (1987). The confusion matrix: A new model. *Human Communication Canada/Communication Humaine Canada*, *11*(4), 7–11. https://cjslpa.ca/files/1987_HumComm_Vol_11/No_04_1-54/Danhauer_Lucks_HumComm_1987.pdf

Deloche, G., & Seron, X. (1987). Numerical transcoding: A general production model. In G. Deloche & X. Seron (Eds.), *Mathematical disabilities: A cognitive neuropsychological perspective* (pp. 137–170). Lawrence Erlbaum.

Dopjans, L., Wallraven, C., & Bulthoff, H. H. (2009). Cross-modal transfer in visual and haptic face recognition. *IEEE Transactions on Haptics*, *2*(4), 236–240. https://doi.org/10.1109/TOH.2009.18

Easton, R. D., Greene, A. J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin and Review*, *4*, 403–410. https://doi.org/10.3758/BF03210801

Easton, R. D., Srinivas, K., & Greene, A. J. (1997). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 153–163. https://doi.org/10.1037//0278-7393.23.1.153

Eberhardt, S. P., Auer, Jr., E. T., & Bernstein, L. E. (2014). Multisensory training can promote or impede visual perceptual learning of speech stimuli: Visual-tactile vs. visual-auditory training. *Frontiers in Human Neuroscience*, *8*, Article 829. https://doi.org/10.3389/fnhum.2014.00829

Ellis, N. C. (2016). Implicit and explicit knowledge about language. In J. Cenoz, D. Gorter, & S. May (Eds.), *Language awareness and multilingualism. Encyclopedia of language and education* (3rd ed., pp. 1–12). Springer. https://doi.org/10.1007/978-3-319-02325-0_7-1

Enns, J. T., & Girgus, J. S. (1985). Developmental changes in selective and integrative visual attention. *Journal of Experimental Child Psychology*, *40*(2), 687–690. https://doi.org/10.1016/0022-0965(85)90093-1

Erdogan, G., Yildirim, I., Jacobs, R. A., & Schrater, P. (2015). From sensory signals to modality-independent conceptual representations: A probabilistic language of thought approach. *PLoS Computational Biology*, *11*(11), Article e1004610. https://doi.org/10.1371/journal.pcbi.1004610

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ettlinger, G., & Wilson, W. A. (1990). Cross-modal performance: Behavioural processes, phylogenetic considerations and neural mechanisms. *Behavioural Brain Research*, *40*(3), 169–192. https://doi.org/10.1016/0166-4328(90)90075-P

Fan, F. Y. (2017). *Poweranalysis: Power analysis in experimental design* (R Package Version 0.2) [Computer software]. https://cran.r-project.org/web/packages/powerAnalysis/powerAnalysis.pdf

Fernández, L. M., & Vadillo, M. A. (2020). Flexibility in reaction time analysis: Many roads to a false positive? *Royal Society Open Science*, *7*(2), Article 190831. https://doi.org/10.1098/rsos.190831

Fischler, I. (1975). Detection and identification of words and letters in simulated visual search of word lists. *Memory and Cognition*, *3*(2), 175–182. https://doi.org/10.3758/BF03212895

Fiset, D., Blais, C., Arguin, M., Tadros, K., Éthier-Majcher, C., Bub, D. N., & Gosselin, F. (2009). The spatio-temporal dynamics of visual letter recognition. *Cognitive Neuropsychology*, *26*(1), 23–35. https://doi.org/10.1080/02643290802421160

Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, *18*(2), 129–155. https://doi.org/10.1007/s11145-004-5892-2

Fraga González, G., Žarić, G., Tijms, J., Bonte, M., & Van der Molen, M. W. (2017). Contributions of letter-speech sound learning and visual print tuning to reading improvement: Evidence from brain potential and dyslexia training studies. *Brain Sciences*, *7*(12), Article 10. https://doi.org/10.3390/brainsci7010010

Fridland, E. (2021). Skill and strategic control. *Synthese*, *199*(3–4), 5937–5964. https://doi.org/10.1007/s11229-021-03053-3

Friedes, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory and deficit. *Psychological Bulletin*, *81*(5), 284–310. https://doi.org/10.1037/h0036331

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326–340. https://doi.org/10.1037/dev0000075

Frings, C., & Spence, C. (2013). Gestalt grouping effects on tactile information processing: When touching hands override spatial proximity. *Attention, Perception, and Psychophysics*, *75*(3), 468–480. https://doi.org/10.3758/s13414-012-0417-6

Geyer, L. H., & DeWald, C. G. (1973). Feature lists and confusion matrices. *Perception and Psychophysics*, *14*(3), 471–482. https://doi.org/10.3758/BF03211185

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is the neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton-Mifflin.

Gori, M., Del Viva, M., Sandini, G., & Burr, D. (2008). Young children do not integrate visual and haptic form information. *Current Biology*, *18*(9), 694–698. https://doi.org/10.1016/j.cub.2008.04.036

Gottfried, A. W., Rose, S. A., & Bridger, W. H. (1977). Cross-modal transfer in human infants. *Child Development*, *48*(1), 118–123. https://doi.org/10.2307/1128889

Hashimoto, R., & Sakai, K. L. (2004). Learning letters in adulthood: Direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, *42*(2), 311–322. https://doi.org/10.1016/S0896-6273(04)00196-5

Hatwell, Y., Orliaguet, J. P., & Brouty, G. (1990). Effects of objects properties, attentional constraints and manual exploratory procedures on haptic perceptual organization: A developmental study. In H. Bloch & B. I. Bertenthal (Eds.), *Sensory-motor organizations and development in infancy and early childhood* (pp. 315–335). Springer.

Hayashi, Y., Schmidt, A. C., & Saunders, K. J. (2013). Effects of letter-identification training on letter naming in prereading children. *Journal of Applied Behavior Analysis*, *46*(4), 838–843. https://doi.org/10.1002/jaba.90

Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., & Sinha, P. (2011). The newly sighted fail to match seen with felt. *Nature Neuroscience*, *14*(5), 551–553. https://doi.org/10.1038/nn.2795

Hilbert, S., Schwaighofer, M., Zech, A., Sarubin, N., Arendasy, M., & Bühner, M. (2017). Working memory tasks train working memory but not reasoning: A material- and operation-specific investigation of transfer from working memory practice. *Intelligence*, *61*, 102–114. https://doi.org/10.1016/j.intell.2017.01.010

Himi, S. A. (2018). *Understanding cognitive structure of multitasking behavior and working memory training effects* [Doctoral Dissertation, Ludwig-Maximilians-Universität München]. Electronic thesis of LMU Munich. https://edoc.ub.uni-muenchen.de/23189/1/Himi_Samsad_Afrin.pdf

Himi, S. A., Bühner, M., Schwaighofer, M., Klapetek, A., & Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition*, *189*, 275–298. https://doi.org/10.1016/j.cognition.2019.04.010

Holender, D., & Peereman, R. (1987). Differential processing of phonographic and logographic single-digit numbers by the two hemispheres. In G. Deloche & X. Seron (Eds.), *Mathematical disabilities: A cognitive neuropsychological perspective* (pp. 43–86). Lawrence Erlbaum.

Howe, M. L., Courage, M. L., Vernescu, R., & Hunt, M. (2000). Distinctiveness effects in children's long-term retention. *Developmental Psychology*, *36*(6), 778–792. https://doi.org/10.1037//0012-1649.36.6.778

Ingles, J. L., & Eskes, G. A. (2008). A comparison of letter and digit processing in letter-by-letter reading. *Journal of the International Neuropsychological Society*, *14*(1), 164–173. https://doi.org/10.1017/S1355617708080119

Islalm, M. S., Rahman, M. M., Rahman, M. H., Arifuzzaman, M., Sassi, R., & Aktaruzzaman, M. (2019, September). Recognition bangla sign language using convolutional neural network. In *2019 international conference on innovation and intelligence for informatics, computing, and technologies (3ICT)* (pp. 1–6). IEEE.

James, K. H., James, T. W., Jobard, G., Wong, A.-N., & Gauthier, I. (2005). Letter processing in the visual system: Different activation patterns for single letters and strings. *Cognitive, Affective, and Behavioral Neuroscience*, *5*(4), 452–466. https://doi.org/10.3758/CABN.5.4.452

Jones, B., & Connolly, K. (1970). Memory effects in cross-modal matching. *British Journal of Psychology*, *61*(2), 267–270. https://doi.org/10.1111/j.2044-8295.1970.tb01243.x

Jones, J. P. (1972). *Intersensory transfer, perceptual shifting, modal preferences and reading*. International Reading Association.

Jones, T. C., & Roediger III, H. L. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology*, *7*(1), 65–80. https://doi.org/10.1080/09541449508520158

Juola, J. F., Leavitt, D. D., & Choe, C. S. (1974). Letter identification in word, nonword, and single-letter displays. *Bulletin of the Psychonomic Society*, *4*(4), 278–280. https://doi.org/10.3758/BF03336725

Karim, A. K. M. R., Proulx, M. J., de Sousa, A. A., & Likova, L. T. (2021a). Do we enjoy what we sense and perceive? A dissociation between aesthetic appreciation and basic perception of environmental objects or events. *Cognitive, Affective, and Behavioral Neuroscience*, *22*, 904–951. https://doi.org/10.3758/s13415-022-01004-0

Karim, A. K. M. R., Proulx, M. J., de Sousa, A. A., & Likova, L. T. (2021b). Neuroplasticity and crossmodal connectivity in the normal, healthy brain. *Psychology and Neuroscience*, *14*(3), 298–334. https://doi.org/10.1037/pne0000258

Karipidis, I. I., Pleisch, G., Brandeis, D., Roth, A., Röthlisberger, M., Schneebeli, M., Walitza, S., & Brem, S. (2018). Simulating reading acquisition: The link between reading outcome and multimodal brain signatures of letter–speech sound learning in prereaders. *Scientific Reports*, *8*(1), Article 7121. https://doi.org/10.1038/s41598-018-24909-8

Kelly, A. C., & Garavan, H. (2005). Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex*, *15*(8), 1089–1102. https://doi.org/10.1093/cercor/bhi005

Kim, J. K., & Zatorre, R. J. (2008). Generalized learning of visual-to-auditory substitution in sighted individuals. *Brain Research*, *1242*, 263–275. https://doi.org/10.1016/j.brainres.2008.06.038

Kim, J. K., & Zatorre, R. J. (2010). Can you hear shapes you touch? *Experimental Brain Research*, *202*(4), 747–754. https://doi.org/10.1007/s00221-010-2178-6

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*(1), 24–38. https://doi.org/10.1037/0033-2909.112.1.24

Klatzky, R. L., Lederman, S. J., & Reed, C. (1989). Haptic integration of object properties: Texture, hardness, and planar contour. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 45–57. https://doi.org/10.1037//0096-1523.15.1.45

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. The Guilford Press.

Komura, H., Nakamura, T., & Ohka, M. (2021). Investigation of tactile illusion based on gestalt theory. *Philosophies*, *6*(3), Article 60. https://doi.org/10.3390/philosophies6030060

Kovács, I. (1996). Gestalten of today: Early processing of visual contours and surfaces. *Behavioural Brain Research*, *82*(1), 1–11. https://doi.org/10.1016/S0166-4328(97)81103-5

Laughery, K. R. (1970). Computer simulation of shortterm memory: A component decay model. In G. T. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 3, pp. 135–200). Academic Press. https://doi.org/10.1016/S0079-7421(08)60399-0

Lederman, S. J., Thorne, G., & Jones, B. (1986). Perception of texture by vision and touch: Multidimensionality and intersensory integration. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(2), 169–180. https://doi.org/10.1037//0096-1523.12.2.169

Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading: XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, *41*(6), 725–743. https://doi.org/10.1016/S0042-6989(00)00295-9

Levy-Tzedek, S., Novick, I., Arbel, R., Abboud, S., Maidenbaum, S., Vaadia, E., & Amedi, A. (2012). Cross-sensory transfer of sensory-motor information: Visuomotor learning affects performance on an audiomotor task, using sensory-substitution. *Scientific Reports*, *2*(1), Article 949. https://doi.org/10.1038/srep00949

Lewis, T. L., Ellemberg, D., Maurer, D., Dirks, M., Wilkinson, F., & Wilson, H. R. (2004). A window on the normal development of sensitivity to global form in glass patterns. *Perception*, *33*(4), 409–418. https://doi.org/10.1068/p5189

Liu, L., Klein, S. A., Xue, F., Zhang, J. Y., & Yu, C. (2009). Using geometric moments to explain human letter recognition near the acuity limit. *Journal of Vision*, *9*(1), Article 26. https://doi.org/10.1167/9.1.26

Lochya, A., Jacquesa, C., Maillarde, L., Colnat-Coulboisg, S., Rossiona, B., & Jonas, J. (2018). Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *PNAS*, *115*(32), E7595–E7604. https://doi.org/10.1073/pnas.1718987115

Loomis, J. M. (1982). Analysis of tactile and visual confusion matrices. *Perception and Psychophysics*, *31*(1), 41–52. https://doi.org/10.3758/BF03206199

Ludwig, V. U., Adachi, I., & Matzuzawa, 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*, *108*(51), 20661–20665. https://doi.org/10.1073/pnas.1112605108

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, *289*(5482), 1206–1208. https://doi.org/10.1126/science.289.5482.1206

MacDonald, J. A., Balakrishnan, J. D., Orosz, M. D., & Karplus, W. J. (2002). Intelligibility of speech in a virtual 3-D environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(2), 272–286. https://doi.org/10.1518/0018720024497934

Madec, S., Goff, K. L., Riès, S. K., Legou, T., Rousselet, G., Courrieu, P., Alario, F.-X., Grainger, J., & Rey, A. (2016). The time course of visual influences in letter recognition. *Cognitive, Affective, and Behavioral Neuroscience*, *16*(3), 406–414. https://doi.org/10.3758/s13415-015-0400-5

Mahar, D., Mackenzie, B., & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception*, *23*(11), 1369–1386. https://doi.org/10.1068/p231369

Mate, J., & Baqués, J. (2009). Short article: Visual similarity at encoding and retrieval in an item recognition task. *Quarterly Journal of Experimental Psychology*, *62*(7), 1277–1284. https://doi.org/10.1080/17470210802680769

Maurer, D., Stager, C. L., & Mondloch, C. J. (1999). Cross-modal transfer of shape is difficult to demonstrate in one-month-olds. *Child Development*, *70*(5), 1047–1057. https://doi.org/10.1111/1467-8624.00077

McCloskey, M., & Schubert, T. (2014). Shared versus separate processes for letter and digit identification. *Cognitive Neuropsychology*, *31*(5–6), 437–460. https://doi.org/10.1080/02643294.2013.869202

McGovern, D. P., Astle, A. T., Clavin, S. L., & Newell, F. N. (2016). Task-specific transfer of perceptual learning across sensory modalities. *Current Biology*, *26*(1), R20–R21. https://doi.org/10.1016/j.cub.2015.11.048

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, *282*(5737), 403–404. https://doi.org/10.1038/282403a0

Merkley, R., Conrad, B., Price, G., & Ansari, D. (2019). Investigating the visual number form area: A replication study. *Royal Society Open Science*, *6*(10), Article 182067. https://doi.org/10.1098/rsos.182067

Min, Y., & Bennett, M. (2020). *Causal explanations of image misclassifications*. arXiv preprint arXiv:2006.15739.

Morgan, B. J. T. (1973). Cluster analyses of two acoustic confusion matrices. *Perception and Psychophysics*, *13*(1), 13–24. https://doi.org/10.3758/BF03207229

Mueller, S. T., & Weidemannb, C. T. (2012). Alphabetic letter identification: Effects of perceivability, similarity, and bias. *Acta Psychologica*, *139*(1), 19–37. https://doi.org/10.1016/j.actpsy.2011.09.014

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2005). Grabbing your ear: Rapid auditory–somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral Cortex*, *15*(7), 963–974. https://doi.org/10.1093/cercor/bhh197

Myers, T., & Szücs, D. (2015). Arithmetic memory is modality specific. *PLoS ONE*, *10*(12), Article e0145614. https://doi.org/10.1371/journal.pone.0145614

Nanay, B. (2018). Multimodal mental imagery. *Cortex*, *105*, 125–134. https://doi.org/10.1016/j.cortex.2017.07.006

Naumer, M. J., Doehrmann, O., Müller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical plasticity of audio-visual object representations. *Cerebral Cortex*, *19*(7), 1641–1653. https://doi.org/10.1093/cercor/bhn200

Navon, D. (1977). Forest before the tree: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Nayar, K., Franchak, J., Adolph, K., & Kiorpes, L. (2015). From local to global processing: The development of illusory contour perception. *Journal of Experimental Child Psychology*, *131*, 38–55. https://doi.org/10.1016/j.jecp.2014.11.001

Neiworth, J. J., Gleichman, A. J., Olinick, A. S., & Lamp, K. E. (2006). Global and local processing in adult humans (*Homo sapiens*), 5-year-old children (*Homo sapiens*), and adult cotton-top tamarins (*Saguinus oedipus*). *Journal of Comparative Psychology*, *120*(4), 323–330. https://doi.org/10.1037/0735-7036.120.4.323

Norman, L. J., & Thaler, L. (2021). Perceptual constancy with a novel sensory skill. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(2), 269–281. https://doi.org/10.1037/xhp0000888

Overvliet, K. E., Krampe, R. T., & Wagemans, J. (2012). Perceptual grouping in haptic search: The influence of proximity, similarity, and good continuation. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 817–821. https://doi.org/10.1037/a0029222

Pal, U., Roy, K., & Kimura, F. (2009). A lexicon-driven handwritten city-name recognition scheme for Indian postal automation. *IEICE Transactions on Information and Systems*, *E92-D*(5), 1146–1158. https://doi.org/10.1587/transinf.E92.D.1146

Paraskevopoulos, E., & Herholz, S. C. (2013). Multisensory integration and neuroplasticity in the human cerebral cortex. *Translational Neuroscience*, *4*(3), 337–348. https://doi.org/10.2478/s13380-013-0134-1

Park, J., Hebrank, A., Polk, T. A., & Park, D. C. (2012). Neural dissociation of number from letter recognition and its relationship to parietal numerical processing. *Journal of Cognitive Neuroscience*, *24*(1), 39–50. https://doi.org/10.1162/jocn_a_00085

Pêcheux, M.-G., Lepecq, J.-C., & Salzarulo, P. (1988). Oral activity and exploration in 1- to 2-month-old infants. *British Journal of Developmental Psychology*, *6*(3), 245–256. https://doi.org/10.1111/j.2044-835X.1988.tb01098.x

Perea, M., Jiménez, M., Martín-Suesta, M., & Gómez, P. (2015). Letter position coding across modalities: Braille and sighted reading of sentences with jumbled words. *Psychonomic Bulletin and Review*, *22*(2), 531–536. https://doi.org/10.3758/s13423-014-0680-8

Peters, L., De Smedt, B., & Op de Beeck, H. P. (2015). The neural representation of Arabic digits in visual cortex. *Frontiers in Human Neuroscience*, *9*, Article 517. https://doi.org/10.3389/fnhum.2015.00517

Picard, D. (2007). Tactual, visual, and cross-modal transfer of texture in 5- and 8-year-old children. *Perception*, *36*(5), 722–736. https://doi.org/10.1068/p5575

Pick, H. L. Jr. (1974). Visual coding of non-visual spatial information. In R. B. MacLeod & H. L. Pick (Eds.), *Perception: Essays in honor of James J. Gibson* (pp. 153–165). Cornell University Press.

Pilling, M., & Thomas, S. (2011). Audiovisual cues and perceptual learning of spectrally distorted speech. *Language and Speech*, *54*(4), 487–497. https://doi.org/10.1177/0023830911404958

Poldrack, R. A. (2000). Imaging brain plasticity: Conceptual and methodological issues—A theoretical review. *NeuroImage*, *12*(1), 1–13. https://doi.org/10.1006/nimg.2000.0596

Polk, T. A., Stallcup, M., Aguirre, G. K., Alsop, D. C., D'Esposito, M., Detre, J. A., & Farah, M. J. (2002). Neural specialization for letter recognition. *Journal of Cognitive Neuroscience*, *14*(2), 145–159. https://doi.org/10.1162/089892902317236803

Prieto, A., Mayas, J., & Ballesteros, S. (2014). Time course of grouping by proximity and similarity in a haptic speeded orientation task. In *International conference on human haptic sensing and touch enabled computer applications* (pp. 379–385). Springer.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*, *134*(1), 73–92. https://doi.org/10.1037/0096-3445.134.1.73

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2017). Learning and retrieving holistic and componential visual–verbal associations in reading and object naming. *Neuropsychologia*, *98*, 68–84. https://doi.org/10.1016/j.neuropsychologia.2016.09.025

Rabby, A. S. A., Haque, S., Islam, S., Abujar, S., & Hossain, S. A. (2018). Bornonet: Bangla handwritten characters recognition using convolutional neural network. *Procedia Computer Science*, *143*, 528–535. https://doi.org/10.1016/j.procs.2018.10.426

Rączy, K., Czarnecka, M., Paplińska, M., Hesselmann, G., Knops, A., & Szwed, M. (2020). Tactile to visual number priming in the left intraparietal cortex of sighted braille readers. *Scientific Reports*, *10*(1), Article 17571. https://doi.org/10.1038/s41598-020-72431-7

Rączy, K., Czarnecka, M., Zaremba, D., Izdebska, K., Paplińska, M., Hesselmann, G., Knops, A., & Szwed, M. (2020). A shared code for Braille and Arabic digits revealed by cross-modal priming in sighted Braille readers. *Acta Psychologica*, *202*, Article 102960. https://doi.org/10.1016/j.actpsy.2019.102960

Rahman, M., Mahmud, I., Uddin, M. P., Afjal, M. I., Habib, A., & Kabir, F. (2019). Handwritten Bengali character recognition through geometry based feature extraction. *Journal of Theoretical and Applied Information Technology*, *97*(23), 3570–3582. https://www.researchgate.net/publication/337945937_Handwritten_BengaliCharacter_Recognition_through_Geometry_Based_Feature_Extraction

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, *28*(2), 617–625. https://doi.org/10.1016/S0896-6273(00)00138-0

Rao, C., Vaid, J., Srinivasan, N., & Chen, H. C. (2011). Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. *Reading and Writing*, *24*(6), 679–695. https://doi.org/10.1007/s11145-010-9256-9

R Development Core Team. (2015). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Reales, J. M., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 644–663. https://doi.org/10.1037/0278-7393.25.3.644

Rock, I., & Harris, C. S. (1967). Vision and touch. *Scientific American*, *216*(5), 96–104. https://doi.org/10.1038/scientificamerican0567-96

Rose, S. A. (1984). Developmental changes in hemispheric specialization for tactual processing in very young children: Evidence from cross-modal transfer. *Developmental Psychology*, *20*(4), 568–574. https://doi.org/10.1037/0012-1649.20.4.568

Rose, S. A., Feldman, J. F., Futterweit, L. R., & Jankowski, J. J. (1998). Continuity in tactual–visual cross-modal transfer: Infancy to 11 years. *Developmental Psychology*, *34*(3), 435–440. https://doi.org/10.1037/0012-1649.34.3.435

Rose, S. A., Gottfried, A. W., & Bridger, W. H. (1981a). Cross-modal transfer and information processing by the sense of touch in infancy. *Developmental Psychology*, *17*(1), 90–98. https://doi.org/10.1037/0012-1649.17.1.90

Rose, S. A., Gottfried, A. W., & Bridger, W. H. (1981b). Cross-modal transfer in 6-month-old infants. *Developmental Psychology*, *17*(5), 661–669. https://doi.org/10.1037/0012-1649.17.5.661

Rose, S. A., Gottfried, A. W., & Bridger, W. H. (1983). Infant's cross-modal transfer from solid objects to their graphic representations. *Child Development*, *54*(3), 686–694. https://doi.org/10.2307/1130056

Rose, S. A., & Orlian, E. K. (1991). Asymmetries in infant cross-modal transfer. *Child Development*, *62*(4), 706–718. https://doi.org/10.2307/1131172

Rubinstein, L., & Gruenberg, E. M. (1971). Intramodal and crossmodal sensory transfer of visual and auditory temporal patterns. *Perception and Psychophysics*, *9*(5), 385–390. https://doi.org/10.3758/BF03210235

Ruff, H. A., & Kohler, C. J. (1978). Tactual–visual transfer in six-month-old infants. *Infant Behavior and Development*, *1*, 259–264. https://doi.org/10.1016/S0163-6383(78)80037-X

Ruffino, C., Truong, C., Dupont, W., Bouguila, F., Michel, C., Lebon, F., & Papaxanthis, C. (2021). Acquisition and consolidation processes following motor imagery practice. *Scientific Reports*, *11*(1), Article 2295. https://doi.org/10.1038/s41598-021-81994-y

Santee, J. L., & Egeth, H. E. (1980). Interference in letter identification: A test of feature-specific inhibition. *Perception and Psychophysics*, *27*(4), 321–330. https://doi.org/10.3758/BF03206121

Santee, J. L., & Egeth, H. E. (1982a). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance*, *8*(4), 489–501. https://doi.org/10.1037//0096-1523.8.4.489

Santee, J. L., & Egeth, H. E. (1982b). Independence versus interference in the perceptual processing of letters. *Perception and Psychophysics*, *31*(2), 101–116. https://doi.org/10.3758/BF03206210

Santens, S., Roggeman, C., Fias, W., & Verguts, T. (2010). Number processing pathways in human parietal cortex. *Cerebral Cortex*, *20*(1), 77–88. https://doi.org/10.1093/cercor/bhp080

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, *11*(1), 53–60. https://doi.org/10.1038/nrn2737

Schmidt, T. T., & Blanenburg, F. (2019). The somatotopy of mental tactile imagery. *Frontiers in Human Neuroscience*, *13*, Article 10. https://doi.org/10.3389/fnhum.2019.00010

Shams, L., Wozny, D., Kim, R. S., & Seitz, A. R. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*, *2*, Article 264. https://doi.org/10.3389/fpsyg.2011.00264

Shepherd, G. M. (2012). *Neurogastronomy: How the brain creates flavor and why it matters*. Columbia University Press.

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M., Rangarajan, V., Winawer, J., Miller, K. J., & Parvizi, J. (2013). A brain area for visual numerals. *Journal of Neuroscience*, *33*(16), 6709–6715. https://doi.org/10.1523/JNEUROSCI.4558-12.2013

Silverston, R. A., & Deichmann, J. W. (1975). Sense modality research and the acquisition of reading skills. *Review of Educational Research*, *45*(1), 149–172. https://doi.org/10.3102/00346543045001149

Skocaj, D., Leonardis, A., & Kruijff, G-J. M. (2012). Cross-modal learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 861–864). Springer.

Speed, L. J., Croijmans, I., Dolscheid, S., & Majid, A. (2021). Crossmodal associations with olfactory, auditory, and tactile stimuli in children and adults. *i-Perception*, *12*(6), 1–34. https://doi.org/10.1177/20416695211048513

Spence, C., & Deroy, O. (2013). Crossmodal mental imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory imagery* (pp. 157–183). Springer.

Spence, C., Senkowski, D., & Röder, B. (2009). Crossmodal processing. *Experimental Brain Research*, *198*(2–3), 107–111. https://doi.org/10.1007/s00221-009-1973-4

Staller, J. D., & Lappin, J. S. (1979). Word and nonword superiority effects in a letter detection task. *Perception and Psychophysics*, *25*(1), 47–54. https://doi.org/10.3758/BF03206109

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255–266. https://doi.org/10.1038/nrn2331

Stoltz-Loike, M., & Bornstein, M. H. (1987). The roles of imagery, language, and metamemory in cross-modal transfer in children. *Psychological Research*, *49*(1), 63–68. https://doi.org/10.1007/BF00309204

Stoycheva, P., Kauramäki, J., Newell, F. N., & Tiippana, K. (2021). Haptic recognition memory and lateralisation for verbal and nonverbal shapes. *Memory*, *29*(8), 1043–1057. https://doi.org/10.1080/09658211.2021.1957938

Streri, A. (1987). Tactile discrimination of shape and intermodal transfer in 2- to 3-month-old infants. *British Journal of Developmental Psychology*, *5*(3), 213–220. https://doi.org/10.1111/j.2044-835X.1987.tb01056.x

Streri, A., & Gentaz, E. (2004). Cross-modal recognition of shape from hand to eyes and handedness in human newborns. *Neuropsychologia*, *42*(10), 1365–1369. https://doi.org/10.1016/j.neuropsychologia.2004.02.012

Streri, A., & Molina, M. (1993). Visual–tactual and tactual–visual transfer between objects and pictures in 2-month-old infants. *Perception*, *22*(11), 1299–1318. https://doi.org/10.1068/p221299

Streri, A., & Pecheux, M. G. (1986). Tactual habituation and discrimination of form in infancy: A comparison with vision. *Child Development*, *57*(1), 100–104. https://doi.org/10.2307/1130641

Suzuki, M., & Gyoba, J. (2008). Visual and tactile cross-modal mere exposure effects. *Cognition and Emotion*, *22*(1), 147–154. https://doi.org/10.1080/02699930701298382

Tabassum, T., Mahmud, I., Uddin, M. P., Ali, E., Afjal, M. I., & Nitu, A. M. (2020). Enhancement of single-handed Bengali sign language recognition based on hog features. *Journal of Theoretical and Applied Information Technology*, *98*(5), 743–756. http://www.jatit.org/volumes/Vol98No5/2Vol98No5.pdf

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., Reiman, E. M., & Liu, Y. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academy of Sciences*, *103*(28), 10775–10780. https://doi.org/10.1073/pnas.0604416103

Taylor, J. S. H., Rastle, K., & Davis, M. H. (2014). Distinct neural specializations for learning to read words and name objects. *Journal of Cognitive Neuroscience*, *26*(9), 2128–2154. https://doi.org/10.1162/jocn_a_00614

Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences*, *103*(21), 8239–8244. https://doi.org/10.1073/pnas.0509704103

Tokita, M., & Ishiguchi, A. (2016). Precision and bias in approximate numerical judgment in auditory, tactile, and cross-modal presentation. *Perception*, *45*(1–2), 56–70. https://doi.org/10.1177/0301006615596888

Tounsi, M., Moalla, I., Lebourgeois, F., & Alimi, A. M. (2018). *Multilingual scene character recognition system using sparse auto-encoder for efficient local features representation in bag of features*. arXiv preprint arXiv:1806.07374.

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception and Psychophysics*, *9(1)*, 40–50. https://doi.org/10.3758/BF03213026

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271–282. https://doi.org/10.1016/j.neuron.2004.06.025

Van der Heijden, A. H. C., Wolters, G., Fleur, E., & Hommels, J. G. M. (1992). Single-letter recognition accuracy benefits and position information. *Bulletin of the Psychonomic Society*, *30*(2), 101–104. https://doi.org/10.3758/BF03330409

van Ede, F., de Lange, F. P., & Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *Journal of Neuroscience*, *32*(30), 10408–10412. https://doi.org/10.1523/JNEUROSCI.1337-12.2012

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). Dividing attention between multiple targets: Simultaneous or sequential allocation? *Journal of Vision*, *7*(9), 642–642. https://doi.org/10.1167/7.9.642

Venkataramanan, A., Laviale, M., Figus, C., Usseglio-Polatera, P., & Pradalier, C. (2021, September). Tackling inter-class similarity and intra-class variance for microscopic image-based classification. In *International conference on computer vision systems* (pp. 93–103). Springer.

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung*, *18*(1), 299–342. https://doi.org/10.1007/BF02409636

Wallraven, C., Bülthoff, H. H., Waterkamp, S., van Dam, L., & Gaißert, N. (2014). The eyes grasp, the hands see: Metric category knowledge transfers between vision and touch. *Psychonomic Bulletin and Review*, *21*(4), 976–985. https://doi.org/10.3758/s13423-013-0563-4

Wismeijer, D. A., Gegenfurtner, K. R., & Drewing, K. (2012). Learning from vision-to-touch is different than learning from touch-to-vision. *Frontiers in Integrative Neuroscience*, *6*, Article 105. https://doi.org/10.3389/fnint.2012.00105

Witelson, S. F. (1974). Hemispheric specialization for linguistic and nonlinguistic tactual perception using a dichotomous stimulation technique. *Cortex*, *10*(1), 3–17. https://doi.org/10.1016/S0010-9452(74)80034-1

Wood, L. E. (1977). Modality-specific coding in matching letters. *Perception and Psychophysics*, *21*(4), 329–335. https://doi.org/10.3758/BF03199482

World Medical Association. (2013). World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Xerri, C., & Zennou-Azogui, Y. (2021). Interplay between primary cortical areas and crossmodal plasticity. In T. Heinbockel & Y. Zhou (Eds.), *Connectivity and functional specialization in the brain* (pp. 1–174). IntechOpen. https://doi.org/10.5772/intechopen.95450

Xu, W., Kolozsvari, O. B., Monto, S. P., & Hämäläinen, J. A. (2018). Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Frontiers in Human Neuroscience*, *12*, Article 304. https://doi.org/10.3389/fnhum.2018.00304

Xu, W., Kolozsvari, O. B., Oostenveld, R., & Hämäläinen, J. A. (2020). Rapid changes in brain activity during learning of grapheme-phoneme associations in adults. *NeuroImage*, *220*, Article 117058. https://doi.org/10.1016/j.neuroimage.2020.117058

Yildirim, I., & Jacobs, R. A. (2013). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition*, *126*(2), 135–148. https://doi.org/10.1016/j.cognition.2012.08.005

Yildirim, I., & Jacobs, R. A. (2015). Learning multisensory representations for auditory-visual transfer of sequence category knowledge: A probabilistic language of thought approach. *Psychonomic Bulletin and Review*, *22*(3), 673–686. https://doi.org/10.3758/s13423-014-0734-y

Zhang, J., Wermter, S., Sun, F., Zhang, C., Engel, A. K., Röder, B., Fu, X., & Xue, G. (2022). Cross-modal learning: Adaptivity, prediction and interaction. *Frontiers in Neurorobotics*, *62*, Article 889911. https://doi.org/10.3389/fnbot.2022.889911

Zhang, Z., & Yuan, K.-H. (2018). *Practical statistical power analysis using webpower and R*. ISDSA Press.

(*Appendix follows*)

Appendix

Table A1

Descriptive Statistics for Eight Behavioral Measures After Arcsine or Log Transformations of the Data in Experiment 1

Measures	*M*	*SD*	Skewness	Kurtosis
	Arcsine recognition score			
Tactilely presented English letter	0.823	0.233	0.487	−0.778
Tactilely presented Bangla letter	0.819	0.212	−0.469	−0.309
Visually presented English letter	1.508	0.146	−2.053	2.702
Visually presented Bangla letter	1.396	0.245	−0.905	−0.804
	Log reaction time			
Tactilely presented English letter	1.302	0.292	−0.438	−0.456
Tactilely presented Bangla letter	1.359	0.254	−0.903	0.108
Visually presented English letter	0.357	0.196	1.457	1.768
Visually presented Bangla letter	0.445	0.185	0.545	−1.018

Table A2

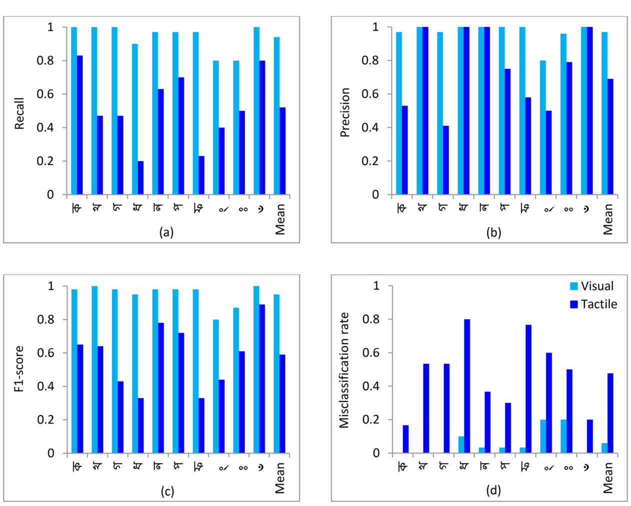
Descriptive Statistics for Eight Behavioral Measures After Arcsine or Log Transformations of the Data in Experiment 2

Measures	*M*	*SD*	Skewness	Kurtosis
	Arcsine recognition score			
Tactilely presented English digit	0.908	0.237	0.926	3.222
Tactilely presented Bangla digit	0.994	0.206	−0.676	−0.027
Visually presented English digit	1.571	0.00 [a]	—	—
Visually presented Bangla digit	1.503	0.157	−2.018	2.436
	Log reaction time			
Tactilely presented English digit	1.257	0.218	−1.047	1.197
Tactilely presented Bangla digit	1.308	0.163	−0.936	0.419
Visually presented English digit	0.3	0.192	1.089	1.272
Visually presented Bangla digit	0.254	0.182	2.048	3.573

[a] All 30 values were the same after arcsine transformations, and therefore no skewness and kurtosis values were produced.

Figure A1

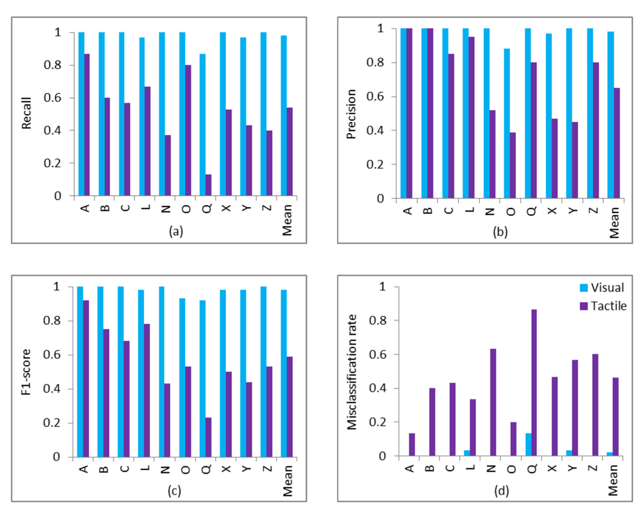
Metrics for Classification Performance and Misclassification Rate Calculated From Confusion Matrices for Bangla Letters Presented in Visual and Tactile Modalities



Note. See the online article for the color version of this figure.

Figure A2

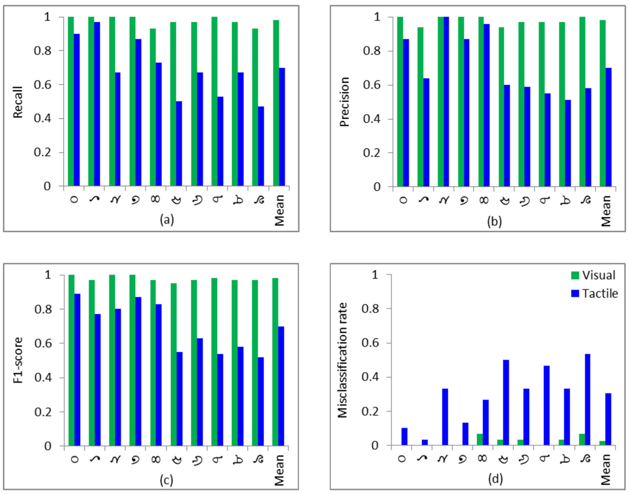
Metrics for Classification Performance and Misclassification Rate Calculated From Confusion Matrices for English Letters Presented in Visual and Tactile Modalities



Note. See the online article for the color version of this figure.

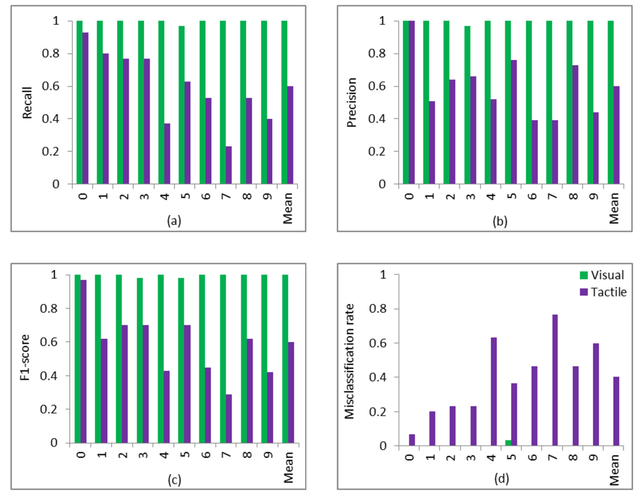
(*Appendix continues*)

Figure A3
Metrics for Classification Performance and Misclassification Rate Calculated From Confusion Matrices for Bangla Digits Presented in Visual and Tactile Modalities



Note. See the online article for the color version of this figure.

Figure A4
Metrics for Classification Performance and Misclassification Rate Calculated From Confusion Matrices for English Digits Presented in Visual and Tactile Modalities



Note. See the online article for the color version of this figure.

Received October 3, 2021
Revision received January 21, 2023
Accepted January 23, 2023 ■