Received: 18 February 2019 | Revised: 25 November 2020 | Accepted: 22 December 2020

DOI: 10.1002/icd.2217

RESEARCH ARTICLE

WILEY

A cross-linguistic study of multisensory perceptual narrowing in German and Swedish infants during the first year of life

Katharina Dorn[1] | Elodie Cauvet[2] | Sabine Weinert[1]

[1]Department of Developmental Psychology, Otto-Friedrich University, Bamberg, Germany

[2]Department of Women's and Children's health, Karolinska Institute of Neurodevelopmental Disorders (KIND), Stockholm, Sweden

Correspondence
Katharina Dorn, Department of Developmental Psychology, Otto-Friedrich University, Bamberg, Germany.
Email: katharina.dorn@uni-bamberg.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Number: Deutscher Akademischer Austauschdienst - IPID4all

Abstract

Four-and-a-half-month-olds look longer at silent mouth movements corresponding to a language they previously listened to. The *perceptual narrowing* hypothesis suggests this general ability to decline as a consequence of experience with the infant's native language. We tracked eye-gaze of German and Swedish infants longitudinally in an *intersensory matching procedure* at 4.5 and 6 months of age. Infants watched and listened sequentially to side-by-side presentations of visual and corresponding auditory fluent speech in their respective native or the non-native language. Looking times indicated that 4.5-month-old infants preferred the respective language they previously listened to, either native or non-native. However, at 6 months of age they only audio-visually matched their native language and kept looking at chance level after listening to the non-native language – suggesting that the intersensory perception of languages narrows down before 6 months of age even in same-rhythm-class languages. Intriguingly, the 6-month-old German and Swedish samples showed different patterns of preference after listening to their native language. Whereas the German infants looked significantly longer to the German visual speech, the Swedish infants looked significantly shorter to the Swedish visual speech. Different explanations and practical implications for early hearing aids are discussed within the frame of perceptual narrowing.

Highlights

- First study to examine perceptual narrowing in same-rhythm-class languages.
- In an intersensory matching procedure audio-visual matching of languages narrows down before 6 months of age in same-rhythm-class languages.
- The empirical findings might have crucial practical implications for early hearing aids.

KEYWORDS

audio-visual speech perception, cross-linguistic study, eye-tracking, multisensory perceptual narrowing, same-rhythm-class languages

1 | INTRODUCTION

From birth on infants are exposed to an audio-visual environment, leading to a close binding between these multimodal stimuli. The ability to integrate multisensory information is a fundamental ability emerging very early in life (Maurer & Mondloch, 1996; Sai, 2005; Streri, Coulon, & Guellaï, 2013). A variety of studies have established that infants look longer at a face articulating the vowel they had just listened to, thus indicating an early developing sensitivity to match audio-visual stimuli. In particular, infants at 4.5–5 months of age preferred looking at the respective face that matched the synchronously presented sound, hence showing awareness of the congruence between speech and lip movements (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999; Yeung & Werker, 2013). This integration is also evidenced by the *McGurk*-effect, a conflict appearing when the auditory and visual speech input are incongruent, resulting in illusory perceptions in adults as well as in infants (Burnham & Dodd, 2004; Dodd, 1979; Kushnerenko, Teinonen, Volein, & Csibra, 2008; McGurk & MacDonald, 1976).

An *intersensory matching procedure* has commonly been used to examine audio-visual matching abilities. This method pairs a couple of visual stimuli, for example two faces (mouth movements), with one auditory stimulus, such as an auditory syllable that matches only one of the presented visual stimuli. Two different versions of this procedure can be distinguished, that is, the stimuli are presented either simultaneously or sequentially. The synchronous presentation could simplify the audio-visual matching since infants may rely on temporal synchrony cues that might enhance the attention of an infant to a stimulus (Bahrick & Lickliter, 2000; Bahrick, Lickliter, & Flom, 2004) with low working memory load (Kubicek et al., 2014). To determine whether infants can detect, extract and use intersensory relations in a more sophisticated way (e.g., phonetic and phonological information), the modalities have to be presented separately, that is, sequentially (Lewkowicz, 2014). This sequential intermodal presentation (SIP) has been suggested as the most promising design to get insights into the processing mechanisms of stimulus perception and intersensory matching (Guihou & Vauclair, 2008). Recently, several studies have applied this procedure in the field of audio-visual speech perception (Kubicek et al., 2014; Lewkowicz & Pons, 2013; Pons, Lewkowicz, Soto-Faraco, & Sebastián-Gallés, 2009). For instance, using this procedure, 6-month-old English- and Spanish-learning infants have been shown to match sequentially presented auditory and visual syllables like /*ba*/ and /*va*/, indicating that temporal synchrony is not necessary (Pons et al., 2009). However, at 11 months of age only the English-learning infants still matched the visual and auditory input appropriately. The authors concluded that the homophonic character of /b/ and /v/ in the Spanish language leads the older Spanish-learning infants to fail to perceive this phonological contrast.

This phenomenon is called *perceptual narrowing* and describes the tendency of infants to maintain or refine perceptual abilities according to their native language attributes, while the discrimination of non-native attributes declines (Scott, Pascalis, & Nelson, 2007). This phenomenon does not only emerge in the field of language acquisition such as auditory language discrimination (Bosch & Sebastián-Gallés, 1997; Nazzi, Jusczyk, & Johnson, 2000), phonetic differentiation (Kuhl, Tsao, & Liu, 2003), visual language discrimination (Weikum et al., 2007) and audio-visual syllable matching (Pons et al., 2009). It is also well established in face discrimination (Kelly et al., 2007; Pascalis, de Haan, & Nelson, 2002) and face-voice perception across species (Lewkowicz & Ghazanfar, 2006). It is important to mention that this narrowing does not end up in a persistent loss of this function, but rather in a reorganization (Maurer & Werker, 2014; Werker & Tees, 2005). Initially, infants are broadly open to all kinds of language input due to the capacity of their developing brain; they link multisensory cues based on shared statistical characteristics (e.g., location, timing, intensity; Lewkowicz, 2014; Murray, Lewkowicz, Amedi, & Wallace, 2016). This enables them to match a variety of non-specific auditory and visual information (not only human but also simian audible and visible speech sounds), before it paves the way for more sophisticated multisensory representations, eventually becoming specific to their native language through their daily experience.

This phenomenon has been studied extensively in the context of segmented speech (syllables). However, in their daily life, infants are confronted with fluent speech and not only speech segments. Presented with more ecologically valid stimuli, 10- to 12-month-old English-learning infants looked longer at the non-native (here: Spanish) mouth movements, although they previously listened to English fluent speech, indicating a novelty preference (Lewkowicz & Pons, 2013). In contrast, 6- to 8-month-old infants showed no preference at all. However, this study contained some questionable methodological issues. For instance, the authors chose short familiarization trials of 20 seconds per trial, which has later been shown to be too short for infants at that age (Kubicek et al., 2014). Additionally, the use of 6- to 8-month-old infants as one broad age group can be called into question. This becomes essential, when considering that another study demonstrated that 6-month-old but not 8-month-old infants were able to detect relevant visual cues to discriminate visually presented speech (Weikum et al., 2007).

Although some research determined the time of origin of this phenomenon in the speech domain emerging later during the second half of the first year for syllables (Pons et al., 2009) as well as for fluent speech (Lewkowicz, Minar, Tift, & Brandon, 2015; Lewkowicz & Pons, 2013), empirical evidence points towards an earlier development of perceptual narrowing in fluent speech when reconsidering some methodological issues. A recent study elaborated these issues and provided evidence for 4.5-month-old German infants to be able to audio-visually match sequentially presented native (German) as well as non-native fluent speech (French), while 6-month-old German infants only audio-visually matched their native language and kept looking at chance level after listening to the non-native language (Kubicek et al., 2014). The authors interpreted this familiarity preference and the change in perception as an indication of perceptual narrowing emerging between 4.5 and 6 months of age.

However, speech perception might vary according to the distance between the languages; hence, it is important to take this variable as well into account (Mehler et al., 1988; Nazzi et al., 2000). Traditionally, languages have been classified into three categories, according to their predominant rhythmic structure (Abercombie, 1967; Pike, 1945); syllable-timed languages (e.g., French, Italian and Spanish), stress-timed languages (e.g., English, Swedish and German) and mora-based languages (e.g., Japanese). More recently, it has been argued that languages are better positioned along a continuum (Beckman, 1992; Dauer, 1983). In line with studies examining the relative proportions of their vocal and consonant intervals (Grabe & Low, 2002; Nazzi, Bertoncini, & Mehler, 1998; Ramus, Nespor, & Mehler, 1999), languages may be described as stress-timed (e.g., German, English) if they have shorter vocalic intervals and high variability in the duration of consonant bundles, as syllable-timed (e.g., French, Spanish) if they have intermediate values for the proportion of vocalic intervals and for consonant cluster variability, and as mora-timed (e.g., Japanese) if they have longer vocalic intervals and low variability in the duration of consonant bundles (Kubicek, Gervain, Lœvenbruck, Pascalis, & Schwarzer, 2018). Considering languages that come from the same rhythm class in the frame of perceptual narrowing processes provides insights into the important question of which factors are guiding the infants attention and hence leading them to narrow down to their mother tongue.

With respect to prosodic cues, the existing body of research has shown that young infants are able to differentiate speech from prosodically distant languages even before birth (DeCasper & Spence, 1986). This provides evidence that foetuses are able to hear by the third trimester; as newborns, they differentiate different-rhythm-class languages relying on prosodic cues like rhythm and intonation (Mehler et al., 1988). At about 4–5 months of age, they are able to discriminate their mother tongue even from other same-rhythm-class languages, for instance Spanish and Catalan (Bosch & Sebastián-Gallés, 1997) or British and American English (Nazzi et al., 2000). At the same time, their ability to differentiate non-native languages declines. While the aforementioned studies focused on auditory cues, fewer researchers took visually perceivable speech properties for language discrimination into account, even though they contribute substantively to our language identity (Munhall & Vatikiotis-Bateson, 2004). Adults are able to discriminate two languages from the same rhythm class (Spanish and Catalan) by watching a sequence of separately presented faces articulating sentences either in their native or a non-native language silently (Soto-Faraco et al., 2007). They distinguish them correctly, provided one of them is native. As early as 4 and further with 6 months of age, monolingual English-learning infants are able to detect relevant visual information to discriminate between two visual speeches from different-rhythm-class languages (English and French; Weikum et al., 2007). At 8 months of age, only bilingual infants still succeeded to do so, while monolingual infants had lost this ability. Regarding visually presented same-rhythm-class languages, female 6-month-old German infants have been shown to distinguish them (English and German; Kubicek et al., 2018). Taken together, when it comes to sensitively processing information from more than one modality, articulatory features, finding expression in subtle jaw, lip and cheek movements have to be considered as well in terms of perceptual salience (Munhall & Vatikiotis-Bateson, 2004; Rosenblum, 2008; Rosenblum, Schmuckler, & Johnson, 1997).

The above-mentioned studies on intersensory matching (audio-visual) only refer to different-rhythm-class languages, varying in their overall prosodic characteristics (English and Spanish, Lewkowicz & Pons, 2013; or German and French, Kubicek et al., 2014). Only one cross-linguistic study compared same-rhythm-class languages (German and Swedish) that differ only in subtle language properties (phonetic and phonological attributes; Dorn, Weinert, & Falck-Ytter, 2018). The authors presented audio-visual fluent speech stimuli sequentially and provided evidence that German and Swedish 4.5-month-old infants are able to extract, remember and integrate these fine-grained audio-visual speech cues in their native as well as in a non-native language. The 4.5-month-old infants looked longer to the mouth movements that referred to the language they previously listened to. However, no study has shed light on the trajectory of perceptual narrowing in a multimodal context for same-rhythm-class languages. As already mentioned, comparing same-rhythm-class languages, such as German and Swedish that both belong to the stress-timed languages but differ in auditory and visually perceivable phonetic and phonological features (Dorn et al., 2018; Lindqvist, 2007), provides insights into the important question which factors are guiding the infants attention and lead them to narrow down to their mother tongue. After examining the ability of 4.5-month-old infants to extract subtle language properties and match audio-visual speech cues of their native and a non-native language (Dorn et al., 2018), the aim of the present study is to track the subsequent perceptual narrowing towards the infants' native language.

Hence, we investigated the trajectory of infants' ability to process, extract and integrate subtle audio-visual language properties in same-rhythm-class languages, such as German and Swedish that differ mainly in phonetic and phonological attributes. The specific purpose of our study was to extend the aforementioned empirical findings of 4.5-month-old infants in the study of Dorn et al. (2018) longitudinally, in order to now examine subsequent perceptual narrowing processes in fluent speech processing at 6 months of age. We used the same method and age groups as in the study of Kubicek et al. (2014), who found empirical evidence for 6 months of age as a critical time point for the emergence of perceptual narrowing in different-rhythm-class languages. We aimed to replicate these results and extend them to same-rhythm-class languages. Specifically, we presented German and Swedish infants (cross-linguistic design) side-by-side videos of German and Swedish silently talking faces articulating semantically identical speech streams before and after they listened to one of the languages, first at 4.5 and then at 6 months of age.

Infants watched and listened sequentially to side-by-side presentations of visual mouth movements and corresponding auditory fluent speech in their respective native or a non-native language. During the presentation, we recorded how long the infants looked to the audio-visually matched face before (baseline) and after (test phase) they listened to one of the two languages. In comparison to the first measurement point at age 4.5 months, we expected the 6-month-old infants to still show an attentional shift between the native and the non-native silently talking face after listening to their respective native language, indicating audio-visual matching abilities, but to keep looking at chance level after listening to the respective non-native language.

2 | METHOD

2.1 | Participants

Participants were recruited in Germany and Sweden.[1] The sample consisted of 45 German-learning infants (female = 24) and 37 Swedish-learning infants (female = 19) who were tested twice in our labs (*names of the labs masked to preserve blinding*), first at 4.5 months and again at 6 months of age. The respective characteristics of the German and Swedish samples are listed in Table 1. As reported by the parents, all infants were full term (38–41 gestation weeks) without any visual or auditory impairment. The data of 14 additional infants in the longitudinal setting were excluded from the analyses, due to too little looking time (criteria adapted from Dorn et al., 2018; Kubicek et al., 2014; baseline <7.5 s at each of the two faces; test phase <3 s at each of the two faces, n = 9), fussiness (n = 3), technical failure (n = 1) and parental influence (n = 1) at either the first or the second visit. Informed written consent was obtained from the respective parent of each infant before any assessment or data collection. The experiment was conducted according to the guidelines laid down in the *Declaration of Helsinki*. All procedures were conducted according to the regulations of the Institutional Review Boards of the *Deutsche Forschungsgemeinschaft* (DFG) and the *Deutsche Gesellschaft für Psychologie* (DGPs) in Germany and the *Regional Ethics Board* in Stockholm in Sweden. Prior to testing, we asked the parents which language they usually used at home and whether the infants had regular contact with individuals speaking another language. Hence, the sample only consisted of monolingual German-or Swedish-learning infants, respectively.

2.2 | Stimuli

We recorded the stimuli at the (name masked to preserve blinding). Visual stimuli were silent video clips of two bilingual adult females – German-Swedish. The women sat in front of a black background, looking directly at a camera with a neutral facial expression. They recited Swedish and German common and semantically identical sentences in a shortened and repeated manner (3 × 10 s episode of utterances – German: “Hallo mein Baby, geht es dir gut? Du

TABLE 1 Characteristics of the German and Swedish 4.5- and 6-month-old samples, considered separately and together

Sample	Age (months)	*N*	Gender (female/male)	*M*	*SD*	Range (days)
German	4.5	45	24/21	139.31	5.33	128–154
	6			184.91	5.37	175–199
Swedish	4.5	37	19/18	138.54	2.72	133–146
	6			182.68	3.97	176–191
German and Swedish	4.5	82	43/39	138.96	4.35	128–154
	6			183.9	4.89	175–199

bist ein hübsches Baby! Wie schön dich zu sehen. Bis bald!", Swedish: "Hej mitt barn, hur mår du? Du är ett vackert barn! Vad trevligt att se dig. Vi ses snart!"). These sentences have been adapted from Kubicek et al. (2014) and were previously used in the study of Dorn et al. (2018). With the aid of a teleprompter, we ensured that the speech rate of the two women was matched in both languages. All videos were matched in size and duration according to the original study (Kubicek et al., 2014). Each of the 30-s video clips presented a full-face image of the respective woman and measured 20.6 cm x 18 cm. The two simultaneously played videos were separated by an 11 cm gap. Both videos, Swedish and German, were edited as to make sure that both started on a closed mouth whereupon the first mouth opening was synchronized. The auditory stimuli corresponded to the extracted 30-seconds soundtracks from the video recordings, resulting in two different voices both either speaking Swedish or German. Sound was presented at conversational sound pressure level (65 dB ±5 dB).

2.3 | Procedure and apparatus

We tested each German infant individually in the (name masked to preserve blinding) and each Swedish infant individually at the (name masked to preserve blinding). The environmental settings were kept constant, for example, size of the room, lighting conditions, screen size). The infant was sitting on the lap of the parent and the parent was instructed not to point at the screen, nor to talk to the infant or get into contact unless signs of distress were appearing. To avoid potential parental influence on the infant's looking behaviour, they were instructed to wear sunglasses, so that the eyetracker would not detect the parents' gaze, and to wear headphones. Infants were placed at approximately 60 cm to the 24-in. monitor (resolution: 1920 × 1,080 pixels). Stimuli were presented with *Tobii Studios software* (Tobii Technology, Sweden) while the eye-tracking data were captured by a *Tobii X60* eye tracker with a sampling rate of 60 Hz. In order to check the videos afterwards for any distracted behaviours (too less looking time, fussiness or parental influence), we used an additional video camera (specialized for low light conditions, type *Logitech*) above the screen. Before the video started, infants completed an infant adapted five-point calibration. Calibration was checked for accuracy – at least three of the five points on each eye were required for the calibration to be deemed as valid. If necessary, the calibration was repeated up to three times.

After showing a calibration video (star moving to five positions on the screen) to later evaluate the accuracy of the recorded eye movements, an attention grabber (walking penguin with a sound) appeared and finally the *intersensory matching procedure* started (Dorn et al., 2018; Kubicek et al., 2014; Pons et al., 2009). In this procedure, the auditory and the visual stimuli were presented sequentially in order to ensure that audio-visual synchrony, such as temporal cues, are not responsible for the matching abilities. The procedure consisted of six trials, each lasting 30 s (see Figure 1). The first two represented the baseline condition (60 s in total) in which the infants saw two side-by-side silent video clips with one bilingual woman speaking the semantically identical utterances in Swedish on one side and in German on the other side. The position of the language appearing on the screen was reversed in the second trial. The third trial outlined the auditory familiarization trial, where the infants listened to the utterances while a static yellow circle appeared on the screen. The infants were randomly assigned to either the native or the non-native auditory familiarization group (German infants: native auditory familiarization N = 21, non-native auditory familiarization N = 24; Swedish infants: native auditory familiarization N = 16, non-native auditory familiarization N = 21). The test phase started in the fourth trial, where the initial silent videos were presented again. The fifth and sixth trials displayed a repetition of trials 3 and 4, with reversed face position. This split test procedure seeks to avoid any side preferences and justifies these two test trials (Kubicek et al., 2014; Lewkowicz & Pons, 2013; Pons et al., 2009). The whole procedure lasted 2 min in total (each familiarization and test phase, respectively, 30 seconds and performed twice).

Based on the assumption that infants directly match the previously heard speech with the corresponding silently talking face (Kubicek et al., 2014), each auditory trial preceded a visual trial in the test phase. To exclude any side preferences, the position of the first language appearing on the left side was counterbalanced across the infants as well as the female bilingual women. Notably, the woman the infants listened to during the auditory familiarization

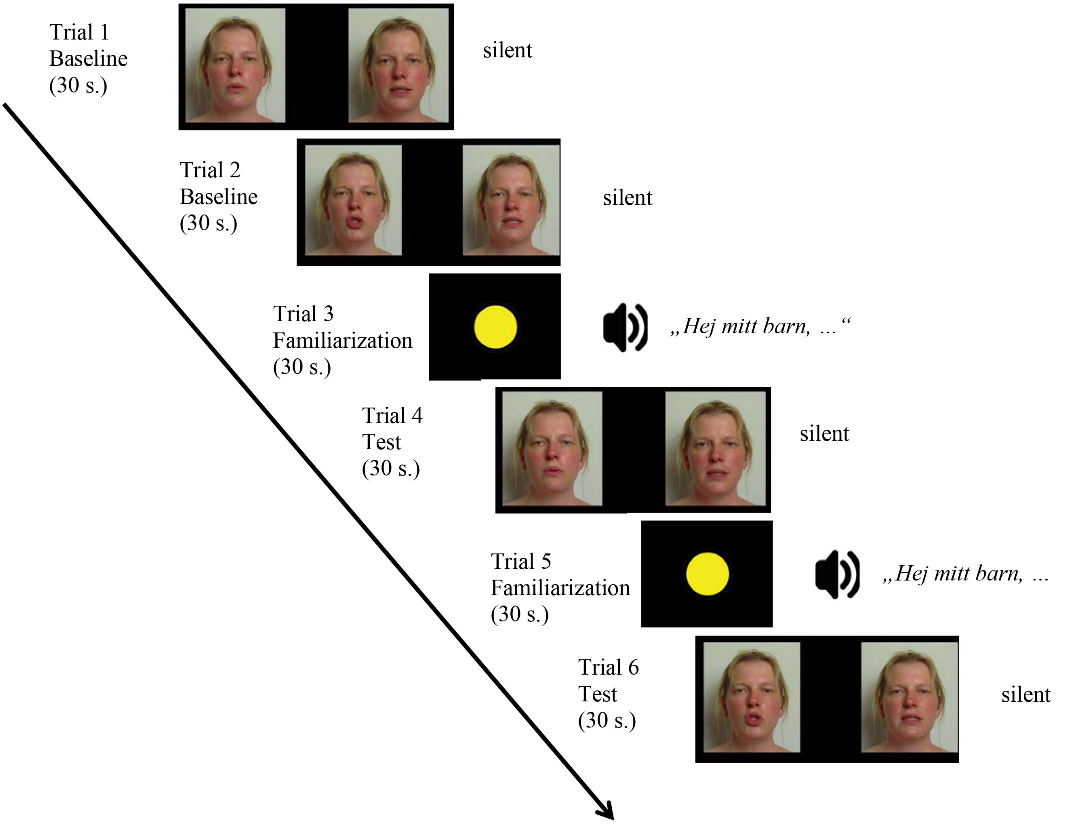

FIGURE 1 Schematic representation of the *intersensory matching procedure*. Only the Swedish familiarization condition is shown. The model has given written informed consent to publication of herphotograph

trials (third and fifth) was different from the woman they saw during the silent videos—baseline phase (first and second) as well as the test phase (fourth and sixth). This procedure ensured that any cross-modal preference was not due to any idiosyncratic aspects (e.g., pronunciation, facial expression) of the particular woman in one of the languages (Lewkowicz & Pons, 2013). We broadened this precaution by means of presenting two different women instead of one to limit the possible idiosyncratic aspects the bilingual women might have in only one of the languages (Kubicek et al., 2014).

2.4 | Data analysis

To determine how much time an infant spent looking at each of the two faces, we created two principal areas of interest (AOIs), one framing the left and the other framing the right face on the screen. Every AOI covered one half of the screen to be comparable to the study of Kubicek et al. (2014), which used hand-coding to distinguish only between left and right looking behaviour. To be considered in the final analyses, we adopted the same criteria as in the study of Dorn et al. (2018). Every infant had to look at each of the two faces for a minimum duration of 7.5 seconds during the baseline trials (criteria adapted from Dorn et al., 2018; Kubicek et al., 2014). When summarized over both baseline trials, this total amount of seconds resulted in at least 25% of the total looking time during baseline. Furthermore, every infant had to look at each of the two faces for a minimum duration of 3 s during the test phase. When summarized over both test trials, this total amount resulted in at least 10% of the total looking time during the test phase. Both criteria assured that the infants have processed both visual languages. Nine infants did not meet this criterion and consequently were not considered in the analyses.

The dependent variable was the *proportion of total looking time score* (PTLT-score), that is the looking time to the face that corresponded to the previously heard language, divided by the looking time to both faces. This measure was obtained in both, the baseline and the test phase respectively for the German and Swedish auditory familiarization. In the baseline phase, the infants had not yet heard the audio; hence, chance level performance was expected (50%). In the test phase, infants had heard the auditory signal before; visual preference could therefore potentially be affected by the language they had listened to. The change of looking behaviour between these two phases is important since it indicates the effect of the auditory input on the looking time to the corresponding face. To account for the looking duration, these scores were converted into percentages. To be considered significant, we used an alpha level of .05 for the statistical analyses.

Since preliminary analyses did not reveal any significant effects of infants' gender ($p > .22$), nor of the woman's identity ($p > .36$), nor of the position of the visual language (either Swedish or German first appearing on the left side; $p > .51$) in the German as well as in the Swedish sample for each age group, these three factors were excluded from the following analyses. No side bias could be detected, when considering all infants together ($p > .86$) nor when considering German and Swedish samples separately ($p > .81$).

3 | RESULTS

3.1 | Face preference during baseline and test phase

We checked whether the infants initially preferred one of the two silent talking faces during baseline, by calculating one-sample t-tests against chance level (50%) for both age groups and both samples. Overall, neither the 4.5-month-old infants (German infants: $M = 50.84$, $SD = 13.18$, $t(44) = 0.43$, $p > .05$; Swedish infants: $M = 53.15$, $SD = 10.59$, $t(36) = 1.81$, $p > .05$) nor the 6-month-old infants (German infants: $M = 51.05$, $SD = 8.94$, $t(44) = 0.79$, $p > .05$; Swedish infants: $M = 51.68$, $SD = 10.82$, $t(36) = 0.94$, $p > .05$) showed any preference during baseline. But as we checked additionally for any baseline preference for the later heard language, the analyses revealed that only the 6-month-old Swedish infants who later heard German already preferred the German visual language at baseline ($M = 55.87$, $SD = 9.91$, $t[20] = 2.72$, $p < .01$), while the other samples did not prefer any language (Table 2).

TABLE 2 Mean of preference scores (%) towards the German visual speech stream (silent-talking face) during baseline

Age (months)	Sample	Auditory familiarization	*M*	*SD*	*t* Value (test vs. chance)	*p*-Value
4.5	German	Native	46.98	15.11	−0.92	$p > .05$
		Non-native	54.20	10.41	1.98	$p > .05$
	Swedish	Native	54.50	9.94	1.58	$p > .05$
		Non-native	49.08	9.35	−0.45	$p > .05$
6	German	Native	52.67	8.15	1.50	$p > .05$
		Non-native	49.64	9.52	−0.19	$p > .05$
	Swedish	Native	46.17	9.65	−1.59	$p > .05$
		Non-native	55.87	9.91	2.72	$p < .01^{**}$

$^{*}p < .05$;
$^{**}p < .01$;
$^{***}p < .001$.

To further see whether during the test phase the infants preferred to look at the matching visual speech stream significantly more or less than what was expected by chance, we calculated one-sample *t*-tests against chance level (50%, see Table 3). At the measurement point of the 4.5-month-old infants, only the Swedish infants who listened to the non-native language showed a significant looking behaviour above chance level to the corresponding German visual speech during test phase ($M = 56.68$, $SD = 12.12$, $t[21] = 2.53$, $p < .05$). The German infants who listened to their native language only showed a marginally significant looking behaviour above chance level to the corresponding German visual speech during test phase ($M = 56.84$, $SD = 17.47$, $t[20] = 1.8$, $p < .10$). Neither the Swedish infants who listened to their native language showed a significant looking behaviour above chance level to the corresponding Swedish visual speech ($M = 54{,}04$, $SD = 20.59$, $t[20] = -.79$, $p > .05$), nor the German infants who listened to the non-native language showed a significant looking behaviour above chance level to the corresponding Swedish visual speech ($M = 54.08$, $SD = 16.29$, $t[20] = -1.23$, $p > .05$), showed a significant looking behaviour above chance level during test phase. At the measurement point of the 6-month-old infants, all infants who listened to their non-native language looked significantly above chance level to the corresponding face during test phase (German infants to Swedish visual speech: $M = 44.53$, $SD = 12.76$, $t(20) = 2.10$, $p < .05$; Swedish infants to German visual speech: $M = 40.01$, $SD = 10.37$, $t(20) = 4.42$, $p > .001$), but only the German infants who listened to their native language looked significantly above chance level to the corresponding German visual speech during test phase ($M = 59.84$, $SD = 11.60$, $t(20) = 3.89$, $p < .001$) compared to the Swedish infants who listened to their native language but did not look significantly above chance level to the corresponding Swedish visual speech ($M = 46.60$, $SD = 11.72$, $t(20) = 1.16$, $p > .05$; see Table 3).

3.2 | Audio-visual matching behaviour

To test for the hypothesized perceptual narrowing effect between 4.5 and 6 months, we calculated a mixed four-way-ANOVA on PTLT-scores as dependent variable for the whole sample with *group* (German and Swedish infants), *auditory familiarization* (native and non-native speech stream) as between-subject factors and *phase* (baseline and test phase) and *age* (4.5 and 6 months) as within-subject factors. This ANOVA revealed no main effects. As predicted, the analysis revealed an age x phase x auditory familiarization interaction effect ($F[1,78] = 17.40$, $p < .001$, $\eta^2 = .18$), indicating that the ability to audio-visually match the language, the infants had previously listened to, depended on the age of the infants and the language they were familiarized with (see Figure 3). In addition, an age x auditory

TABLE 3 Mean of preference scores (%) towards the German visual speech stream (silent-talking face) during test phase

Age (months)	Sample	Auditory familiarization	*M*	*SD*	*t* Value (test vs. chance)	*p*-Value
4.5	German	Native	56.84	17.47	1.8	$p < .10^{+}$
		Non-native	45.92	16.29	−1.23	$p > .05$
		Native	45.96	20.59	−.79	$p > .05$
	Swedish	Non-native	56.68	12.12	2.53	$p < .05^{*}$
6	German	Native	59.84	11.60	3.89	$p < 001^{***}$
		Non-native	55.47	12.76	2.10	$p < .05^{*}$
		Native	53.40	11.72	1.16	$p > .05$
	Swedish	Non-native	59.99	10.37	4.42	$p < .001^{***}$

$^{+}p < .10$;
$^{*}p < .05$;
$^{**}p < .01$;
$^{***}p < .001$.

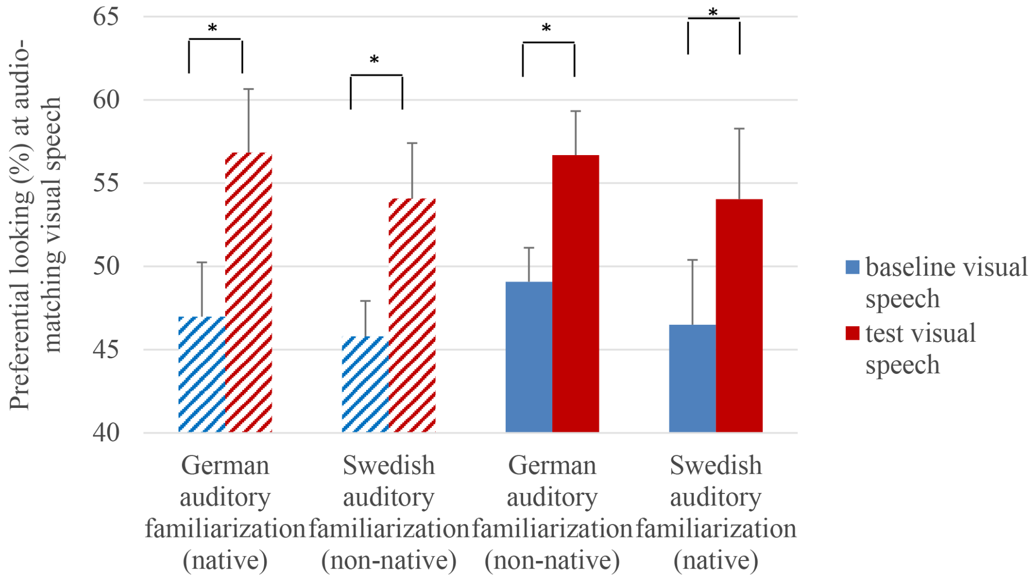

FIGURE 2 Audio-visual matching in 4.5-month-old German (striped) and Swedish infants (solid) after German and Swedish auditory familiarization, respectively. Mean of preference scores at the audio-matching visible speech during baseline (blue bar) and test trials (red bar). Error bars indicate the standard error of the mean

familiarization interaction effect emerged ($F[1,78] = 10.87$, $p < .01$, $\eta^2 = .12$), indicating that depending on age, the infants perceived the auditory familiarization differently. In contrast to the 4.5-month-old measurement point when the infants increased their looking time to the respective audio-matching visual speech (German and Swedish infants considered together as one sample – German auditory familiarization: $M = 48.03$; $SD = 12.45$ to $M = 56.76$; $SD = 14.85$; Swedish auditory familiarization: $M = 44.08$; $SD = 10.32$ to $M = 54.06$; $SD = 17.88$), at the 6-month-old measurement point, the German infants increased their looking time after listening to their native language ($M = 52.67$; $SD = 8.15$ to $M = 59.84$; $SD = 11.60$), whereas Swedish infants decreased their looking time after listening to their native language ($M = 53.83$; $SD = 9.65$ to $M = 46.60$; $SD = 11.72$). Instead of a matching pattern (familiarity effect), the last subgroup showed a mis-matching pattern (novelty effect). Additionally, no preference was found after listening to the non-native language ($p > .05$). Furthermore, an age x phase interaction effect ($F[1,78] = 9.15$, $p < .01$, $\eta^2 = .11$) was found. This interaction raised from the fact that at the 4.5-month-old measurement point the infants matching of the respective auditive input and the visual face increased from baseline to test phase (see Tables 2 and 3 for detailed numbers and 4 for distinct differences scores among the age groups), whereas at the 6-month-old measurement point not much change could be found between the baseline and the test phase (see Tables 2 and 3 for detailed numbers and 4 for distinct differences scores among the age groups).

To further clarify the *three-way interaction age x phase x auditory familiarization* and to determine whether the infants prefer the audio-matching mouth movement after they were familiarized with either native or non-native speech, we calculated paired two-tailed t-tests. Here, we compared the looking preference to the audio-matching visual speech (mouth movements) during baseline to the one during test trials. The results for the 4.5- and 6-month-old infants are illustrated in Figures 2 and 3, respectively, each figure separated by language group (German or Swedish). After listening to either their native or a non-native language, the 4.5-month-old infants looked longer to the respective visual speech afterwards (German infants: German auditory familiarization ($t(20) = 2.24$, $p < .05$; $d = .60$); Swedish auditory familiarization ($t[23] = 2.97$, $p < .05$; $d = .58$); Swedish infants: German auditory familiarization ($t(20) = 2.54$, $p < .05$; $d = .70$); Swedish auditory familiarization ($t[15] = 2.58$, $p < .05$, $d =$

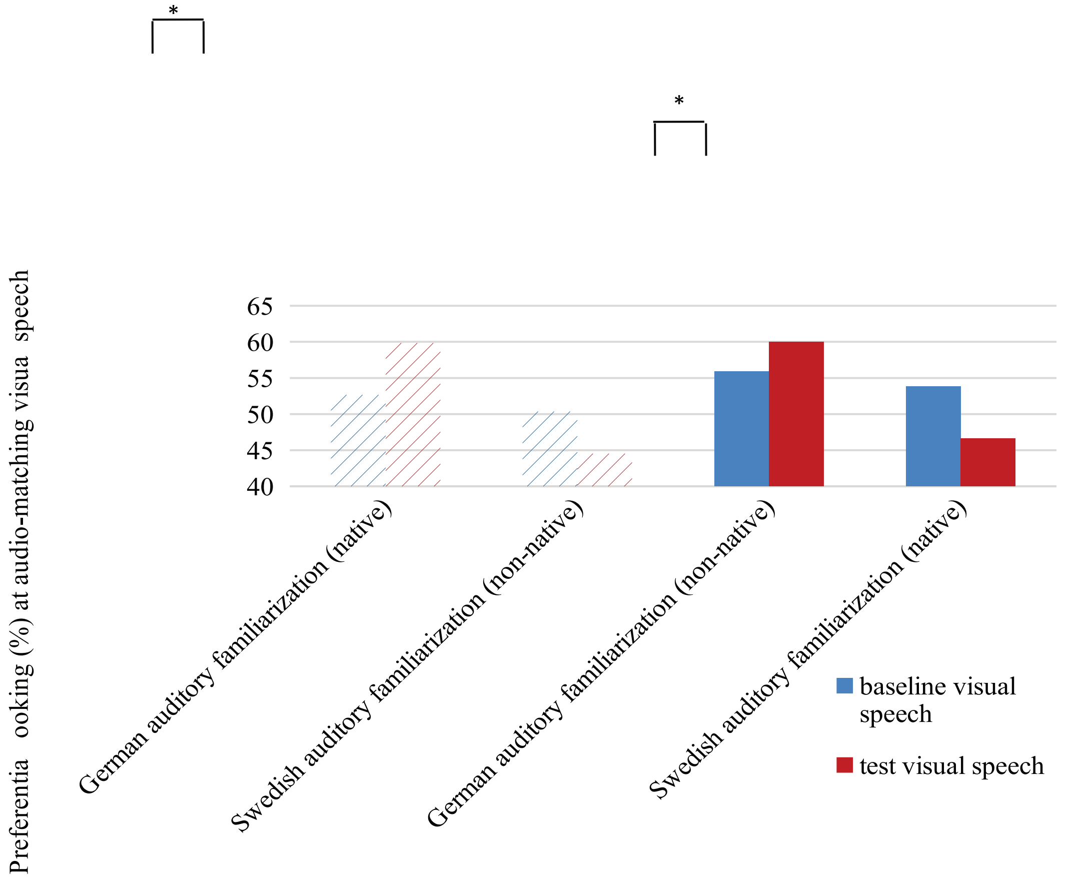

FIGURE 3 Audio-visual matching in 6-month-old German (striped) and Swedish infants (solid) after German and Swedish auditory familiarization, respectively. Mean of preference scores at the audio-matching visible speech during baseline (blue bar) and test trials (red bar). Error bars indicate the standard error of the mean

.73; Figure 2). This supports the assumption that 4.5-month-old infants are able to audio-visually match the language they previously listened to.[2] If perceptual narrowing occurs between 4.5 and 6 months, 6-month-old infants should display a different pattern – that is, they should differentiate between the native and the non-native visual speech by shifting their visual attention correspondingly after having listened to their respective native language, but keep looking at chance level after having listened to the non-native language. The results support this assumption: While both groups looked equally long on both faces after having listened to the non-native language (German infants: Swedish auditory familiarization ($t[23] = 1.84$, $p = .08$); Swedish infants: German auditory familiarization: ($t(20) = 1.3$, $p = .21$), both groups shifted their attention thereby differentiating the two faces after having been presented with their respective native language (German infants: German auditory familiarization ($t(20) = 2.60$, $p < .05$; $d = .73$); Swedish infants: Swedish auditory familiarization: ($t[15] = -2.26$, $p < .05$, $d = .69$, Figure 3). Although different patterns of preferences were observed (familiarity and novelty effect), discrimination abilities are only reported after listening to the respective native language, while looking behaviour stays at chance level after listening to the respective non-native language, pointing to perceptual narrowing. Furthermore, we calculated change scores, using the advantage of our longitudinal perspective. We presented the within-infant developmental changes from baseline to test phase for each infant (PTLT_baseline—PTLT_test phase) and separated the data by auditory familiarization as can be seen in Figure 4.

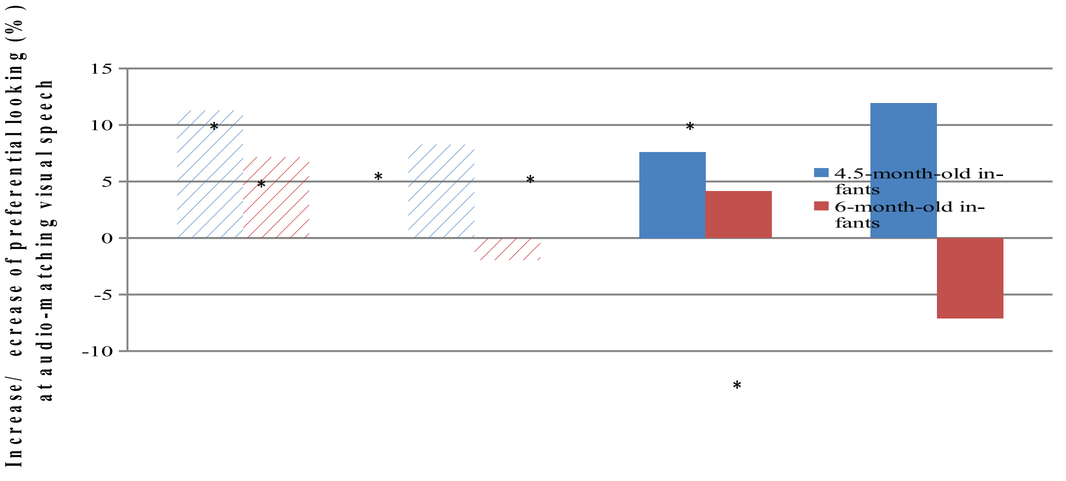

FIGURE 4 Difference scores from baseline to test phase of the 4.5- (blue) and 6-month-old infants' (red) preferential looking at the audio-matching visual speech of the German (striped) and Swedish infants (solid) after German and Swedish auditory familiarization, respectively. Error bars indicate the standard error of the mean

4 | DISCUSSION

Our study aimed to trace the trajectory of the infants' ability to process, extract and integrate subtle audio-visually perceivable language properties in same-rhythm-class languages (German and Swedish) to test the multisensory perceptual narrowing in fluent speech. In a cross-linguistic design, we tracked the gaze pattern of German and Swedish infants longitudinally, first at 4.5 months and then at 6 months of age. By using an *intersensory matching procedure*, infants watched and listened sequentially to side-by-side presentations of visual mouth movements and corresponding auditory fluent speech in their native or a non-native language both of which belonged to the same rhythm-class.

In agreement with our hypothesis, in both samples of 6-month-old infants, the looking pattern remained at chance level after listening to the respective non-native language, while they discriminated the two silent mouth movements after listening to their respective native language. In comparison with previous empirical findings of Dorn et al. (2018) and our present data that provided evidence for the ability of 4.5-month old infants to audio-visual match a non-native language, this points to a perceptual narrowing phenomenon, that is to say, the decline in discriminating non-native attributes (Maurer & Werker, 2014; Scott et al., 2007). The second part of this phenomenon refers to the maintaining or refining of the perceptual abilities with regard to the respective native language attributes. The results showed that both samples shifted their attention and preference concerning native and non-native mouth movements after listening to their native language, indicating audio-visual matching abilities. Nevertheless, considering the attentional shift in more detail an unexpected result occurred in the present study. Whereas the German-learning infants increased their looking time to the German mouth movements after listening to their native language (expected familiarity effect), the Swedish-learning infants decreased their looking time to the Swedish mouth movements after listening to their native language (unexpected novelty effect). The former familiarity effect replicates the previous finding of Kubicek et al. (2014) in 6-month-old infants and extends them from different- to same-rhythm-class languages. The authors have shown that 6-month-old infants prefer to look at the native visual speech they previously listened to (familiarity preference) and looked at chance level after listening to a non-native speech. But what about the latter finding of a novelty preference? At first glance, it seems to be contradictory; but note that any divergence from random looking behaviour is indicative of the infants' ability to discriminate the presented stimuli (Houston-Price & Nakai, 2004). Especially in the field of multisensory and visual perception literature, a novelty effect is neither new nor rare (Gottfried, Rose, & Bridger, 1977; Pascalis et al., 2002). Such a novelty effect has also been shown in 10- to 12-month-old English-learning infants in the same intersensory matching procedure; these infants had been familiarized with English utterances but looked longer at the non-native, non-matching Spanish visual speech (Lewkowicz & Pons, 2013). The authors assumed that perceptual narrowing might have occurred since the infants only performed this gaze pattern in response to their native speech, as it is the case in our present study. In contrast to the overall audio-visual matching abilities at 4.5 months of age (Dorn et al., 2018; Kubicek et al., 2014), this looking pattern indicates that the infants pass through an initial stage of being broadly open to all kinds of language input (Kuhl, 2004). This might be due to structural and functional immaturity before it paves the way for more sophisticated multisensory representations, becoming more and more attuned towards their native language attributes, driven by their daily experience (Lewkowicz, 2014; Murray et al., 2016).

In order to understand the guiding factor(s) of a specific novelty preference of Swedish 6-month-old, it might be helpful to have a closer look at the special environmental conditions. Remarkably, we only found one baseline preference in our study when considering the later heard language, namely the Swedish 6-month-old infants who already preferred the German visual speech. That exactly this group afterwards still prefers the German face, although they listened to their native Swedish language, might point to specific acoustic characteristics that might have attracted the Swedish infants' attention more to the German visual speech. Supporting this line of reasoning, the other Swedish 6-month-old infants who later heard German also tend to look longer to the German face during baseline, albeit not significantly. For instance, more vowels produced with lip protrusion might be more salient and attractive for the infants, leading to more attention to the respective stimuli (Kubicek et al., 2014). Especially, the Swedish language is, among other language features, characterized by long vowels tending to diphtongizations (e.g., /*e*/ is pronounced like an /*ea*/) or particular lip roundings such as pursed lips that does not exist in the German language (e.g., /*u*/ more like a compound of /*i*/ and /*ü*/; see Lindqvist, 2007 for a review). These examples for visemes, might explain how infants can distinguish between the two visual speeches (for more linguistic analyses see Lindqvist, 2007). This existence of long vowels and their interplay with consonants might display a great amount of multiple and concurrent sensory cues, the infant may draw on in terms of early language recognition and discrimination. After the infants gained prenatal listening experience in utero (DeCasper & Spence, 1986) as well as postnatal listening experience with their native language, different responses to these visual speeches in the German and Swedish samples might have been evoked. Why only the Swedish 6-month-old sample was more attracted to the German silent-talking face needs to be further examined by analysing specific acoustic characteristics of these two languages (e.g., pitch changes, syllable duration, mouth openings). All in all, similar asymmetrical effects, that is to say, different gaze pattern preferences in several subsamples of infants, have been interpreted to be indicative of language discrimination (e.g., Bosch & Sebastián-Gallés, 1997, 2001; Molnar, Gervain, & Carreiras, 2014; Moon, Cooper, & Fifer, 1993). For instance, monolingual Basque and bilingual Basque-Spanish 3.5-month-old infants discriminated Spanish and Basque in a visual habitiation paradigm independent of the language they were familiarized with, while monolingual Spanish infants only discriminated the languages after listening to Basque (Molnar et al., 2014). The authors interpreted both outcomes as showing discrimination abilities and reasoned that the infants' behaviour reflects a possible overhearing of an L2 in the first months of life that alters their language-processing skills (either recognition or discrimination). Sweden is often considered as a kind of bilingual nation and characterized by statements that there is no common language or that from birth onward, the young people are bilingual in some way, despite they are born in Sweden (Johansson, Davis, & Geijer, 2007; Lindberg, 2007). This might point to a diverse linguistic background in the Swedish infants that could have evoked a different pattern of preference. Future studies may examine the influence of a diverse linguistic background (overheard L2) and add further cognitive measures such as pupil dilation, in order to examine this distinct looking pattern and the associated cognitive processes more precisely.

Either specific acoustic characteristics that in particular attract the Swedish infants' attention or the diverse linguistic background – we finally argue in support of an indication of perceptual narrowing. The Swedish infants only demonstrated this gaze pattern in response to their native speech, just as the infants in the study of Lewkowicz and Pons (2013) did. In contrast, after listening to a non-native speech their looking pattern remained at chance level.

Hence, the perceptual narrowing phenomenon is constituted by this distinctive looking behaviour (Scott et al., 2007). Generally, it is of crucial interest to consider both directions as possible discrimination evidence and interpret the looking behaviour separately (Houston-Price & Nakai, 2004). However, one limitation of our study results in the impossibility to draw a conclusion on why only 6-month-old Swedish infants enrolled in the German familiarization are affected differently in their looking/recognition pattern. For this reason, it is of importance to interpret these results cautiously and to further analyse the speech characteristics of these two languages and the effect of a diverse linguistic background more precisely to figure out the guiding factor leading to these (different) looking patterns.

Despite most research providing evidence that perceptual narrowing in the speech domain appears later in the first year of life (Lewkowicz et al., 2015; Maurer & Werker, 2014; Pons et al., 2009), our findings lend support to the empirical results of Kubicek et al. (2014) which showed that under specific circumstances (e.g., fluent prosodically rich speech), this tuning process might emerge earlier and within the first 6 months of life. Due to the use of different levels of cues, for instance Hindu syllables (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1984) homophone syllables (Pons et al., 2009), phonetic continuum of speech sounds (Maye, Werker, & Gerken, 2002) or even fluent speech (Kubicek et al., 2014), it is not surprising that various studies set the emergence of perceptual narrowing differently. Our stimuli consisted of fluent ecological audio-visual speech, characterized as prosodically rich, lively and common in everyday life that possess multiple and concurrent sensory cues. Hence, infants seem to benefit from these various multisensory cues in this situation. In particular, it is important to mention that vowels have been identified to induce an earlier emergence of perceptual narrowing (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). The Swedish language possesses long vowels tending to diphtongizations (e.g., */e/* is pronounced like an */ea/*) or particular lip roundings such as pursed lips that does not exist in the German language (e.g., */u/* more like a compound of */i/* and */ü/*; see Lindqvist, 2007 for a review). This existence of long vowels and their interplay with consonants might display a great amount of multiple and concurrent sensory cues, the infant may draw on in terms of early language recognition and discrimination.

In the studies of Lewkowicz and Pons (2013) and Kubicek et al. (2014), the infants listened to two different-rhythm-class languages, whereas in our study, the infants were presented with two same-rhythm-class languages. Thus, we strengthened and extended their findings, providing empirical indication of perceptual narrowing shortly before 6 months of age even in same-rhythm-class languages. Up to a certain timeframe, infants perceive these subtle language properties (Dorn et al., 2018), whereas afterwards they decline in processing non-native attributes due to their everyday experience. This provides further evidence for the *native language acquisition hypothesis* (Nazzi & Ramus, 2003), stating that infants learn the specific features of their native language rhythm rather than for the rhythm class as a whole. What enables infants to perform these discriminations is an innate sensitivity since birth and a growing knowledge of the features that allow them to discriminate same-rhythm-class languages. In future studies, it would be of interest to track the same infants in processing different- as well as same-rhythm-class languages to specify certain attributes of the languages that account for the specific looking patterns.

Our study highlights an important time range between 4.5 and 6 months of age, in which crucial steps occur to acquire language. Especially in the field of deaf and hearing-impaired infants, this knowledge could be important to set the starting point for interventions at the time point when infants can mostly benefit from. A growing body of literature recognizes the importance of early implantation, resulting in better overall outcome patterns for children with cochlear implants or even the possibility that these affected infants catch up with their typically developing peers (Colletti, Mandalà, Zoccante, Shannon, & Colletti, 2011; Houston, Stewart, Moberly, Hollich, & Miyamoto, 2012; Nikolopoulos, Archbold, & O'Donoghue, 1999). For instance, empirical findings revealed that deaf born children who have recovered their hearing ability with cochlear implants before 2.5 years are able to acquire adequate audio-visual speech integration later on (Schorr, Fox, van Wassenhove, & Knudsen, 2005) and infants with an implantation before 12 months of age benefited in terms of improved auditory, speech language and cognitive performances (Colletti et al., 2011). One single study provided evidence for cochlear implantation as early as 2–6 months of age being associated with improved speech perception, receptive vocabulary and speech production approximately identical to the level of normal-hearing children without more complications (Colletti, Mandalà, &

Colletti, 2012). During the first year of life infants run through a series of critical periods with respect to their phonological development and each has cascading effects on the following one. If implantation emerges after these critical periods, their brains might have already been affected by this absence of auditory stimulation (Werker & Hensch, 2015). As our present study, these findings suggest an earlier sensitive period to be of importance in starting interventions in deaf and hearing-impaired infants. Future studies should track these affected infants in their cognitive and language development to determine the most beneficial starting point of interventions, such as cochlear implantations, providing them with the best requirements for language acquisition.

5 | CONCLUSION

In conclusion, following evidence for perceiving, extracting and integrating subtle language properties on the phonological and phonetic level (Dorn et al., 2018), the present study is the first one tracing the development of multisensory perceptual narrowing processes in same-rhythm-class languages longitudinally. In a cross-linguistic design, the results provided empirical indication for this phenomenon occurring before 6 months of age, similar to different-rhythm-class languages (Kubicek et al., 2014). Nevertheless, the results have to be interpreted cautiously due to different patterns of looking preferences in the different language samples. This might have been caused by specific acoustic characteristics, potentially evoked by a diverse linguistic background in the Swedish sample. These findings could have crucial implications for the temporal benefit of cochlear implantations in infancy.

ACKNOWLEDGEMENTS

The authors would particularly like to thank Terje Falck-Ytter, who enabled a great cooperation between our two baby labs for this cross-linguistic study. Furthermore, we would like to thank the German as well as the Swedish parents who took part in our study, supporting this cross-linguistic cooperation project. In addition, we were grateful to have financial support by the funding program *IPID4all* of the Federal Ministry of Education and Research (Germany) for data collection in the Uppsala Child and Baby Lab at the Uppsala Universitet in Sweden that made this cross-linguistic data collection possible.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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

ORCID

Katharina Dorn https://orcid.org/0000-0001-5454-5143

ENDNOTES

[1] The data of the first measurement point when children were 4.5-month-old overlap with the cross-sectional study of *names masked to preserve blinding* (N = 96). That study only focused on 4.5-month-old infants' fine perception of subtle language properties to audio-visual match languages, whereas the present study examined the trajectory to 6-month-old infants and perceptual narrowing processes. Only those infants who participated at both time points (4.5 and 6 months) are included in the present study (N = 82).

[2] The data of the first measurement point when children were 4.5-month-old overlap with the cross-sectional study of *names masked to preserve blinding* (N = 96). That study only focused on 4.5-month-old infants' fine perception of subtle language properties to audio-visual match languages, whereas the present study examined the trajectory to 6-month-old infants and perceptual narrowing processes. Only those infants who participated at both time points (4.5 and 6 months) are included in the present study (N = 82).

REFERENCES

Abercombie, D. (1967). *Elements of general phonetics*, Edinburgh: Aldine Publications.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*, 190–201. https://doi.org/10.1037//0012-1649.36.2.190

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science*, *13*, 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x

Beckman, M. E. (1992). Evidence for speech rhythms across languages. In: Y. Tohkura et al. (Eds.), *Speech Perception, Production and Linguistic Structure*, 457–463.Tokyo: IOS Press.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, *65*, 33–69. https://doi.org/10.1016/S0010-0277(97)00040-1

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

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*, 204–220. https://doi.org/10.1002/dev.20032

Colletti, L., Mandalà, M., & Colletti, V. (2012). Cochlear implants in children younger than 6 months. *Otolaryngology–Head and Neck Surgery*, *147*, 139–146. https://doi.org/10.1177/0194599812441572

Colletti, L., Mandalà, M., Zoccante, L., Shannon, R. V., & Colletti, V. (2011). Infants versus older children fitted with cochlear implants: Performance over 10 years. *International Journal of Pediatric Otorhinolaryngology*, *2011*, 504–509.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*, 51–62.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, *9*, 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Dodd, B. (1979). Lip Reading in infants: Attention to speech presented in- and out-of-synchrony. *Cognitive Psychology*, *1979*, 478–484.

Dorn, K., Weinert, S., & Falck-Ytter, T. (2018). Watch and listen—A cross-cultural study of audio-visual-matching behavior in 4.5-month-old infants in German and Swedish talking faces. *Infant Behavior & Development*, *52*, 121–129. https://doi.org/10.1016/j.infbeh.2018.05.003

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

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *2002*, 515–546.

Guihou, A., & Vauclair, J. (2008). Intermodal matching of vision and audition in infancy: A proposal for a new taxonomy. *European Journal of Developmental Psychology*, *5*, 68–91. https://doi.org/10.1080/17405620600760409

Houston, D. M., Stewart, J., Moberly, A., Hollich, G., & Miyamoto, R. T. (2012). Word learning in deaf children with cochlear implants: Effects of early auditory experience. *Developmental Science*, *15*(3), 448–461.

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, *13*, 341–348. https://doi.org/10.1002/icd.364

Johansson, O., Davis, A., & Geijer, L. (2007). A perspective on diversity, equality and equity in Swedish schools. *School Leadership and Management*, *2007*, 21–33.

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *2007*, 1084–1089.

Kubicek, C., Gervain, J., Lœvenbruck, H., Pascalis, O., & Schwarzer, G. (2018). Goldilocks versus Goldlöckchen: Visual speech preference for same-rhythm-class languages in 6-month-old infants. *Infant and Child Development*, *27*, e2084. https://doi.org/10.1002/icd.2084

Kubicek, C., Hillairet de Boisferon, A., 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. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews. Neuroscience*, *5*, 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138–1141.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, *7*, 361–381. https://doi.org/10.1016/S0163-6383(84)80050-8

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., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experiencealters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 11442–11445. https://doi.org/10.1073/pnas.0804275105

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, *56*, 292–315. https://doi.org/10.1002/dev.21197

Lewkowicz, D. J., & Ghazanfar, A. A. (2006). The decline of cross-species intersensory perception in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 6771–6774. https://doi.org/10.1073/pnas.0602027103

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

Lindberg, I. (2007). Multilingual education: A Swedish perspective. Education in 'Multicultural' Societies–Turkish and Swedish Perspectives, 71–90.

Lindqvist, C. (2007). *Schwedische Phonetik für Deutschsprachige*, Hamburg: Buske Verlag.

Maurer, D., & Mondloch, C. (1996). Synesthesia: A stage of normal infancy. In Proceedings of the 12th meeting of the International Society for Psychophysics, Padua (Italy).

Maurer, D., & Werker, J. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*, 154–178. https://doi.org/10.1002/dev.21177

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111. https://doi.org/10.1016/S0010-0277(01)00157-3

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *1976*, 746–748.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Betoncini, J., & Amiel-Tison, C. (1988). A precursor to language acquisition in young infants. *Cognition*, *29*(2), 143–178.

Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy*, *19*, 326–337. https://doi.org/10.1111/infa.12041

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *1993*, 495–500.

Munhall, K. G., & Vatikiotis-Bateson, E. (2004). Spatial and Temporal Constraints on Audiovisual Speech Perception. Retrieved from https://scholar.google.de/scholar?hl=de&as_sdt=0%2C5&q=Munhall+%26+Vatikiotis-Bateson%2C+2004&btnG

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences*, *39*, 567–579. https://doi.org/10.1016/j.tins.2016.05.003

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 756–766.

Nazzi, T., Jusczyk, P., & Johnson, E. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*, *43*, 1–19. https://doi.org/10.1006/jmla.2000.2698

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*, 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1

Nikolopoulos, T. P., Archbold, S. M., & O'Donoghue, G. M. (1999). The development of auditory perception in children following cochlear implantation. *International Journal of Pediatric Otorhinolaryngology*, *49*, S189–S191. https://doi.org/10.1016/S0165-5876(99)00158-5

Pascalis, O., de Haan, M., & Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science*, *2002*, 1321–1323.

Patterson, M. L., & Werker, J. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*, 237–247. https://doi.org/10.1016/S0163-6383(99)00003-X

Pike, K. L. (1945). *The intonation of American English*, Ann Arbor: University of Michigan Press.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435.

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

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*, AD3–AD30. https://doi.org/10.1016/S0010-0277(00)00101-3

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*, 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*, 347–357. https://doi.org/10.3758/BF03211902

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development*, *14*, 29–50. https://doi.org/10.1002/icd.376

Schorr, E. A., Fox, N. A., van Wassenhove, V., & Knudsen, E. I. (2005). Auditory–visual fusion in speech perception in children with cochlear implants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 18748–18750.

Scott, L. S., Pascalis, O., & Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Current Directions in Psychological Science*, *16*, 197–201. https://doi.org/10.1111/j.1467-8721.2007.00503.x

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

Streri, A., Coulon, M., & Guellaï, B. (2013). The foundations of social cognition: Studies on face/voice integration in newborn infants. *International Journal of Behavioral Development*, *37*, 79–83. https://doi.org/10.1177/0165025412465361

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

Werker, & Tees. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, *7*, 49–63.

Werker, & Tees. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, *46*, 233–251. https://doi.org/10.1002/dev.20060

Werker, J., Gilbert, J. H. V., Humphrey, K., & Tees, R. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *1981*, 349–355.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196.

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

How to cite this article: Dorn K, Cauvet E, Weinert S. A cross-linguistic study of multisensory perceptual narrowing in German and Swedish infants during the first year of life. *Inf Child Dev*. 2021;30:e2217. https://doi.org/10.1002/icd.2217