Bridging ears and eyes when learning spoken words: On the effects of bilingual experience at 30 months

Havy, Mélanie; Zesiger, Pascal Eric

How to cite

HAVY, Mélanie, ZESIGER, Pascal Eric. Bridging ears and eyes when learning spoken words: On the effects of bilingual experience at 30 months. In: Developmental science, 2020, vol. 24, n° 1. doi: 10.1111/desc.13002

This publication URL: https://archive-ouverte.unige.ch/unige:165488

Publication DOI: 10.1111/desc.13002

Received: 1 October 2019 | Revised: 15 May 2020 | Accepted: 15 May 2020
DOI: 10.1111/desc.13002

PAPER

Developmental Science WILEY

Bridging ears and eyes when learning spoken words: On the effects of bilingual experience at 30 months

Mélanie Havy | **Pascal E. Zesiger**

Faculty of Psychology and Educational Sciences, Geneva University, Geneva, Switzerland

Correspondence
Mélanie Havy, FAPSE, Université de Genève, Boulevard du Pont d'Arve 28, 1211 Genève 4, Switzerland.
Email: melanie.havy@gmail.com

Funding information
This work was funded by a SNSF grant (100014_159402) to MH and PZ.

Abstract

From the very first moments of their lives, infants selectively attend to the visible orofacial movements of their social partners and apply their exquisite speech perception skills to the service of lexical learning. Here we explore how early bilingual experience modulates children's ability to use visible speech as they form new lexical representations. Using a cross-modal word-learning task, bilingual children aged 30 months were tested on their ability to learn new lexical mappings in either the auditory or the visual modality. Lexical recognition was assessed either in the same modality as the one used at learning ('same modality' condition: auditory test after auditory learning, visual test after visual learning) or in the other modality ('cross-modality' condition: visual test after auditory learning, auditory test after visual learning). The results revealed that like their monolingual peers, bilingual children successfully learn new words in either the auditory or the visual modality and show cross-modal recognition of words following auditory learning. Interestingly, as opposed to monolinguals, they also demonstrate cross-modal recognition of words upon visual learning. Collectively, these findings indicate a bilingual edge in visual word learning, expressed in the capacity to form a recoverable cross-modal representation of visually learned words.

KEYWORDS
audio-visual speech perception, bilingualism, language development, sensory format of lexical representations, word learning

1 | INTRODUCTION

Children raised in a bilingual household from birth (simultaneous bilinguals) are growing in number. Dual language experience represents a unique challenge, as children have to contend with two languages, while receiving minimal exposure and extent variation for each individual input (Bijeljac-Babic, 2018). One source of information that alleviates perceptual uncertainties is seeing the visible orofacial movements accompanying speech. During social interactions, seeing talking faces provides information that is highly redundant and complementary to the auditory signal (Miller & Nicely, 1955). In monolingual infants and adults, it enhances phonetic perception (Samuel & Lieblich, 2014; Ter Schure, Junge, & Boersma, 2016) and assists word recognition in both normal and adverse listening conditions (Buchwald, Winters, & Pisoni, 2009; Havy, Foroud, Fais, & Werker, 2017). In simultaneous bilingual populations, converging evidence documents an early contribution of visible speech to phonetic perception (Pons, Bosch, & Lewkowicz, 2015; Soto-Faraco et al., 2007; Weikum et al., 2013). Yet, it is unknown whether monolingual and bilingual children differ in their use of visible speech as they establish new lexical representations. Here, we ask to what extent early bilingual experience modulates 30-month-old children's attention to visible speech as they build their nascent lexicon.

In the literature, there is tremendous evidence that irrespective of the language environment (monolingual vs. bilingual), infants are natively endowed with basic perceptual skills to apprehend the multisensory nature of speech. From the very first moments of their lives, infants preferentially orient their attention toward talking faces (Schonberg, Sandhofer, Tsang, & Johnson, 2014) and selectively attend to the orofacial movements of their social partners (Hunnius & Geuze, 2004). Infants can link auditory and visual speech events across a variety of temporal (Baart, Vroomen, Shaw, & Bortfeld, 2014; Lewkowicz & Pons, 2013) and spectral dimensions (Kuhl, Williams, & Meltzoff, 1991). For instance, when viewing two side-by-side displays of two identical faces silently uttering two distinct sounds, or sound sequences, 2- to 12-month-old infants look longer at the video matching the concurrently heard sound(s) (Danielson, Bruderer, Kandhadai, Vatikiotis-Bateson, & Werker, 2017; Lewkowicz, Minar, Tift, & Brandon, 2015; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009; Streri, Coulon, Marie, & Yeung, 2016). Infants build on these sensitivities to learn the phonology of their native language. For instance, they can identify the language in use just from watching someone silently talking as early as 4 months (Kubicek et al., 2014; Lewkowicz & Pons, 2013; Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; Weikum et al., 2007) and can distinguish certain complex auditory phonetic contrasts by simultaneously following the corresponding orofacial information at just 6 months (Kuhl, Tsao, & Liu, 2003; Teinonen, Aslin, Alku, & Csibra, 2008; Ter Schure et al., 2016). By 8 months, their babbling productions are dominated by the most visible segments (Rvachew & Alhaidary, 2018).

As they advance in age and gain language experience, these capacities become increasingly refined. Yet, they develop differently as a function of the ambient language environment. In this field, a great deal of research has revealed that early bilingual exposure fine-tunes infants' attention to the visible aspects of speech. First, studies have found that, presumably to compensate for the challenge of learning two languages and tease them apart, bilingual infants reliably start to attend to the redundant visible speech cues inherent in a speaker's mouth at an earlier age (by 4 months and perhaps earlier) than their monolingual counterparts (between 4 and 8 months; Pons et al., 2015). Studies have also revealed that despite an increased attentiveness to the mouth of talking faces in monolinguals during critical periods of phonological (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012) and lexical development (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2018), 8- to 18-month-old bilingual infants demonstrate a greater interest in orofacial cues than their peers. This bilingual edge is observed in linguistic communicative situations (Birulés, Bosch, Brieke, Pons, & Lewkowicz, 2018; Pons et al., 2015) but also more extensively in non-linguistic situations (i.e., faces portraying different emotional states such as crying or laughing, Ayneto & Sebastián-Gallés, 2017; non-speech movements involving eyebrow raise or lip protrusion, Fort, Ayneto-Gimeno, Escrichs, & Sebastian-Galles, 2018), suggesting a broad-based attention bias. Converging evidence has bolstered this view, by showing that infants growing up with two languages are more proficient at discerning languages visually. Indeed, although monolingual and bilingual infants are equally able to discriminate their native language from an unfamiliar one at 4 to 6 months, only bilingual infants retain the capacity to do so with languages that are unfamiliar to them at 8 months (Sebastián-Gallés et al., 2012; Weikum et al., 2007). Together, these observations suggest that in the early stages of language acquisition, infants raised in bilingual homes may be more sensitive to the visible features of speech than monolingual infants. Yet, it is unknown if these early differences influence the way they appreciate visible speech information as they learn their first words.

Research Highlights

- Like monolingual children, bilingual children aged 30 months successfully learn new word-object mappings when listening to auditory instantiations of the words.
- Like monolingual children, bilingual children successfully learn new word-object mappings when watching the articulatory instantiations of the spoken words in silence.
- Like monolingual children, bilingual children can visually recognize words previously experienced in the auditory modality.
- As opposed to monolinguals, bilingual children can auditorily recognize words previously experienced in the visual modality.

In the literature, the visual recognition of spoken words has been well documented in monolingual children and empirical evidence has revealed that visible speech is stored in lexical representations early in development. This is supported by studies showing that monolingual children are sensitive to mispronunciations of familiar words in both the auditory and visual modalities as early as 1 year (Weatherhead & White, 2017) and can use visible speech to foster lexical recognition in normal (Davies, Kidd, & Lander, 2009) and adverse listening conditions by 4–14 years (Grieco-Calub & Olson, 2015; Lalonde & Holt, 2015; Ross et al., 2011). Yet, extant evidence has also revealed that the use of visible speech in a lexical context is not always straightforward. First, in many instances, words differ by phonemes that share visual articulatory characteristics and are virtually indistinguishable from each other on the sole basis of visible speech, a phenomenon called 'homopheny' that represents half of the words in English (i.e., 'path', 'bath', Berger, 1977). Second, there is evidence that attending to the lexical content of a spoken word is demanding in young children and can easily compromise the visual processing of the subtle phonological detail of the word form. For instance, Jerger, Damian, Tye-Murray, and Abdi (2014), Jerger, Damian, Tye-Murray, and Abdi (2017) found that the presence of visible speech in adverse listening condition improved phoneme identification to a greater extent for pseudowords than for words in 4-year-old monolingual children (Jerger, Damian, Tye-Murray, & Abdi, 2014, 2017).

To date, all research effort has gone into describing the capacities of monolingual children with no empirical consideration for bilingual children. In monolinguals, much of the attention was focused on determining whether visible speech is contained in early lexical representations and only one research project led by Havy et al. (2017; Havy & Zesiger, 2017) has attempted to explore how visible speech becomes part of new lexical representations (Havy & Zesiger, 2017; Havy et al., 2017). In this work, Havy et al. (2017; Havy & Zesiger, 2017 focused on 18- to 30-month-old monolingual children and examined two individual mechanisms. (a) First, they asked whether visible speech could be stored directly from the visually available information alone. (b) Second, they asked whether the information from either modality could be encoded indirectly through cross-modal translation of the input. In this work, children were taught new lexical mappings (Object A: 'byp' and Object B: 'var') in either the auditory (acoustic form of the word with no accompanying face) or the visual modality (talking face with no sound), and thereafter tested for recognition ('Look at the "var!"') in either the same modality as at learning ('same modality' condition) or in the other modality ('cross-modality' condition). The pseudo-words were maximally distinct in both the auditory and visual modalities and were easily discriminable at 18 and 30 months (Pons et al., 2009; Streri et al., 2016; Werker & Curtin, 2005; Yeung & Werker, 2013).

The results from the 'same modality' condition revealed that children's word learning capacities are subject to an initial auditory dominance expressed by the capacity to learn from auditory but not visual input at 18 months. This dominance wanes as children advance in age and language experience, and by 30 months, children reliably learn from visual input. Of interest, the results from the 'cross-modality' condition indicate that despite difficulties in learning in the visual modality, 18-month-old monolingual infants show visual recognition of auditorily-learned words. This pattern suggests that although monolingual infants primarily attend to the auditory speech signal to guide lexical learning, they can incorporate visible speech in their representations through cross-modal translation of the auditory input. At 30 months, monolingual children more reliably learn lexical mappings from visible speech. Yet unlike adults, they do not show auditory recognition upon visual learning. This indicates that by 30 months, learning from visual input is still demanding and that only auditory experience engenders a recoverable cross-modal representation of the words.

Following on Havy and Zesiger (2017), the purpose of the current study is to determine whether 30-month-old bilingual children are faced with a similar challenge of forming cross-modal representations of words following visual learning. We reason that bilingually raised children may be better placed than their monolingual peers as they demonstrate stronger preferences for the mouth in the early stages of language acquisition (Ayneto & Sebastián-Gallés, 2017; Pons et al., 2015) and better use of visible speech in language discrimination tasks (Sebastián-Gallés et al., 2012; Weikum et al., 2007). These proclivities may benefit visual word learning, especially given the literature showing that a larger number of fixations to the speaker's mouth is associated with greater visual recognition of spoken words in monolingual adults (Marassa & Lansing, 1995) and greater expressive vocabulary size in 6- to 24-month-old monolingual (Altvater-Mackensen, Mani, & Grossmann, 2016; Tenenbaum, Sobel, Sheinkopf, Malle, & Morgan, 2015) and bilingual infants (Tsang, Atagi, & Johnson, 2018).

Bilingual experience also imparts advantages in a wide array of executive functions that are relevant to visual word learning and cross-modal recognition. For instance, bilingual infants and children have been found to have greater working memory retrieval than their monolingual peers (6 months–7 years: Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Singh et al., 2015), a capacity that happens to be associated with greater auditory (16–20 months: Vlach & Johnson, 2013) and visual spoken word recognition in monolingual infants and children (7–14 years: Heikkilä, Lonka, Ahola, Meronen, & Tiippana, 2017; Tye-Murray, Hale, Spehar, Myerson, & Sommers, 2014). Bilingual infants and children also demonstrate greater attention shifting (7 months–7 years: Byers-Heinlein, Morin-Lessard, & Lew-Williams, 2017; Crivello et al., 2016) and memory generalization across contents (18–24 months: Bialystok, 2017; Brito, Sebastián-Gallés, & Barr, 2015), two core capacities that develop along with frequent language switching and that are foundational to cross-modal recognition.

Bridging these series of evidence together, the above literature suggests a bilingual edge in a variety of capacities that are relevant to visual word learning and cross-modal recognition. Yet, alternative evidence exists that paints a more nuanced picture. First, in the field of audio–visual speech perception, the greater bilingual interest for the mouth has not been consistently found in infancy and early childhood (14 months–6 years: Birulés et al., 2018; Morin-Lessard, Poulin-Dubois, Segalowitz, & Byers-Heinlein, 2019). Importantly, the reported preferences have been documented with tasks that involved watching speakers silently uttering sentences, without distinction between the pre-lexical (no consideration for the associated meaning) and lexical mechanisms (consideration for the associated meaning) that are possibly involved (Ronquest, Levi, & Pisoni, 2010). It is therefore unclear whether the observed proclivities pertain to pre-lexical processes more than lexical ones. Additionally, there is evidence that although mouth gazing is developmentally linked to lexical proficiency, this relationship is not stronger in bilingual infants (6–12 months: Tsang et al., 2018), toddlers and preschoolers (2–5 years: Morin-Lessard et al., 2019) relative to monolinguals. Finally, one adult study revealed a monolingual advantage in the use of visible speech during non-native phonological discrimination (Burfin et al., 2014).

Taking the issue at the lexical level, the literature has also yielded a mixed narrative. In this field, the contribution of visible speech to the lexical domain has never been explored in bilinguals. Yet, bilingual lexical achievement has been well documented in the auditory modality, with empirical evidence revealing that the sensitivity of bilinguals to the phonological detail of auditory word forms is sometimes comparable (Fennell & Byers-Heinlein, 2014; Ramon-Casas, Fennell,

TABLE 1 Demographic information including the participant's identification (ID), the age (months; days), the gender (male, female), the MCDI estimation of the expressive vocabulary size and the socioeconomic status (SES) as well as the mean and standard deviation (*SD*) for each group

	Age	Gender	MCDI	Type of language		Percent of exposure			
ID	Months; days	Male, female	Production	L2	L3	French	L2	L3	SES scores
Auditory learning group									
M (*SD*)	30;19 (0.26)	9 Males	434 (151)			52.59 (16.71)	43.57 (16.59)	3.84 (4.58)	23.22 (6.96)
1	30;00	Female	Unknown	Dutch	None	70	30	0	24
2	30;00	Female	129	Spanish	None	70	30	0	11
3	31;03	Female	657	Italian	English	60	30	10	23
s4	31;09	Female	335	Romanian	None	60	40	0	18
5	31;09	Male	325	Romanian	None	60	40	0	18
6	30;02	Male	532	Italian	None	68	32	0	23
7	31;06	Female	463	Spanish	Portuguese	50	42	8	26
8	30;02	Male	Unknown	Arabic	None	80	20	0	26
9	31;29	Female	496	English	Italian	36	54,8	9,2	18
10	31;01	Female	343	Italian	Greek	29	62,8	8,2	22
11	29;05	Male	Unknown	English	Greek	40	50	10	26
12	30;03	Female	Unknown	English	Spanish	30	60	10	22
13	30;01	Female	542	Spanish	None	40	60	0	45
14	30;00	Male	396	English	Turkish	60	30	10	25
15	29;28	Female	567	English	None	20	80	0	26
16	30;22	Male	563	Wolof	None	26	74	0	16
17	32;30	Male	Unknown	Turkish	None	64	36	0	28
18	28;25	Male	480	Spanish	None	59,5	40,5	0	Unknown
19	30;12	Male	531	Spanish	Danish	69	25	6	Unknown
20	29;26	Female	70	Spanish	Italian	69	21	10	Unknown
21	29;10	Female	473	German	None	40	60	0	Unknown
22	31;10	Female	511	Russian	None	52	48	0	Unknown
23	30;24	Female	405	German	Italian	57	36	7	21
Visual learning group									
M (*SD*)	30;11 (0.28)	12 Males	447 (197)			55.32 (19.63)	43.04 (18.75)	1.68 (3.56)	21.73 (4.21)
1	31;23	Female	552	German	None	69	31	0	28
2	29;08	Female	225	Japanese	None	57	43	0	17
3	29;11	Female	470	Italian	None	72	28	0	24

(Continues)

TABLE 1 (Continued)

	Age	Gender	MCDI	Type of language		Percent of exposure			
ID	Months; days	Male, female	Production	L2	L3	French	L2	L3	SES scores
4	29;10	Female	496	Spanish	None	60	40	0	24
5	31;28	Male	534	Polish	English	52	40	8	24
6	31;28	Female	897	Arabic	None	63	37	0	18
7	30;21	Male	210	English	Polish	70	20	10	26
8	31;15	Female	349	Italian	None	60	40	0	24
9	30;07	Female	425	Armenian	None	80	20	0	22
10	28;30	Female	352	Spanish	None	80	20	0	20
11	29;30	Male	329	Spanish	None	80	20	0	22
12	30;14	Female	159	Cantonese	None	80	20	0	14
13	30;14	Male	346	Spanish	None	67	33	0	18
14	30;30	Male	293	Spanish	None	33	67	0	15
15	29;29	Male	509	Spanish	None	40	60	0	Unknown
16	29;13	Male	177	Spanish	Italian	30	62	8	13
17	29,26	Male	206	English	Polish	24	71	5	26
18	31;25	Female	458	English	None	33	67	0	22
19	31;02	Male	387	Russian	None	43	57	0	23
20	31;08	Female	693	Portuguese	None	22	79	0	23
21	30;06	Male	894	Spanish	None	45	55	0	23
22	29;22	Female	526	German	Spanish	69	31	0	26
23	29;22	Female	528	German	Spanish	55	45	0	26
24	29;24	Male	633	English	Croatian	23	66	11	Unknown
25	30;01	Male	526	English	None	76	24	0	Unknown

& Bosch, 2017), inferior (Ramon-Casas, Swingley, Sebastián-Gallés, & Bosch, 2009; Wewalaarachchi, Wong, & Singh, 2017) or superior (Singh, Fu, Tay, & Golinkoff, 2018) to that of monolinguals. However, it should be noted that any difference reported in the timeline of acquisition is small and moderated by many elements, including the relative similarity of the phonological inventories of the two languages and language exposure history, with greater outcome in children learning languages with similar acoustic realization of the phonological contrasts (Havy, Bouchon, & Nazzi, 2016), and in children tested on their dominant language with an input that matches their language learning environment (Fennell & Byers-Heinlein, 2014). Critically, in the sole study that reported a developmental advance in encoding the segmental detail of novel words, the authors used a vowel contrast that was highly salient by access to visual cues (Singh et al., 2018). Although the study only presented the words auditorily, the authors argued that phonemic contrasts that are visually salient may be preferentially available on the account of prior experience with visible speech. Taken as such, children raised in bilingual environment may have more robust visual representations of words than their monolingual peers.

Overall, the literature on audio–visual speech perception and word learning does not provide clear-cut predictions as to whether monolingual and bilingual children differ in their use of visible speech as they learn new words at 30 months. The purpose of this study is to elucidate this issue by investigating it in a twofold manner. First we want to evaluate whether, like their monolingual peers, French-learning bilingual children aged 30 months successfully learn new words in either the auditory or the visual modality. Second, we want to determine whether learning engenders a recoverable cross-modal representation of the words. These questions are addressed using the exact same stimuli and word-learning design as in Havy and Zesiger (2017). In our sample, bilingual children are exposed to French and another language from birth and receive input in each language for a minimum of 20% and a maximum of 80%, respectively (Bosch & Sebastián-Gallés, 2001). In the current study, bilingual children are randomly assigned to either one of the two learning groups: one that learns the words in the auditory modality and one that learns the words in the visual modality. In each learning group, children take part in two test conditions: (a) a 'same modality' test condition in which lexical recognition is tested in the same modality as the one used at learning (auditory after auditory learning, visual after visual learning) and (b) a 'cross-modality' test condition in which lexical recognition is tested in the other modality (visual after auditory learning, auditory after visual learning). The predictions are straightforward: If bilingual experience promotes the use of visible speech during lexical learning and fosters the cross-modal recognition of words, then bilingual children should succeed in every condition, including the most challenging one in which monolinguals have been found to fail: the visual-to-auditory cross-modal condition. If, however, language exposure does not exert any influence, then bilingual children should behave like their monolingual peers and demonstrate success in every condition except in the visual-to-auditory cross-modal condition. If bilinguals lag behind their monolingual peers, different patterns are possible. Given the extensive literature reporting successful auditory learning of distinct words in bilingual toddlers (Byers-Heinlein, Fennell, & Werker, 2013; Westermann & Mani, 2017), it is unlikely that 30-month-old bilinguals fail at learning words in the auditory modality. Yet, they may experience difficulties in using visible speech to guide lexical learning, as is the case in younger monolingual children (Havy & Zesiger, 2017), and fail in the 'same modality' and 'cross-modality' trials of the visual learning condition. Another possibility is that they may experience difficulties in establishing a recoverable cross-modal representation of words. In this case, failure is expected only in the 'cross-modality' trials of the auditory and visual learning conditions.

2 | METHOD

2.1 | Participants

A total of 48 children aged 30 months and living in Geneva completed the study, Table 1. Participants were recruited through birth records provided by the Canton of Geneva, Switzerland and enrolled upon parents' consent on behalf of their children. All children were carried full term with no diagnosed impairments in hearing, vision or language. Children were assigned to either one of two learning conditions (auditory, visual). Twenty-three children formed the auditory learning group (9 males). Another 25 children formed the visual learning group (12 males). Children were raised in a dual language environment from birth and considered to be simultaneous bilinguals. Language background included French and another language. Due to the geopolitical situation of Geneva, a wide constellation of languages was represented. Language exposure was assessed using the LEAT (DeAnda, Bosch, Poulin-Dubois, Zesiger, & Friend, 2016) with an 80% cut-off (Bosch & Sebastián-Gallés, 2001). Overall, the auditory and visual learning groups had comparable exposure to French, $t < 1$, two-tailed, with respectively 15 and 16 children showing a French dominance. Exposure to a third language, if present was sporadic and below 11%. Both groups came from mid-upper socioeconomic status families; $t < 1$, two-tailed (Havy et al., 2016), and had comparable word learning capacities (MCDI in production, Kern, 2003; $t < 1$, two-tailed). Twenty-two additional children participated but were excluded from the final sample due to excessive fussiness/crying ($n = 5$), failure to complete each test condition ($n = 7$), calibration issues ($n = 4$) or tracking ratio lower than 30% ($n = 6$). This yielded an attrition rate of 33%.

2.2 | Stimuli

2.2.1 | Speech stimuli

The speech stimuli were adapted from Havy and Zesiger (2017) and consisted of four pairs of French-sounding words. The pseudo-words were monosyllables with a CVC structure (/byp/-/var/,/rik/-/fal/,/fyf/-/gɛl/,/mum/-/tit/). These contrasted by at least two features on each segment (manner, place and voicing for the consonants, backness, height and roundness for the vowels) and were easily discriminable by monolingual (Streri et al., 2016; Werker & Curtin, 2005) and bilingual children aged 30 months (Albareda-Castellot, Pons, & Sebastián-Gallés, 2011; Burns, Yoshida, Hill, & Werker, 2007).

The pseudo-words were uttered by a native French female speaker in a child friendly directed register, at a slow speed and with slightly exaggerated intonation. These were instantiated in a carrier phrase: determiner ('un') + pseudo-word, so as to clarify the referential status of the word (Fennell & Waxman, 2010) and ensure that children effectively attended to the speech information prior to the word onset. For each pseudo-word, two tokens (two for each pseudo-word) served to the familiarization, learning and test phases and another one (one for each pseudo-word) served to the learning and test phases solely. For each token, three media sequences were generated: an audio–visual sequence, an auditory only sequence, a visual only sequence. The video sequences were presented at the center of the screen, with a display size corresponding roughly to a visual angle of 16° × 16° at a viewing distance of 60 cm.

2.2.2 | Object stimuli

The object stimuli were taken from Havy and Zesiger (2017). These consisted of four pairs of colorful novel objects rotating along a vertical axis against a black background. The objects stood alone at the center of the screen for the learning phase and together side by side for the test phase. At a viewing distance of 60 cm, the objects subtended approximately 16° × 16° of visual angle for the learning phase and 10° × 10° for the test phase. There was a gap of roughly six visual degrees between the objects at test. Each pair of objects was randomly assigned to a unique pair of pseudo-word. A smooth undulating shape with a display size of 3° × 3° of visual angle was used to sustain the children's interest.

2.3 | Apparatus

The visual stimuli were displayed on a 22-in. Dell E2209W monitor with a resolution of 1,680 × 1,050 pixels per inch and a refresh rate of 60 Hz. The auditory stimuli were delivered through loudspeakers located to the left-right sides of the monitor at a conversational level. A Dell Latitude E6520 laptop was used to handle calibration and stimuli presentation. The experiment was prepared and executed using I-view (I-view, version 2.8.26) and Experiment Center (Experiment Center, version 3.2.17) native to SMI (SensoMotoric Instruments GmbH). Children's eye-gaze was monitored by mean of a stand-alone SMI RED500 eye-tracker at a sampling rate of 60 Hz.

2.4 | Procedure

Testing took place in a sound attenuated and dimly illuminated room at the University of Geneva. During the experiment, children were sitting on the lap of one of their parents, approximately 60 cm away from a 22-in. eye-tracker-monitor set-up. The parents were instructed to keep their eyes closed and not to talk, point at the screen or influence the child's attention. We used the SMI eye tracker's five-point calibration routine to calibrate each child's gaze behavior. Once the calibration routine was completed, each child viewed four videos, one video per each pair of pseudo-words. Each video comprised three phases: (a) a familiarization phase, (b) a learning phase and (c) a test phase (Figure 2).

1. During the familiarization phase, the children were introduced to two pseudo-words uttered twice in alternation in their audio–visual mode (see and hear a talking face; i.e., token 1 of pseudo-word 1; token 1 of pseudo-word 2; token 2 of pseudo-word 1, token 2 of pseudo-word 2). This Audio–visual presentation of pseudo-words was done in order to direct the children's attention toward the multisensory aspect of speech.
2. Immediately after familiarization, the children completed a lexical learning phase. During lexical learning, the same two pseudo-words were presented in conjunction with two distinct novel objects. Word-object pairs consisted of a pseudo-word immediately followed by the corresponding object. Word-object pairs were spaced by a 1 s interval and arranged in the following order: three occurrences of one pair followed by three occurrences of the other, two occurrences of the first pair followed by two occurrences of the other and lastly one occurrence of each pair. To sustain the children's interest, three different tokens (two from the familiarization phase and one novel) of each pseudo-word were used and iterated twice across the pairs. Children enrolled in the auditory learning condition experienced the words in the auditory modality (they could hear the sound but saw a black screen, Figure 1). Children enrolled in the visual learning condition experienced the words in the visual modality (they saw the talking face in silence, Figure 2).
3. Immediately after the learning phase, the test phase began. The test phase was comprised of a pre-naming, a naming and a post-naming period. During the pre-naming and post-naming periods, children were presented with the two previously seen objects side by side in silence for 4 s. During the naming period, the objects were withdrawn and the label of one of them was played three times, each time with a different instantiation (pseudo-word 1: token 1, pseudo-word 1: token 2, pseudo-word 1: token 3). Modality of labeling varied upon the test condition. In the 'same modality' test condition, labeling happened in the same modality as the one used at learning: auditory after auditory learning (Figure 1a), visual after visual learning (Figure 2a). In the 'cross-modality' test condition, labeling happened in the other modality as the one used at learning: visual after auditory learning (Figure 1b); auditory after visual learning (Figure 2b). Each video

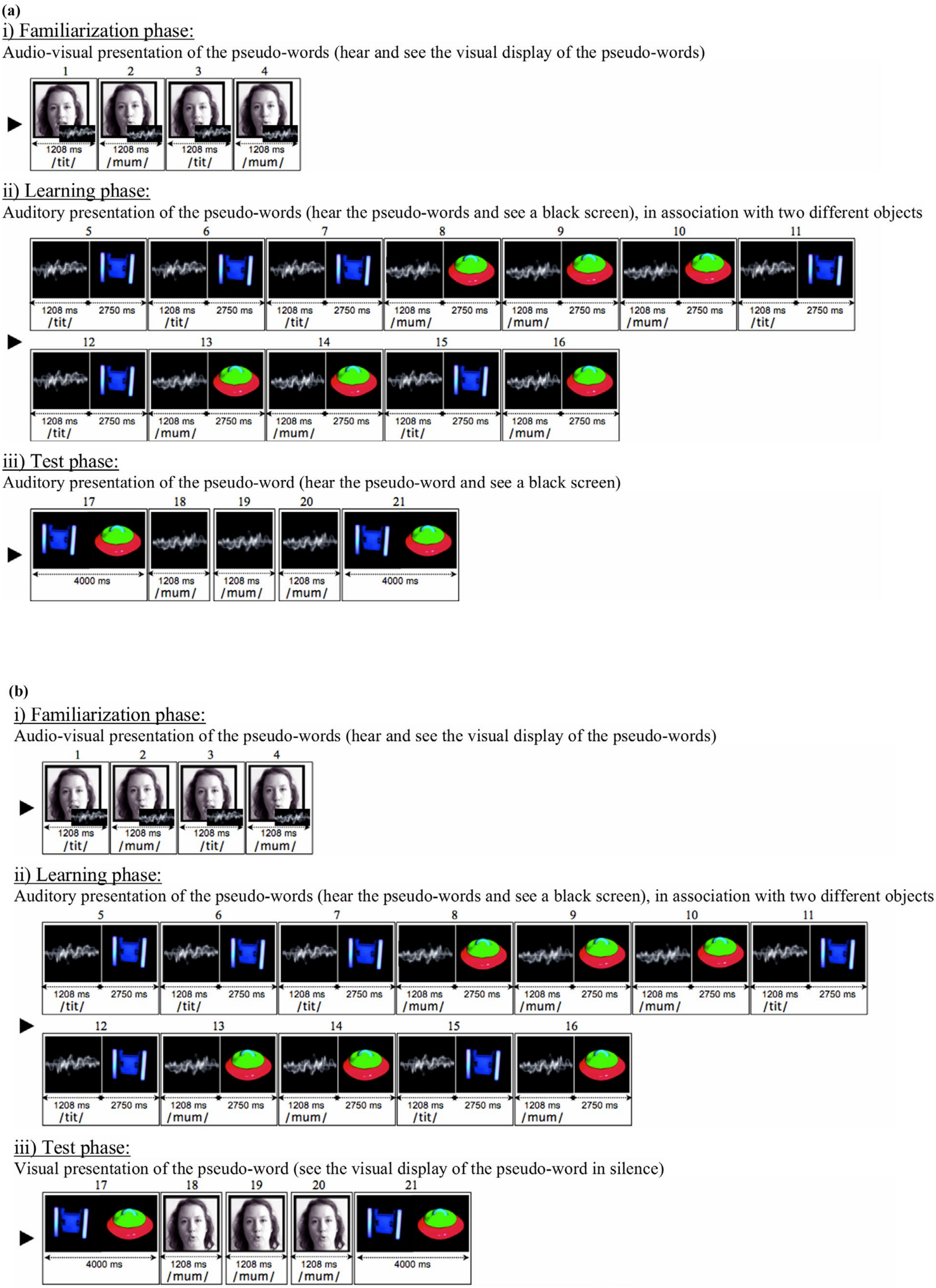

FIGURE 1 Schematic of a 'same modality' (a) and a 'cross-modality' (b) test trial in the auditory word-learning condition. The sequences of events are depicted in their actual order (from 1 to 21) for each experimental test trial. The sine wave represents the sound heard and is not actually seen on the screen

FIGURE 2 Schematic of a 'same modality' (a) and a 'cross-modality' (b) test trial in the visual word-learning condition. The sequences of events are depicted in their actual order (from 1 to 21) for each experimental test trial

tested one test condition only: either the 'same modality' or the 'cross-modality' condition. During the session, children received two videos testing the 'same modality' condition and two videos testing the 'cross-modality' condition. A smooth undulating shape was used between each video to reorient children to the screen. Sessions lasted approximately 20 min, after which each family received a small compensation.

2.5 | Counterbalancing

Participants of each learning group were randomly assigned to one out of three protocols.

Protocols varied the order of presentation of each word-object pair and for each word-object pair, the modality of labeling at test ('same modality' vs. 'cross-modality') and which of the two objects was labeled (object A vs. object B). All protocols started with a 'same modality' test trial to facilitate children's understanding of the task.

3 | RESULTS

3.1 | Data analyses

Analyses were performed using BeGaze (BeGaze, version 3.2.28). Gaze data were monitored and coded with respect to two areas of interest (AOIs) – one corresponding to the location of the target object, the other corresponding to the location of the distractor. AOIs were defined by dividing the entire screen in half. We adopted large AOIs to adjust for variations in object size across trials and reduce artefacts in how BeGaze interpolates eye position (Hessels & Hooge, 2019). Gaze data consisted of the sum of durations for all saccades and fixations hitting the AOIs. Following Havy et al. (2017; Havy & Zesiger, 2017, we analyzed the 4-s periods prior to (pre-naming period) and after naming (post-naming period) and computed for each period the proportion of target looking times (PLT), that is the amount of looking time devoted to the target object (T) over the amount of looking time to the target (T) and the distractor object (D): $T/(T + D)$. The data were aggregated by test condition, thus giving two naming scores per child, one for the 'same modality' condition and one for the 'cross-modality' condition.

3.2 | Data cleaning

Data cleaning consisted of a series of six filters successively applied to the initial dataset. The dataset initially included 220 trials for 55 participants (auditory learning: n = 27; visual learning n = 28) and upon filtering consisted of 170 trials for 48 participants (auditory learning: n = 23; visual learning n = 25). Filters were applied following Havy et al. (2017). We first trimmed three trials from the initial dataset on which children were not fixating on the monitor during the familiarization and learning phases (Filter 1: auditory learning: 3/110; visual learning: 0/110). We then removed six visual test trials (auditory learning: cross-modality; visual learning: same modality) on which children were not looking at the model during the visual-only naming period of the test phase (Filter 2: auditory learning: 4/55 trials; visual learning: 2/55 trials). Based on criteria commonly used in 2-choice word learning and word recognition tasks (Havy et al., 2017; Swingley & Aslin, 2000), we discarded twenty-three additional trials on which children were not fixating on the monitor during the pre-naming and/or post-naming periods of the test phase (Filter 3: auditory learning: 10/110; visual learning: 13/110). We then controlled for imbalance in spontaneous objects preferences and identified sixteen trials in which children attended either one of the objects during the pre-naming period (auditory learning: 8/110; visual learning: 8/110). Of these trials, eight were removed from our analyses (auditory learning: 4/110; visual learning: 4/110), as the bias endured beyond the post-naming period (Filter 4, Delle Luche, Durrant, Poltrock, & Floccia, 2015). The exclusion of these 40 trials led to the exclusion of seven participants, who no longer contributed to at least one data point in each of the test conditions ('same modality' vs. 'cross-modality'; Filter 5: auditory learning: n = 4; visual learning n = 3; Fernald, Perfors, & Marchman, 2006). The exclusion of the four participants resulted in the removal of ten additional trials (Filter 6: auditory learning: 4/110; visual learning: 6/110). A total of 50 trials (auditory learning: 25/110; visual learning: 25/110) were thus removed from the original dataset.

3.3 | Statistical analyses

Statistical analyzes were conducted using SPSS (SPSS, version 26.0.0.0). Data were log transformed to reduce skewness at the extremes of the distribution (DeCoster, 2001).

To appreciate children's PLT performance, several analyses were conducted separately for each learning group. Starting with the auditory learning condition, we ran a linear mixed effects model, using the percentage of target looking time per trial and per child as a dependent measure. As fixed effects, we entered in the model the test condition, the naming period as well as the interaction between the two predictors. The test condition corresponded to the modality of labeling at test ('same modality' vs. 'cross-modality'). The naming period corresponded to the period of time prior to and after labeling one of the two objects ('pre-naming' vs. 'post-naming'). As random effects, we initially entered random intercepts for participants and items and random slopes allowing for the effects of the naming period and test condition to differ across participants. However, the random slopes were subsequently removed as the model failed to converge. Results reported here stem from a by-participant and by-item intercept-only model. The model was fitted using maximum likelihood estimation. Estimates, standard errors and t-values are reported with $t > 2$ being interpreted as significant. T-tests were also performed to compare mean proportion of looking times (averaged over the trials of each condition) against chance (set at 50% since each response involved a choice between two equally probable possibilities).

The results of the mixed effects model yielded a significant main effect of naming period (t(134.45) = 2.70, p < .01), Tables 2 and 3. Overall, children in the auditory learning group were able to identify the target object upon naming. They demonstrated similar attention to either object prior to naming, t < 1; but a significant preference for the target object after naming, t(22) = 4.38, p < .01, d = 1.87. Other effects and interactions did not reach significance. Labeling effects surfaced in both the 'same modality' and 'cross-modality' test conditions, Figure 3a. In both conditions, there was no looking preference for either object prior to naming: 'same modality', t < 1; 'cross-modality', t < 1. After naming, there was a significant preference for the target object: 'same modality', t(22) = 3.86, p < .01, d = 1.65; 'cross-modality', t(22) = 2.11, p = .046, d = 0.90. Wald Z statistics revealed that the variation on the participants' (Wald Z = 0.84, p = .40) and items' intercepts was not confounded with our effects of primary interest (Wald Z = 0.94, p = .35).

We then considered the PLT responses of the visual learning group. These were submitted to a linear mixed effects model, using the exact same parameters as in the preceding analysis. The results of the mixed effects model revealed a significant main effect of naming period (t(142.63) = 2.01, p = .047), Tables 2 and 4. Children in the visual learning group were able to identify the target object upon naming. They devoted similar attention to either object prior to naming, t(24) = 1.27, p = .22, d = 0.52; but a significant preference for the target object after naming, t(24) = 4.06, p < .01, d = 1.66. Other effects and interactions were not significant. Just like above, labeling effects surfaced in both the 'same modality' and 'cross-modality' test conditions, Figure 3b. In both conditions, there was no looking preference for either object prior to naming: 'same modality', t(24) = 1.03, p = .31, d = 0.42; 'cross-modality', t(24) = 1.26, p = .22, d = 0.51. After naming, there was a significant preference for the target object: 'same modality', t(24) = 2.91, p < .01, d = 1.19; 'cross-modality', t(24) = 2.10, p = .046, d = 0.86. Wald Z statistics revealed no significant variation on the participants' (Wald Z = 0.82, p = .41) and items' intercepts (Wald Z = 1.66, p = .10).

Overall, the results indicate that bilingual children achieve successful lexical mapping, and apparently do so even in the condition in which monolinguals have been found to fail: the visual-to-auditory cross-modality condition.

To ascertain the effects of bilingual experience, we conducted further analyses using the monolingual dataset from Havy and Zesiger (2017). This dataset was collected with monolingual children of the same age in the exact same experimental conditions. Once again, separate analyses were run for each learning group. Starting with the auditory learning group, we submitted children's PLT responses to a linear mixed effects model, using the percentage of target looking time per trial and per child as a dependent measure. As fixed effects, we entered in the model the language status, the test condition, the naming period as well as all interactions between the three predictors. The language status referred to whether the participant was monolingual or bilingual. As random effects, we initially entered random intercepts for participants and items and random slopes allowing for heterogeneity in the effects of naming period and test condition across individuals. As before, due to lack of convergence, we retained an intercept-only model. The model was fitted using maximum likelihood estimation.

The analysis yielded a main effect of naming period (t(240.53) = 3.04, p < .01), indicating that children successfully recognize auditorily learned lexical mappings, Table 5. Of interest, this naming effect was not moderated by the language status or test condition, thus suggesting little influence of bilingual exposure on auditory word learning. Wald Z statistics revealed no significant variation

TABLE 2 Reporting bilingual children's performance in the different test conditions of auditory and visual word learning. Parameter estimates include the mean performance (M) in the different conditions, the standard deviation (SD) and T-test statistics comparing performance against chance level. Significant effects are bolded

Parameters			Parameters estimates: *M (SD)*	Parameters estimates: *T-test statistics*
Auditory learning				
	Overall	Pre-naming	51.11 (10.93)	t < 1
		Post-naming	62.80 (13.73)	**t(22) = 4.38, p < .01, d = 1.87**
	Same modality	Pre-naming	53.35 (13.59)	t < 1
		Post-naming	65.35 (18.11)	**t(22) = 3.86, p < .01, d = 1.65**
	Cross-modality	Pre-naming	48.87 (17.55)	t < 1
		Post-naming	60.25 (20.62)	**t(22) = 2.11, p = .046, d = 0.90**
Visual learning				
	Overall	Pre-naming	46.79 (15.44)	t(24) = 1.27, p = .22, d = 0.52
		Post-naming	59.50 (11.65)	**t(24) = 4.06, p < .01, d = 1.66**
	Same modality	Pre-naming	47.40 (17.37)	t(24) = 1.03, p = .31, d = 0.42
		Post-naming	58.64 (14.96)	**t(24) = 2.91, p < .01, d = 1.19**
	Cross-modality	Pre-naming	46.17 (20.30)	t(24) = 1.26, p = .22, d = 0.51
		Post-naming	60.36 (22.14)	**t(24) = 2.10, p = .046, d = 0.86**

TABLE 3 Results of a maximum likelihood estimated model predicting bilingual children's performance in the different test conditions of auditory word learning. Parameter estimates include the mean performance (*M*) in the different conditions, the estimated coefficient (Estimate) of the fixed effects and *t*-test statistics, the variance of the random effects and Wald *Z* statistics. Significant main effects and interactions are bolded

Parameters	Parameters estimates Estimate (*SE*)	*T*-test statistics
Fixed effects		
Main effects and interactions		
Test condition	−2.21 (3.68)	*t*(16.85) = 0.60, *p* = .56
Naming period	8.08 (2.99)	**t(134.45) = 2.70, p < .01**
Naming period × test condition	0.39 (4.26)	*t*(134.45) = 0.09, *p* = .93
	Variance (*SE*)	**Wald *Z* statistics**
Random effects		
Participants on intercepts	0.09 (0.11)	Wald *Z* = 0.84, *p* = .40
Items on intercepts	0.09 (0.09)	Wald *Z* = 0.94, *p* = .35

on the participants' (Wald *Z* = 1.21, *p* = .23*)* and items' intercepts (Wald *Z* = 1.54, *p* = .12).

We then regarded the PLT responses of the visual learning group. These were submitted to a linear mixed effects model, using the exact same parameters as before. The analysis revealed a main effect of naming period (*t*(260.03) = 2.54, *p* < .01), Table 6. This naming effect was moderated by the test condition and the language status (*t*(260.03) = 2.14, *p* = .03), suggesting variation across test conditions in how bilingual exposure influences visual word learning, Figure 4. Decomposition of the interaction revealed that bilingual exposure had no effects on naming outcomes in the 'same-modality' test condition (*t*(152.31) = −0.53, *p* = .60) but a significant influence in the 'cross-modality' test condition (*t*(105.79) = 2.52, *p* = .01). Other fixed effects did not reach significance. Wald *Z* statistics revealed no significant variation on the participants' intercepts (Wald *Z* = 0.82, *p* = .42) but a significant variation on the items' intercepts (Wald *Z* = 2.03, *p* = .04).

Overall, the current results demonstrate a bilingual edge in forming cross-modal representations of visually learned words but also pinpoint that this edge is moderated by the items being chosen.

4 | DISCUSSION

The purpose of the current study was to trace the influence of bilingual experience on how children appreciate visible orofacial speech information during lexical learning. We examined two core mechanisms whereby visible speech may become part of lexical representations. First, we asked whether 30-month-old bilingual children draw on visible speech alone to guide lexical learning. Secondly, we examined whether information from either auditory or visual modalities can be part of new lexical representations through cross-modal translation of the input. To test this, we employed the same word learning task as in Havy et al. (2017). Children were taught new lexical mappings in either the auditory or the visual modality and tested on recognition either in the same modality as the one used at learning ('same modality' condition) or the other modality ('cross-modality' condition).

First, we found that along with their monolingual peers of the same age, bilingual children are able to learn new word-object mappings from solely hearing or watching someone talking. This pattern is consistent with the monolingual literature showing that by 30 months children reliably attend to the visible aspects of speech as they establish new lexical representations (Havy & Zesiger, 2017). Of interest, our results document bilinguals' audio–visual speech perception skills at 30 months and fuel a growing literature that indicates that bilingual children precociously attend to the visible aspects of speech (Pons et al., 2015). Our results go even further and provide the first demonstration that bilingual children appropriately apply their exquisite visible speech perception sensitivities to the service of lexical learning. This is a noticeable developmental achievement especially given the difficulties monolinguals experience recruiting these sensitivities at earlier ages and sometimes later on when mapping words onto objects (Havy et al., 2017; Jerger et al., 2014, 2017). Here we clearly demonstrate that by 30 months, in the current word learning task, bilingual children, like their monolingual peers, consider visible speech as informative enough to guide lexical learning.

Another aim of our study was to determine whether bilingual children form a recoverable cross-modal representation upon auditory and visual word learning. The purpose was to determine whether unisensory experience of a word form evokes a multisensory representation. We found that along with monolingual children of the same age (Havy & Zesiger, 2017), bilingual children visually recognize words previously experienced in the auditory modality. This pattern indicates that regardless of the language environment, children are capable of cross-modal translation upon auditory learning and that the representation they form has a sensory format that contains or is accessible from the visual modality. Of interest, and to the difference of monolinguals, bilingual children auditorily recognize words previously experienced in the visual modality. This pattern suggests a bilingual edge in forming cross-modal representations of words upon visual input.

This bilingual edge, impressive in itself, rests on many antecedent developmental achievements. At first, there is the possibility that bilingual infants are more inclined to learn from the visible orofacial speech information and that cross-modal recognition arises as a result of greater confidence placed in this learning. This is supported by evidence showing that bilingual infants demonstrate stronger proclivities for orofacial cues than their peers (Fort, Lammertink, et al., 2018; Pons et al., 2015) and are better at discerning any two languages visually (Sebastián-Gallés et al., 2012). Yet, greater

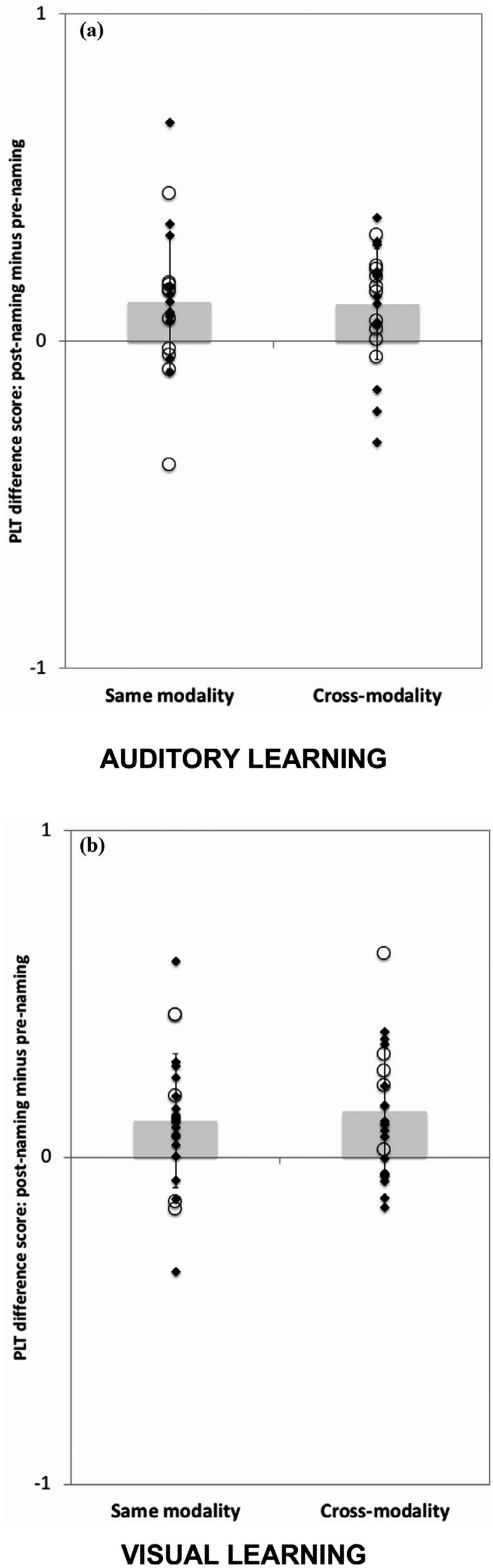

FIGURE 3 Mean difference score between the PLT (proportion of target looking times) during the pre-naming period and the PLT during the post-naming period (PLT in post-naming period minus PLT in pre-naming period), with the PLT being defined as target/(target + distractor). Positive values indicate greater looking to the target object rather than the distractor upon naming. Individual data points are overlaid on group means for the auditory learning group (a) and the visual learning group (b) in the 'same modality' and the 'cross-modality' test conditions. Black diamonds refer to bilingual participants. White circles refer to trilingual participants. Error bars represent the standard deviation of the mean

bilingual interest for the mouth region of speakers has not been consistently found (Mercure et al., 2019; Morin-Lessard et al., 2019) and as far as monolingual and bilingual children have been compared, there was no difference in their ability to visually recognize visually–learned words. Therefore, it is unclear as to whether the observed difference in the cross-modal condition stems primarily from greater visual word learning capacities.

The apparent bilingual advantage may be underpinned by a more effective way of handling uncertainty about lexical identity. Unlike auditory input, visible speech often matches more than one phonological template. Bilinguals may simply be better at activating and retaining multiple phonologically close lexical candidates and/or better at making a decision among them when trying to find a match with the test item. This possibility is supported by evidence showing that bilingual infants and adults have a greater capacity to retain information and update memory content as compared to monolinguals (Brito et al., 2015; Singh et al., 2015). Yet, such memory advantage has not been consistently found (Paap, Johnson, & Sawi, 2015) and in the very few auditory studies that examined lexical decisions among phonological neighbors there was no clear effect of bilingualism in toddlers (Canseco-Gonzalez et al., 2010; Von Holzen, Fennel, & Mani, 2019). Future research should control for these issues by manipulating the number of alternative candidates for a same visual word form.

Still in the lexical domain, the bilingual edge may result from a greater attentiveness to the discrete phonological detail of the word form during the establishment of word-object mapping. This greater accuracy may narrow down the number of auditory correspondences to draw from a same visual form and hone cross-modal recognition. This possibility is supported by auditory word comprehension and production studies that suggest that bilinguals are more accurate in perceiving (Singh et al., 2018) and producing the phonological details of words (Kehoe, 2018). However, some other evidence describes considerable variability in how they appreciate the fine-grained phonological details of words, with capacities that are sometimes comparable or inferior to those of monolinguals (perception: Fennell & Byers-Heinlein, 2014; Havy et al., 2016; Wewalaarachchi et al., 2017; production: MACLeod, Laukys, & Rvachew, 2011). One direction for future researches would be to test the influence of bilingual experience on children's ability to visually treat words that are minimally distinct.

TABLE 4 Results of a maximum likelihood estimated model predicting bilingual children's performance in the different test conditions of visual word learning. Parameter estimates include the mean performance (*M*) in the different conditions, the estimated coefficient (Estimate) of the fixed effects and *t*-test statistics, the variance of the random effects and Wald *Z* statistics. Significant main effects and interactions are bolded

Parameters	Parameters estimates Estimate (*SE*)	*T*-test statistics
Fixed effects		
Main effects and interactions		
Test condition	−3.70 (5.35)	$t(15.40) = 0.69, p = .50$
Naming period	6.29 (3.13)	**$t(142.63) = 2.01, p = .047$**
Naming period × test condition	2.47 (4.51)	$t(142.63) = 0.55, p = .59$
	Variance (*SE*)	**Wald *Z* statistics**
Random effects		
Participants on intercepts	0.11 (0.13)	Wald $Z = 0.82, p = .41$
Items on intercepts	0.45 (0.27)	Wald $Z = 1.66, p = .10$

TABLE 5 Results of a maximum likelihood estimated model comparing monolingual and bilingual children's performance in the different test conditions of auditory word learning. Parameter estimates include the estimated coefficient (Estimate) of the fixed effects and *t*-test statistics, the variance of the random effects and Wald *Z* statistics. Significant main effects and interactions are bolded

Parameters	Parameters estimates Estimate (*SE*)	*T*-test statistics
Fixed effects		
Main effects and interactions		
Language status	0.45 (4.11)	$t(32.57) = 0.11, p = .91$
Test condition	−1.67 (4.13)	$t(34.73) = −0.40, p = .69$
Naming period	9.88 (3.25)	**$t(240.53) = 3.04, p = .003$**
Language status × test condition	−0.50 (5.67)	$t(31.04) = −0.09, p = .93$
Naming period × test condition	3.00 (4.60)	$t(240.53) = 0.65, p = .52$
Language status × naming period	−1.80 (4.39)	$t(240.53) = −0.41, p = .68$
Language status × naming period × test condition	−3.39 (6.21)	$t(240.53) = −0.55, p = .59$
	Variance (*SE*)	**Wald *Z* statistics**
Random effects		
Participants on intercepts	0.10 (0.08)	Wald $Z = 1.21, p = .23$
Items on intercepts	0.13 (0.08)	Wald $Z = 1.54, p = .12$

Moving beyond lexical considerations alone, it could be the case that bilingualism hones low-level cross-modal exchanges between the auditory and visual modality. This is supported by brain imagery evidence showing that bilingual children (Della Rosa et al., 2013) and adults (Li, Legault, & Litcofsky, 2014) have more gray matter density and cortical thickness in two multisensory pathways that are highly engaged in audio–visual speech processing and are more inclined to bind auditory and visual features of speech (McGurk illusion, Marian, Hayakawa, Lam, & Schroeder, 2018) and non-speech events (Double flash illusion, Bidelman & Heath, 2019. Sound-color synesthesia, Ward, 2013; Watson et al., 2017). However, bilingual infants do not always outperform monolinguals in bridging their auditory and visual experiences (McGurk illusion, Mercure et al., 2019; Sound symbolism: Fort, Lammertink, et al., 2018; Pejovic & Molnar, 2017; Synesthesia: Watson et al., 2017). Future studies should test whether bilinguals' capacity to perform cross-modal translation upon visual speech is linked to more domain-general multisensory abilities.

In overview, bilingual success in the visual-to-auditory translation may arise from the joint action of all the aforementioned elements including perceptual, cognitive and lexical achievements. It could also stem from their capacity to take advantage of a pitfall of our design. The cross-modal word learning task we used was built up with a familiarization phase that children received prior to entering the lexical learning phase. During the familiarization phase, children had an opportunity to experience the word forms in the audio–visual modality. The familiarization was primarily designed to promote multisensory processing and was carefully controlled for duration to avoid any inferences. In monolinguals, the familiarization did not promote cross-modal recognition upon visual learning. Yet bilinguals have been found to have better attention and memory skills and better capacities to reason and draw complex inferences in the auditory and visual domain (Becker, Prat, & Stocco, 2016; Hara & Tappe, 2016). Therefore, it could be the case that bilinguals exploit this source of information to bridge the gap in the sensory information available during lexical learning and infer the auditory or visual correspondence of the word. Future studies should control this issue by using different words during familiarization and lexical learning.

Importantly, the bilingual edge, in the visual-to-auditory translation, did not surface as strongly in all learning situations as there was significant by-item variation. Such variation was not found in auditory word learning. This calls into question the generality and robustness of the bilingual advantage and its contingence to the material being used. Future studies should replicate these effects using other word-object pairings.

Along these lines, bilingual children were considered as a group, but bilingual children do not all share the same language exposure history and may therefore not approach the task the same way. First

TABLE 6 Results of a maximum likelihood estimated model comparing monolingual and bilingual children's performance in the different test conditions of visual word learning. Parameter estimates include the estimated coefficient (Estimate) of the fixed effects and *t*-test statistics, the variance of the random effects and Wald *Z* statistics. Planned comparisons compare monolingual and bilingual children on the size of the naming effects in each test conditions (language status × naming period). Significant main effects and interactions are bolded

Parameters	Parameters estimates Estimate (*SE*)	*T*-test statistics
Fixed effects		
Main effects and interactions		
Language status	0.06 (0.06)	$t(16.61) = -0.01$, $p = .99$
Test condition	5.45 (7.17)	$t(13.90) = -0.76$, $p = .46$
Naming period	8.61 (3.40)	**$t(260.03) = 2.54$, $p = .01$**
Language status × test condition	−9.21 (8.78)	$t(16.31) = -1.05$, $p = .31$
Naming period × test condition	−11.40 (4.84)	**$t(260.03) = -2.36$, $p = .02$**
Language status × Naming period	−2.32 (4.53)	$t(260.03) = -0.51$, $p = .61$
Language status × naming period × test condition	14.87 (6.48)	**$t(260.03) = 2.14$, $p = .03$**
	Variance (*SE*)	**Wald *Z* statistics**
Random effect		
Participants on intercepts	0.07 (0.08)	Wald $Z = 0.82$, $p = .42$
Items on intercepts	0.40 (0.20)	**Wald $Z = 2.03$, $p = .04$**
	Estimate (*SE*)	***T*-test statistics**
Planned comparisons		
Testing bilingual influence on naming effects		
Same modality	−2.32 (4.41)	$t(152.31) = -0.53$, $p = .60$
Cross-modality	11.55 (4.59)	**$t(105.79) = 2.52$, $p = .01$**

and foremost, there is evidence that bilingual language outcomes are influenced by the amount of input received in each language (Bijeljac-Babic, 2018). Bilingual preschoolers who spend 40% or more of their waking hours since birth in a given language environment are more likely to attain receptive and expressive vocabulary achievement in that language, comparable to that of monolinguals (Thordardottir, 2019; Unsworth, 2016). Here, the majority of participants ($n = 37/48$) received more than 40% of French exposure and post-hoc analyses did not yield any significant effect of this factor ($ts < 2$). Future studies testing more children with lower amounts of French should determine whether there is some threshold in language dominance that critically affects the sensory format of the emerging lexicon.

Along these lines, it must be noted that bilingual input not only varies in quantity but also in quality and that children receiving their dominant language with more heavily accentuated foreign input may be more inclined to search for visually available speech cues to resolve auditory uncertainties. This is supported by evidence showing that monolingual adults attend more to visible speech information when dealing with accentuated speech (Zheng & Samuel, 2019). Future studies should evaluate the language proficiency of parental input and determine its influence on bilinguals' capacity to establish lexical representations upon visible speech.

Importantly, bilingual word learning behavior may also be influenced by the relative similarity of the languages to be learned. This is supported by findings showing that bilingual infants and preschool aged children who learn languages that share similarities at the phonological and lexical level are more likely to attend to the talker's mouth than bilinguals learning more distant languages when viewing someone talking (Birulés et al., 2018). It could hence be the case that these children perform better in our task than the other bilinguals. Given the variety of languages represented in our sample, additional study is needed to explore such effects.

Another point of relevance is the consideration of age at second language exposure. The current study focused its scope on simultaneous bilinguals, which are bilinguals exposed to both languages from birth. But there is the case of sequential bilinguals who are learning one language first and then another one later. The distinction between the two language experiences is of importance as there are maturational epochs in early childhood during which the developing brain is optimally more open to visual experiential influences (Birdsong, 2018; Choi, Black, & Werker, 2018). This is illustrated by adult studies who found that early childhood exposure is crucial for using relevant visual speech information to separate languages visually (Weikum et al., 2013) and boost non-native phonetic discrimination (Burfin et al., 2014). Early childhood exposure also promises more balanced lexical abilities across the two languages (Junker & Stockman, 2002). Future studies should explore whether the bilingual advantage observed in the current study holds for children who have later second language experience.

In the same vein, it should be stressed that our bilingual population was uniquely exposed to spoken languages. But there is also the case of bimodal bilinguals who exhibit a singular form of bilingualism that involves both spoken and sign language. This population is mainly found in families where hearing children are raised by deaf parents and experience one language exclusively in the visual modality. The extent use of the visual modality may convey an edge in attending to the visible orofacial

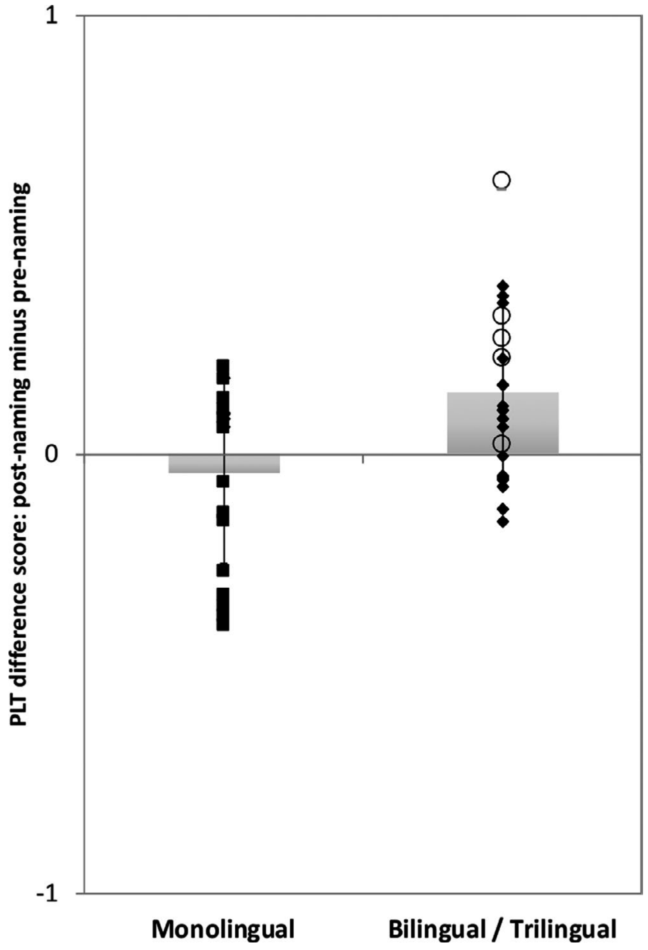

FIGURE 4 Data from monolingual (from Havy & Zesiger, 2017) bilingual and trilingual children in the cross-modal condition of visual word learning. Mean difference score between the PLT (proportion of target looking times) during the pre-naming period and the PLT during the post-naming period (PLT in post-naming period minus PLT in pre-naming period), with the PLT being defined as target/(target + distractor). Positive values indicate greater looking to the target object rather than the distractor upon naming. Individual data points are overlaid on group means for the auditory learning group (a) and the visual learning group (b) in the 'same modality' and the 'cross-modality' test conditions. Monolingual, bilingual, and trilingual participants are referred to respectively by back squares, black diamonds, and white circles. Error bars represent the standard deviation of the mean

information of the spoken language. This is supported by evidence showing that many sign languages use facial expressions (raised eyebrows, distinct mouth configurations) to mark linguistic structure (Zeshan, 2004) and that using sign language enhances the ability to discriminate facial expressions (Stoll et al., 2017) and retain visuo-spatial information (Geraci, Gozzi, Papagno, & Cecchetto, 2008). Besides, bimodal bilinguals frequently produce elements of both languages at the same time without a cost. This cross-modal code blending between languages may give an edge in forming cross-modal associations. However, there is concurrent evidence that the interest for the mouth in bimodal bilinguals is inferior to that of unimodal bilinguals (Mercure et al., 2019). Alongside this, studies have found that attention shifting and generalization of memory content across modalities which are foundational to cross-modal associations are inferior to those of unimodal bilinguals (Emmorey, Luk, Pyers, & Bialystok, 2008). Future studies should address this issue by comparing unimodal and bimodal bilingual populations.

Finally, if being bilingual is advantageous, what about being trilingual. Trilingualism is generally treated in the relevant literature as another type of bilingualism, and theories and findings from studies of bilinguals are often assumed to be applicable to trilinguals by extension, but trilingualism may have distinct consequences (Schroeder & Marian, 2017). Trilingual experience places an increased burden on memory and executive processes. The processing demand associated to acquiring a third language may bolster certain bilingual advantages or excessively drain resources and even fail to elicit the advantages observed in bilinguals, particularly given the low cognitive supply in infants and toddlers. In this respect, the literature documents various outcomes of trilingual experience in diverse executive (Brito et al., 2015; Poarch & Bialystok, 2015) and lexical tasks (Byers-Heinlein et al., 2013; Mieszkowska et al., 2017). In our study, about a third of our participants were exposed to a third language but this exposure was relatively small (5%–11%) and post-hoc analyses did not yield any significant effects of this factor (*ts* < 2). Future studies should evaluate how trilingual children with a more balanced language experience behave in our word learning task.

In sum our study documents a bilingual edge in how children bridge their auditory and visual experience of speech during lexical acquisition. Adopting a developmental lens, future researches should explore whether this advantage holds for younger and older ages.

ACKNOWLEDGEMENTS

The authors thank the children and their families for their participation. The authors also thank E. Stanford, E. Gutierrez, S. Gutierrez, P. Bombart, J. Sauvage, and B. Belle for their insightful comments.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Albareda-Castellot, B., Pons, F., & Sebastián-Gallés, N. (2011). The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*, *14*, 395–401. https://doi.org/10.1111/j.1467-7687.2010.00989.x

Altvater-Mackensen, N., Mani, N., & Grossmann, T. (2016). Audiovisual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis) matches between auditory and visual speech cues. *Developmental Psychology*, *52*, 191–204. https://doi.org/10.1037/a0039964

Ayneto, A., & Sebastián-Gallés, N. (2017). The influence of bilingualism on the preference for the mouth region of dynamic faces. *Developmental Science*, *20*, e12446. https://doi.org/10.1111/desc.12446

Baart, M., Vroomen, J., Shaw, K., & Bortfeld, H. (2014). Degrading phonetic information affects matching of audiovisual speech in adults, but not in infants. *Cognition*, *130*, 31–43. https://doi.org/10.1016/j.cognition.2013.09.006

Becker, T. M., Prat, C. S., & Stocco, A. (2016). A network-level analysis of cognitive flexibility reveals a differential influence of the anterior cingulate cortex in bilinguals versus monolinguals. *Neuropsychologia*, *85*, 62–73. https://doi.org/10.1016/j.neuropsychologia.2016.01.020

Berger, J. (1977). *Ways of seeing 1972*. London, UK: British Broadcasting Corporation and Penguin Books.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*, 233. https://doi.org/10.1037/bul0000099

Bidelman, G. M., & Heath, S. (2019). Enhanced temporal binding of audiovisual information in the bilingual brain. *Bilingualism: Language and Cognition*, *22*, 752–762.

Bijeljac-Babic, R. (2018). L'acquisition du langage au cours de la première année de vie: Le cas des nourrissons bilingues. *Devenir*, *30*, 31–41.

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, *9*, 81. https://doi.org/10.3389/fpsyg.2018.00081

Birulés, J., Bosch, L., Brieke, R., Pons, F., & Lewkowicz, D. J. (2018). Inside bilingualism: Language background modulates selective attention to a talker's mouth. *Developmental Science*, *22*, e12755.

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy*, *2*, 29–49. https://doi.org/10.1207/S15327078IN0201_3

Brito, N. H., Sebastián-Gallés, N., & Barr, R. (2015). Differences in language exposure and its effects on memory flexibility in monolingual, bilingual, and trilingual infants. *Bilingualism: Language and Cognition*, *18*, 670–682. https://doi.org/10.1017/S1366728914000789

Buchwald, A. B., Winters, S. J., & Pisoni, D. B. (2009). Visual speech primes open-set recognition of spoken words. *Language and Cognitive Processes*, *24*, 580–610. https://doi.org/10.1080/01690960802536357

Burfin, S., Pascalis, O., Ruiz Tada, E., Costa, A., Savariaux, C., & Kandel, S. (2014). Bilingualism affects audiovisual phoneme identification. *Frontiers in Psychology*, *5*, 1179. https://doi.org/10.3389/fpsyg.2014.01179

Burns, T. C., Yoshida, K. A., Hill, K., & Werker, J. F. (2007). The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*, *28*, 455–474. https://doi.org/10.1017/S0142716407070257

Byers-Heinlein, K., Fennell, C. T., & Werker, J. F. (2013). The development of associative word learning in monolingual and bilingual infants. *Bilingualism: Language and Cognition*, *16*, 198–205. https://doi.org/10.1017/S1366728912000417

Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 9032–9037. https://doi.org/10.1073/pnas.1703220114

Canseco-Gonzalez, E., Brehm, L., Brick, C. A., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or Cárcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes*, *25*, 669–705. https://doi.org/10.1080/01690960903474912

Choi, D., Black, A. K., & Werker, J. F. (2018). Cascading and multisensory influences on speech perception development. *Mind, Brain, and Education*, *12*, 212–223. https://doi.org/10.1111/mbe.12162

Crivello, C., Kuzyk, O., Rodrigues, M., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2016). The effects of bilingual growth on toddlers' executive function. *Journal of Experimental Child Psychology*, *141*, 121–132. https://doi.org/10.1016/j.jecp.2015.08.004

Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*, *42*, 37–48. https://doi.org/10.1016/j.cogdev.2017.02.004

Davies, R., Kidd, E., & Lander, K. (2009). Investigating the psycholinguistic correlates of speechreading in preschool age children. *International Journal of Language & Communication Disorders*, *44*, 164–174. https://doi.org/10.1080/13682820801997189

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The language exposure assessment tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research*, *59*, 1346–1356. https://doi.org/10.1044/2016_JSLHR-L-15-0234

DeCoster, J. (2001). *Transforming and restructuring data*. Retrieved from http://www.stathelp.com/notes.html

Della Rosa, P. A., Videsott, G., Borsa, V. M., Canini, M., Weekes, B. S., Franceschini, R., & Abutalebi, J. (2013). A neural interactive location for multilingual talent. *Cortex*, *49*, 605–608. https://doi.org/10.1016/j.cortex.2012.12.001

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the Intermodal Preferential Looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development*, *40*, 151–172. https://doi.org/10.1016/j.infbeh.2015.05.005

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science*, *19*, 1201–1206. https://doi.org/10.1111/j.1467-9280.2008.02224.x

Fennell, C. T., & Byers-Heinlein, K. (2014). You sound like mommy: Bilingual and monolingual infants learn words best from speakers typical of their language environments. *International Journal of Behavioral Development*, *38*, 309–316. https://doi.org/10.1177/0165025414530631

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*, 1376–1383. https://doi.org/10.1111/j.1467-8624.2010.01479.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*, 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fort, M., Ayneto-Gimeno, A., Escrichs, A., & Sebastian-Galles, N. (2018). Impact of bilingualism on infants' ability to learn from talking and nontalking faces. *Language Learning*, *68*, 31–57.

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

Geraci, C., Gozzi, M., Papagno, C., & Cecchetto, C. (2008). How grammar can cope with limited short-term memory: Simultaneity and seriality in sign languages. *Cognition*, *106*, 780–804. https://doi.org/10.1016/j.cognition.2007.04.014

Grieco-Calub, T. M., & Olson, J. (2015). Individual differences in real-time processing of audiovisual speech by preschool children. *The Journal of the Acoustical Society of America*, *137*, 2375. https://doi.org/10.1121/1.4920629

Hara, A., & Tappe, H. (2016). Inference generation and text comprehension in bilingual children: A case study. *Literator*, *37*, a1287. https://doi.org/10.4102/lit.v37i2.1287

Havy, M., Bouchon, C., & Nazzi, T. (2016). Phonetic processing when learning words: The case of bilingual infants. *International Journal of Behavioral Development*, *40*, 41–52. https://doi.org/10.1177/0165025415570646

Havy, M., Foroud, A., Fais, L., & Werker, J. F. (2017). The role of auditory and visual speech in word learning at 18 months and in adulthood. *Child Development*, *88*, 2043–2059.

Havy, M., & Zesiger, P. (2017). Learning spoken words via the ears and eyes: Evidence from 30-month-old children. *Frontiers in Psychology*, *8*, 2122. https://doi.org/10.3389/fpsyg.2017.02122

Heikkilä, J., Lonka, E., Ahola, S., Meronen, A., & Tiippana, K. (2017). Lipreading ability and its cognitive correlates in typically developing children and children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *60*, 485–493. https://doi.org/10.1044/2016_JSLHR-S-15-0071

Hessels, R. S., & Hooge, I. T. (2019). Eye tracking in developmental cognitive neuroscience – The good, the bad and the ugly. *Developmental Cognitive Neuroscience*, *40*, 100710. https://doi.org/10.1016/j.dcn.2019.100710

Hillairet de Boisferon, A. H., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: Role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*, e12381.

Hillairet de Boisferon, A. H., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology*, *172*, 189–200. https://doi.org/10.1016/j.jecp.2018.03.009

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*, 231–255. https://doi.org/10.1207/s15327078in0602_5

Jerger, S., Damian, M. F., Tye-Murray, N., & Abdi, H. (2014). Children use visual speech to compensate for non-intact auditory speech. *Journal of Experimental Child Psychology*, *126*, 295–312. https://doi.org/10.1016/j.jecp.2014.05.003

Jerger, S., Damian, M. F., Tye-Murray, N., & Abdi, H. (2017). Children perceive speechonsets by ear and eye. *Journal of Child Language*, *44*, 185–215. https://doi.org/10.1017/S030500091500077X

Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology*, *11*, 381–394. https://doi.org/10.1044/1058-0360(2002/042)

Kehoe, M. (2018). The development of rhotics: A comparison of monolingual and bilingual children. *Bilingualism: Language and Cognition*, *21*, 710–731. https://doi.org/10.1017/S1366728916001279

Kern, S. (2003). Le compte-rendu parental au service de l'évaluation de la production lexicale des enfants français entre 16 et 30 mois. *Glossa*, *85*, 48–62.

Kubicek, C., De Boisferon, A. H., Dupierrix, E., Pascalis, O., Lœvenbruck, H., Gervain, J., & Schwarzer, G. (2014). Cross-modal matching of audio-visual German and French fluent speech in infancy. *PLoS One*, *9*, e89275. https://doi.org/10.1371/journal.pone.0089275

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

Kuhl, P. K., Williams, K. A., & Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using nonspeech auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 829.

Lalonde, K., & Holt, R. F. (2015). Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research*, *58*, 135–150. https://doi.org/10.1044/2014_JSLHR-H-13-0343

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 1431–1436. https://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., Minar, N. J., Tift, A. H., & Brandon, M. (2015). Perception of the multisensory coherence of fluent audiovisual speech in infancy: Its emergence and the role of experience. *Journal of Experimental Child Psychology*, *130*, 147–162. https://doi.org/10.1016/j.jecp.2014.10.006

Lewkowicz, D. J., & Pons, F. (2013). Recognition of amodal language identity emerges in infancy. *International Journal of Behavioral Development*, *37*, 90–94. https://doi.org/10.1177/0165025412467582

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, *58*, 301–324. https://doi.org/10.1016/j.cortex.2014.05.001

MACLeod, A. A. N., Laukys, K., & Rvachew, S. (2011). The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech-Language Pathology*, *13*, 490–499. https://doi.org/10.3109/17549507.2011.578658

Marassa, L. K., & Lansing, C. R. (1995). Visual word recognition in two facial motion conditions: Full-face versus lips-plus-mandible. *Journal of Speech, Language, and Hearing Research*, *38*, 1387–1394. https://doi.org/10.1044/jshr.3806.1387

Marian, V., Hayakawa, S., Lam, T., & Schroeder, S. (2018). Language experience changes audiovisual perception. *Brain Sciences*, *8*, 85. https://doi.org/10.3390/brainsci8050085

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*, e12701. https://doi.org/10.1111/desc.12701

Mieszkowska, K., Łuniewska, M., Kołak, J., Kacprzak, A., Wodniecka, Z., & Haman, E. (2017). Home language will not take care of itself: Vocabulary knowledge in trilingual children in the United Kingdom. *Frontiers in Psychology*, *8*, 1358. https://doi.org/10.3389/fpsyg.2017.01358

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*, 338–352. https://doi.org/10.1121/1.1907526

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*. https://doi.org/10.1037/dev0000750

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278. https://doi.org/10.1016/j.cortex.2015.04.014

Pejovic, J., & Molnar, M. (2017). The development of spontaneous sound shape matching in monolingual and bilingual infants during the first year. *Developmental Psychology*, *53*, 581–586. https://doi.org/10.1037/dev0000237

Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review.*, *35*, 113–124.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, *26*, 490–498. https://doi.org/10.1177/0956797614568320

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián-Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 10598–10602. https://doi.org/10.1073/pnas.0904134106

Ramon-Casas, M., Fennell, C. T., & Bosch, L. (2017). Minimal-pair word learning by bilingual toddlers: The Catalan/e/-/ɛ/contrast revisited. *Bilingualism: Language and Cognition*, *20*, 649–656. https://doi.org/10.1017/S1366728916001115

Ramon-Casas, M., Swingley, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, *59*, 96–121. https://doi.org/10.1016/j.cogpsych.2009.02.002

Ronquest, R. E., Levi, S. V., & Pisoni, D. B. (2010). Language identification from visual-only speech signals. *Attention, Perception, & Psychophysics*, *72*, 1601–1613. https://doi.org/10.3758/APP.72.6.1601

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *The European Journal of Neuroscience*, *33*, 2329–2337. https://doi.org/10.1111/j.1460-9568.2011.07685.x

Rvachew, S., & Alhaidary, A. (2018). The phonetics of babbling. In M. Aronoff (Ed.), *Oxford research encyclopedia of linguistics*. Retrieved from http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-413

Samuel, A. G., & Lieblich, J. (2014). Visual speech acts differently than lexical context in supporting speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1479–1490.

Schonberg, C., Sandhofer, C. M., Tsang, T., & Johnson, S. P. (2014). Does bilingual experience affect early visual perceptual development? *Frontiers in Psychology*, *5*, 1429. https://doi.org/10.3389/fpsyg.2014.01429

Schroeder, S. R., & Marian, V. (2017). Cognitive consequences of trilingualism. *International Journal of Bilingualism*, *21*, 754–773. https://doi.org/10.1177/1367006916637288

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, *23*, 994–999. https://doi.org/10.1177/0956797612436817

Singh, L., Fu, C. S. L., Rahman, A. A., Hameed, W. B., Sanmugam, S., Agarwal, P., & GUSTO Research Team. (2015). Back to basics: A bilingual advantage in infant visual habituation. *Child Development*, *86*, 294–302. https://doi.org/10.1111/cdev.12271

Singh, L., Fu, C. S., Tay, Z. W., & Golinkoff, R. M. (2018). Novel word learning in bilingual and monolingual infants: Evidence for a bilingual advantage. *Child Development*, *89*, e183–e198. https://doi.org/10.1111/cdev.12747

Soto-Faraco, S., Navarra, J., Weikum, W. M., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception & Psychophysics*, *69*, 218–231. https://doi.org/10.3758/BF03193744

Stoll, C., Palluel-Germain, R., Caldara, R., Lao, J., Dye, M. W., Aptel, F., & Pascalis, O. (2017). Face recognition is shaped by the use of sign language. *The Journal of Deaf Studies and Deaf Education*, *23*, 62–70.

Streri, A., Coulon, M., Marie, J., & Yeung, H. H. (2016). Developmental change in infants' detection of visual faces that match auditory vowels. *Infancy*, *21*, 177–198. https://doi.org/10.1111/infa.12104

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*, 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*, 1173–1190. https://doi.org/10.1017/S0305000914000725

Ter Schure, S., Junge, C., & Boersma, P. (2016). Discriminating non-native vowels on the basis of multimodal, auditory or visual information: Effects on infants' looking patterns and discrimination. *Frontiers in Psychology*, *7*, 525. https://doi.org/10.3389/fpsyg.2016.00525

Thordardottir, E. (2019). Amount trumps timing in bilingual vocabulary acquisition: Effects of input in simultaneous and sequential school-age bilinguals. *International Journal of Bilingualism*, *23*, 236–255. https://doi.org/10.1177/1367006917722418

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Tye-Murray, N., Hale, S., Spehar, B., Myerson, J., & Sommers, M. S. (2014). Lipreading in school-age children: The roles of age, hearing status, and cognitive ability. *Journal of Speech, Language, and Hearing Research*, *57*, 556–565. https://doi.org/10.1044/2013_JSLHR-H-12-0273

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Lifespan perspectives on bilingualism* (pp. 136–196). Berlin, Germany: Mouton de Gruyter.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*, 375–382. https://doi.org/10.1016/j.cognition.2013.02.015

Von Holzen, K., Fennel, C. T., & Mani, N. (2019). The impact of cross-language phonological overlap on bilingual and monolingual toddlers' word recognition. *Bilingualism: Language and Cognition*, *22*, 476–499.

Ward, J. (2013). Synesthesia. *Annual Review of Psychology*, *64*, 49–75. https://doi.org/10.1146/annurev-psych-113011-143840

Watson, M. R., Chromý, J., Crawford, L., Eagleman, D. M., Enns, J. T., & Akins, K. A. (2017). The prevalence of synaesthesia depends on early language learning. *Consciousness and Cognition*, *48*, 212–231. https://doi.org/10.1016/j.concog.2016.12.004

Weatherhead, D., & White, K. S. (2017). Read my lips: Visual speech influences word processing in infants. *Cognition*, *160*, 103–109. https://doi.org/10.1016/j.cognition.2017.01.002

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, *316*, 1159. https://doi.org/10.1126/science.1137686

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2013). Age-related sensitive periods influence visual language discrimination in adults. *Frontiers in Systems Neuroscience*, *7*, 86. https://doi.org/10.3389/fnsys.2013.00086

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234. https://doi.org/10.1080/15475441.2005.9684216

Westermann, G., & Mani, N. (2017). *Early word learning*. London, UK: Routledge.

Wewalaarachchi, T. D., Wong, L. H., & Singh, L. (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English-Mandarin toddlers. *Journal of Experimental Child Psychology*, *159*, 16–33. https://doi.org/10.1016/j.jecp.2017.01.009

Yeung, H. H., & Werker, J. F. (2013). Lip movements affect infants' audiovisual speech perception. *Psychological Science*, *24*, 603–612. https://doi.org/10.1177/0956797612458802

Zeshan, U. (2004). Hand, head and face – Negative constructions in sign languages. *Linguistic Typology*, *8*, 1–58. https://doi.org/10.1515/lity.2004.003

Zheng, Y., & Samuel, A. G. (2019). How much do visual cues help listeners in perceiving accented speech? *Applied Psycholinguistics*, *40*, 93–109. https://doi.org/10.1017/S0142716418000462

How to cite this article: Havy M, Zesiger PE. Bridging ears and eyes when learning spoken words: On the effects of bilingual experience at 30 months. *Dev Sci*. 2021;24:e13002. https://doi.org/10.1111/desc.13002