An Amodal Study of Phonological Abstraction in Early Infancy

Eylem Altuntas[1], Catherine T. Best[1], Marina Kalashnikova[2], Antonia Goetz[1], Denis Burnham[1]
[1]MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Sydney, Australia
[2]Basque Center on Cognition, Brain, and Language
eylem.altuntas@westernsydney.edu.au

ABSTRACT

Adults learn non-native phonological contrasts faster if they experience the stimulus language in their first postnatal 6 months than if they never experienced it [1]. The resulting implication that young infants engage in phonological abstraction was examined here. We trained infants to associate words from two artificial mini-languages distinguished by consonant place (labial vs coronal), presented to two subgroups either as audio-only speech (A) or video-only speech (V) followed by different animal images for each of the two mini-languages. In Test, novel words of each mini-language were presented in the opposite modality to Training (A→V or V→A), followed by either the Congruent (trained) or the Incongruent (opposite mini-language) animal. The A→V mode infants attended longer to Congruent than Incongruent Test trials, but the V→A infants did not. Thus, infants show (i) phonological abstraction of the labial vs coronal distinction, (ii) in a cross-modal task, (iii) but only for A→V transfer.

Keywords: infant speech perception, phonological abstraction, language acquisition, artificial language training, perceptual learning

1. INTRODUCTION

While very young infants as early as 1 month discriminate consonant or vowel contrasts from just about any language, later in their first year they begin to pay less attention to foreign, non-native contrasts and attune to the contrasts in their ambient (native) language [1]. This process of perceptual attunement to native speech contrasts occurs around 6 months of age for vowels and around 9-10 months for consonants, and it is a well-established phenomenon [2-4]. As perceptual attunement involves the infant establishing the phonemes of the language spoken around them, speech perception *before* 6 months is generally assumed to involve phonetic-level or even acoustic-level [5] rather than phonemic-level speech perception. An implicit corollary of this assumption is that before 6 months, prior to perceptual attunement, there is no phonological abstraction, i.e., that perceptual attunement is a necessary precursor of phonological abstraction.

This view has recently been challenged by studies which show that adults retain phonologically abstract knowledge that could only have been acquired before 6 months, that is, before perceptual attunement begins. Choi and colleagues [6, 7] found that 30-year-old Dutch speakers who were adopted at 3 to 5 months of age from Korea (no subsequent contact with Korean) learned to discriminate the Korean 3-way fortis, lenis, and aspirated alveolar stop contrast, [t*]-[t]-[th], faster than their Dutch counterparts who had no previous exposure to Korean. In addition, with no further perceptual or production training, the adoptees were significantly better than the Dutch controls at (i) generalising discrimination of the 3-way contrast from the trained alveolar place of articulation to the bilabial and the velar places, and (ii) producing the contrast at all three places.

Those findings indicate that the adoptees had not only retained knowledge of speech heard early in life, but they also stored that knowledge at an abstract phonological level. As that abstract representation must have been laid down before 6 months [9, 10], these results imply that perceptual attunement (which occurs from around 6 months) is not a prerequisite for phonological abstraction. Indeed, it may even be possible that phonological abstraction paves the way for perceptual attunement. However, this evidence for abstract phonological representation prior to 6 months and storage thereafter into adulthood is indirect as it comes from 30-year-old adults. The present study investigates the development of abstract phonological knowledge directly in young infants.

Two forms of abstraction were studied here:

1. Infants first had to learn two mini-languages of word sets that differed only in consonant place of articulation: labial (lips) vs coronal (tongue tip). Infants were exposed to pseudo-words (non-words) paired with animal images that differed between the languages. Tests for this place of articulation learning presented novel words of each mini-language paired with either the same (Congruent) or the opposite (Incongruent) animal as during exposure. *This manipulation tests whether infants can form contrasting mini-language categories based solely on place of articulation, irrespective of the actual consonants manifested at these places.*

2. Infants received the word-animal pairings in the exposure phase in either audio-only (A) or video-only (V: silent talking face) modality. Test trial words

were presented in the opposite modality from the exposure phase (A→V or V→A). *This manipulation tests whether infants can learn the articulatory contrast between word sets (lips vs tongue tip consonants) and generalise it to the opposite modality, and whether this depends on direction of modality change.*

With respect to 1, research suggests that even in their earliest months of life, infants are capable of tracking the distributional statistics of vowels and consonants in their environment, allowing them to begin learning the phonetic categories used in their native language. For instance, studies such as Wanrooij, Boersma, and van Zuijen [8] have shown that infants as young as 2-3 months old can rapidly learn phonetic distinctions, as evidenced by event-related potential (ERP) studies. Additionally, Zacharaki and Sebastian-Galles [9] found evidence that infants can begin to learn and differentiate between the sounds of their native language even prior to the onset of perceptual narrowing. Furthermore, infants can discriminate consonant place of articulation contrasts by as young as young as 1-3 months of age [10]. In addition, infants as young as 2 months old can abstract consonant place contrasts across vowel context changes, as demonstrated by Bertoncini et al. [11]. Hillenbrand's [12] research further suggests that infants can perform this abstraction across contexts and talkers by the age of 6 months.

Regarding 2, it is now well-established that acoustic and dynamic optical information (from the lips, face and head of the talker) yield a collaborative effect on infants' speech perception: audio and video information together facilitate infants' recognition of their mother [13]; infants match an audio phoneme to a silent video of the face articulating it [14]; they also integrate information from the two modalities to perceive a hybrid phoneme [15]. Thus, it appears that infants detect associations between what they hear and what they see, suggesting that infants' speech perception is intermodal.

It is also possible that beyond simple transfer from one modality to another, or intermodal, speech perception may in fact be amodal, *rather than auditory or visual*. That is, speech may be perceived phonologically, which is more abstract than the specific sensory modalities. In this regard, by 4 months of age (prior to perceptual attunement) infants can recognise the particular mouth movements that correspond to specific speech sounds [16] even for phonemes that do not occur in their native language [17]. This is posited to be modality-neutral or amodal in nature, i.e., not being simply auditory, visual, tactile, and/or gestural but above them all [18].

Further evidence for the amodal nature of speech perception comes from another cross-modal study investigating Spanish-learning and English-learning infants [19]. In that study, 6- and 11-month-old infants were familiarised to silent videos (video-only) of the English /b/-/v/ contrast, which is non-existent in Spanish, then habituated to either audio-only /b/ or /v/, followed by a test on the same video-only /b/-/v/ contrast as in the familiarisation period. Six-month-olds in both language groups showed a preference for the video-only consonant they had been habituated to in the audio-only mode. But only the English-learning infants continued to show this video-only preference at 11 months, implying that the Spanish-learning infants showed a developmental decline in cross-modal discrimination of this non-native consonant contrast. This cross-modal matching of audio-only to subsequent video-only speech information is a true instance of infants' amodal speech perception.

The evidence for amodal perception in the above-mentioned study [19] comes from auditory followed by visual (A→V) presentations. There are also studies using a V→A design For instance, Teinonen et al. [20] presented silent videos of Finnish vowels to 4-month-old infants and found that the infants were better able to discriminate between audio-only vowels when they had previously seen the corresponding silent videos. Similarly, Kuhl et al. [21] presented silent videos of Mandarin Chinese syllables to 6-month-old infants and found they could better discriminate between audio-only syllables after viewing the videos. Nonetheless, other studies have not found that silent videos enhance infants' later perception of audio-only speech. Weikum et al. [22] presented silent videos of English or French vowels to 4-month-old infants and found no evidence that the videos improved their perception of audio-only vowels. These inconsistencies suggest that the relationship between visual and auditory information in speech perception is complex and may depend on various factors that require further exploration.

Thus, this study assessed POA and cross-modality (auditory, visual) abstraction, including both A→V and V→A conditions. This was achieved using a task created for a larger project on phonological abstraction (see ICPhS 2023 paper 674). In our POA/amodal task, infants were familiarised with two artificial languages containing non-words that differed solely on corresponding audio and video information – labial POA includes not only audible place information but also visible 'lip' articulatory gestures, and coronal POA includes both audible place information and visible 'tongue tip' gestures. The words were presented in either audio-only (A) or video-only (V) form, followed by tests in the opposite modality (A→V or V→A).

2. EXPERIMENT

Due to the COVID-19 pandemic, we initiated data collection online using the Lookit platform [23] and later continued it in our infant speech perception laboratory. In both contexts, we examined infants' ability to differentiate between mini-languages that differed on a single phonological feature of their consonants. In one mini-language all consonants were articulated using the lips (i.e., the *labial* feature for place of articulation) whereas in the other mini-language all consonants were articulated using the tongue tip (i.e., the *coronal* place feature).

2.1. Method

2.1.1. Participants

Eighty-three monolingual English-learning infants participated: 53 infants completed the A→V modality (exposed to audio-only words then tested on video-only) and 30 infants completed the V→A mode (exposed to video-only and tested on audio-only).

Another 20 participants were excluded from the A→V group due to technical difficulties (n = 5), fussiness (n = 4), experimenter error (n = 4), language background (n = 6), and background noise (n = 1), leading to a success rate of 72%. In the V→A group, another 21 participants were excluded due to technical complications (n = 5), fussiness (n = 3), experimenter error (n = 4), language background (n = 5), background noise (n = 3), and caregiver interference (n = 1), resulting in a success rate of 58%, which was lower than for A→V.

A→V infants had a mean age of 5.49 months (SD = 1.42 months, range = 3.91-8.94 months), and V→A ones had a mean age of 6.31 months (SD = 1.64 months, range = 4.08-8.98 months). 47 infants were North American English-learning monolinguals in the USA and 36 were Australian-English learning monolinguals in Sydney, Australia. In terms of ethnicity, 91.5% of the infants were Caucasian.

All infants were born full-term, without any known risks of cognitive or language delay, and with normal hearing and vision. Demographic information was either collected at registration or on the test day. Families received either a $5 gift card (for online participation) or $20 travel reimbursement (for lab participation), and a certificate of completion.

2.1.2. Stimuli

The two mini-languages were composed of non-words that only differed in place of articulation of the syllable-initial consonants: lips (labial place: /b, v, w/) versus tongue tip (coronal place: /d, z, l/). All words had three consonant-vowel (CV) syllables, with different consonants and vowels in each of the three syllables (e.g., *bi-va-wo* for lips, or *dæ-zu-la* for tongue tip). 240 non-words were generated for each of the two mini-languages.

A female native speaker of Australian English was video- and audio-recorded producing all 480 unique non-words for the two mini-languages in infant-directed speech (IDS). Video recordings consisted of a close-up of the speaker's face from the neck up. Audio-only (A) and video-only (V) stimulus words were separately extracted from the recordings.

2.1.3. Procedure

We used an adaptation of an associative learning task [24], which includes an Exposure Phase and a Test Phase. During the Exposure Phase, infants were exposed to two sets of word-image pairings, with modality of the words depending on Audio-only *or* Video-only modality subgroup: Set A (lips) words were followed by a cartoon image of a jellyfish and Set B (tongue tip) words by a cartoon crab image. Infants completed a set number of 36 Exposure trials divided into 3 Blocks always in the same order: 12 Set A pairings, 12 Set B pairings, then 12 mixed Set A and Set B pairings (6 of each presented in pseudo-random order). Test blocks followed immediately after Exposure with no break. The Test Phase was cross-modal for both subgroups (opposite modality to Exposure: A→V or V→A), and included two Blocks of 12 trials, a total of 24 Test trials. 25% of the Test trials were Incongruent trials [12.5% A-words, 12.5% B-words] such that the word-animal pairings were mismatched with respect to Exposure pairings (Lips-Crab; Tongue Tip-Jellyfish); and 75% of trials were Congruent trials [37.5% A-words, 37.5% B-words], such that the word-animal pairings were the same as in Exposure (Lips-Jellyfish; Tongue Tip-Crab). Test trials were presented in pseudo-random order, such that 3-trial sequences of Incongruent trials were separated by at least two 3-trial sequences of Congruent trials (e.g., Incongruent, Congruent, Congruent, Incongruent).

Different subsets of the nonwords in each mini-language were used for each phase (Training and Test) – the experiment program randomly selected words from the relevant mini-language without replacement. Thus, infants never heard or saw the same word twice, and the words were randomised across task phases and infants.

Infants sat on the parent's lap facing a computer monitor and watched a 6-minute series of the non-words, which they either heard *or* saw in Exposure and Test according to their cross-modal subgroup (A→V or V→A), paired with the designated cartoon animal for each mini-language. Parents wore

headphones and listened to music during the task so as not to inadvertently influence the baby.

2.2. Data Analysis

Infants' looking times were recorded during the experimental session and were coded offline using ELAN. Looking time was averaged over triplets (3-trial sequences) due to individual trials being very short (4.443 seconds). We transformed the raw looking times into proportions of looking time over each 3-trial sequence because there were minor variations in trial length due to internet transmission in the online test context. Statistical analyses were performed on the Test Phase data with linear mixed models in *R* [25], using the *lmer* function from the package *lme4* [26]. The *p* values for the fixed-effects factors were computed using the Kenward-Roger approximation to the degrees of freedom, as recommended by [27] and the *anova* function from the *car* package [28] was used to calculate *F*. The dependent variable was proportion of looking time per triplet. Only the Incongruent triplet and the immediately preceding Congruent triplet in each block were included in the analysis, a total of 2 Congruent-Incongruent sequences. The fixed effects were Condition (Congruent, Incongruent), Blocks (Block 1, Block 2), and Mode (A→V, V→A). Participants was a random effect.

2.3. Results

Results from the linear mixed model analysis are displayed in Figure 1. The main effect of Condition was significant, $F(1, 243) = 4.28$, $p = .03$); infants looked longer in Congruent than Incongruent triplets. There was also a significant interaction between Condition and Mode: $F(1, 243) = 4.46$, $p = .03$. Therefore, we next tested whether Condition was significant in each Mode separately. In the A→V Mode, the main effect of Condition indicates that infants looked significantly longer for Congruent than Incongruent triplets ($F(1, 314) = 27.44$, $p < 0.001$), However, in V→A Mode, Condition was not significant ($F(1, 176) = 0.31$, $p = .5$).

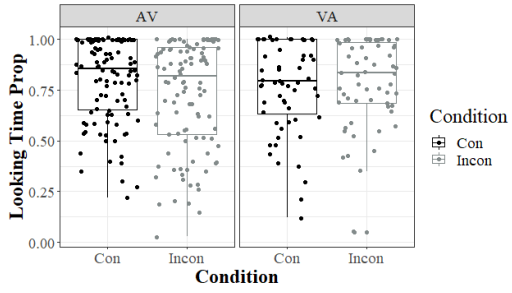

Figure 1: Looking time proportions across Test conditions in A→V and V→A cross-modal subgroups.

3. DISCUSSION AND CONCLUSIONS

This study investigated amodal phonological abstraction in infancy. Specifically, we asked whether 1. infants can learn two artificial languages based on consonants on a consonant place of articulation difference that is both auditorily and visually distinct, and whether 2. they can then transfer this learning across modalities, i.e., from audio-only to video-only (A→V mode) and video-only to audio-only (V→A).

In the A→V, but not the V→A mode, infants showed phonological abstraction of the difference between two artificial languages that differed in using labial vs tongue-tip consonants. Infants generalised this abstract category learning across the A→V speech modality change, but not across the V→A speech modality change. This shows that learning and cross-modal generalisation occur in A→V, but not in V→A. This modality difference could suggest that (i) there is no visual learning and thus no cross-modal transfer or that (ii) there is visual speech learning, which is not amodally represented thus obviating cross-modal transfer to auditory speech. The latter *cannot be correct*, because there *must be* amodal representation of visual speech given the A→V results. Thus, it is possible that in the V→A mode there is no visual learning in the Exposure phase for some more peripheral reason. The most likely candidate is that when the initial trials, in Exposure, consist of visual-only presentations (in V→A), these are not as interesting or attention-getting as are initial auditory-only trials (in A→V) such that there is little to no learning of the word-animal associations. This could be tested by detailed examination of the Exposure phase in the A→V and V→A groups, but this is beyond the scope of this presentation.

These preliminary analyses suggest that young infants show two types of phonological abstraction in the same context (i) phonological abstraction of the labial vs coronal consonant distinction, (ii) additional phonological abstraction in the form of amodal speech perception, but only in the A→V mode. Thus, not only is there phonological abstraction of phonetic /articulatory features of speech, but there is also abstract amodal representation of this phonologically abstract information prior to completion of phonological attunement in the first year of life.

4. REFERENCES

[1] J. F. Werker, J. H. Gilbert, K. Humphrey, and R. C. Tees, "Developmental aspects of cross-language speech perception," *Child development*, vol. 52, no. 1, pp. 349-355, 1981.

[2] J. F. Werker and R. C. Tees, "Influences on infant speech processing: Toward a new synthesis," *Annual review of psychology*, vol. 50, no. 1, pp. 509-535, 1999, doi: 10.1146/annurev.psych.50.1.509.

[3] P. K. Kuhl, "A new view of language acquisition," *Proceedings of the National Academy of Sciences*, vol. 97, no. 22, pp. 11850-11857, 2000, doi: 10.1073/pnas.97.22.11850.

[4] L. Polka and J. F. Werker, "Developmental changes in perception of nonnative vowel contrasts," *Journal of Experimental Psychology: Human perception and performance*, vol. 20, no. 2, p. 421, 1994, doi: 10.1037//0096-1523.20.2.421.

[5] D. K. Burnham, M. D. Tyler, and S. Horlyck, *Periods of speech perception development and their vestiges in adulthood*. 2002.

[6] J. Choi, M. Broersma, and A. Cutler, "Early phonology revealed by international adoptees' birth language retention," *Proceedings of the National Academy of Sciences*, vol. 114, no. 28, pp. 7307-7312, 2017.

[7] J. Choi, A. Cutler, and M. Broersma, "Early development of abstract language knowledge: evidence from perception–production transfer of birth-language memory," *Royal Society Open Science*, vol. 4, no. 1, p. 160660, 2017.

[8] K. Wanrooij, P. Boersma, and T. L. Van Zuijen, "Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study," *Frontiers in psychology*, vol. 5, p. 77, 2014.

[9] K. Zacharaki and N. Sebastian-Galles, "Before perceptual narrowing: The emergence of the native sounds of language," *Infancy*, vol. 27, no. 5, pp. 900-915, 2022.

[10] P. D. Eimas, "Auditory and linguistic processing of cues for place of articulation by infants," *Perception & Psychophysics*, vol. 16, no. 3, pp. 513-521, 1974.

[11] J. Bertoncini, R. Bijeljac-Babic, P. W. Jusczyk, L. J. Kennedy, and J. Mehler, "An investigation of young infants' perceptual representations of speech sounds," *Journal of experimental psychology: General*, vol. 117, no. 1, p. 21, 1988.

[12] J. Hillenbrand, "Speech perception by infants: Categorization based on nasal consonant place of articulation," *The Journal of the Acoustical Society of America*, vol. 75, no. 5, pp. 1613-1622, 1984.

[13] D. Burnham, "Visual recognition of mother by young infants: facilitation by speech," *Perception*, vol. 22, no. 10, pp. 1133-1153, 1993, doi: 10.1068/p221133.

[14] B. Dodd and D. Burnham, "Processing speechread information," *The Volta Review*, 1988.

[15] D. Burnham and B. Dodd, "Auditory-visual speech perception as a direct process: The McGurk effect in infants and across languages," in *Speechreading by humans and machines*: Springer, 1996, pp. 103-114.

[16] P. K. Kuhl and A. N. Meltzoff, "The bimodal perception of speech in infancy," *Science*, vol. 218, no. 4577, pp. 1138-1141, 1982, doi: 10.1126/science.7146899.

[17] G. E. Walton and T. Bower, "Amodal representation of speech in infants," *Infant Behavior and Development*, vol. 16, no. 2, pp. 233-243, 1993.

[18] P. K. Kuhl and A. N. Meltzoff, "Speech as an intermodal object of perception," in *Perceptual development in infacy: The Minnesota symposia on child psychology*, 1988, vol. 20, pp. 235-266.

[19] F. Pons, D. J. Lewkowicz, S. Soto-Faraco, and N. Sebastián-Gallés, "Narrowing of intersensory speech perception in infancy," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10598-10602, 2009, doi: 10.1073/pnas.0904134106.

[20] T. Teinonen, V. Fellman, R. Näätänen, P. Alku, and M. Huotilainen, "Statistical language learning in neonates revealed by event-related brain potentials," *BMC neuroscience*, vol. 10, no. 1, pp. 1-8, 2009.

[21] P. K. Kuhl, K. A. Williams, F. Lacerda, K. N. Stevens, and B. Lindblom, "Linguistic experience alters phonetic perception in infants by 6 months of age," *Science*, vol. 255, no. 5044, pp. 606-608, 1992.

[22] W. M. Weikum, A. Vouloumanos, J. Navarra, S. Soto-Faraco, N. Sebastián-Gallés, and J. F. Werker, "Visual language discrimination in infancy," *Science*, vol. 316, no. 5828, pp. 1159-1159, 2007.

[23] K. Scott and L. Schulz, "Lookit (part 1): A new online platform for developmental research," *Open Mind*, vol. 1, no. 1, pp. 4-14, 2017.

[24] C. Kabdebon and G. Dehaene-Lambertz, "Symbolic labeling in 5-month-old human infants," *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, pp. 5805-5810, 2019.

[25] R. C. Team, "R: The R Project for Statistical Computing [Internet]. 2021 [cited 2022 Aug 22]," *Available from: Available from:* https://www. *r-project. org*.

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823*, 2014.

[27] U. Halekoh and S. Højsgaard, "A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models–the R package pbkrtest," *Journal of Statistical Software*, vol. 59, pp. 1-32, 2014.

[28] J. Fox and S. Weisberg, *An R companion to applied regression*. Sage publications, 2018.