AUDIOVISUAL CUES IN DEVELOPMENTAL SPEECH PERCEPTION: DATA FROM A VOICELESS FRICATIVE PLACE DISCRIMINATION

[1]Laura Bosch, [2]Joan Birulés, [1]Ferran Pons

[1]Universitat de Barcelona, [2]Université Grenoble-Alpes
laurabosch@ub.edu, joanbirules@gmail.com, ferran.pons@ub.edu

ABSTRACT

Infants' perceptual narrowing to native speech sounds has been mainly described from auditory paradigms. Previous results from fricative place of articulation discrimination in infancy offer a non-convergent view on the factors modulating discrimination of noise contrasts. The present research focuses on infants' capacity to discriminate a native voiceless fricative place of articulation contrast, [s] - [f]. The contribution of audiovisual (AV) cues in reaching successful discrimination was explored. Three groups of 6-month-old infants were tested with versions of the familiarization-preference procedure involving two auditory-only conditions (adult-directed and infant-directed speech) and an AV presentation. Discrimination was absent and not favored by the AV format. Positive evidence was only obtained in an older group of 12-month-olds from the auditory-only version. The developmental pattern emerging from these data deviates from the expected maintenance of early language-general skills and suggests a late emergence of this fricative contrast with limited contribution of AV cues.

Keywords:voiceless fricatives, place discrimination, audiovisual cues, infants, perceptual narrowing.

1. INTRODUCTION

The field covering developmental speech perception research in infancy and its connection to later language learning skills has established that this process starts well before infants begin to utter their first words. Infants' attention towards the language spoken in their environment, increases their learning of the specific sound properties of their native language [1]. The speech signal is rich in information and it can be encoded based on a number of different dimensions. Infants' representation of these dimensions emerges from their initial perceptual biases, the acoustic-perceptual saliency of the information, the regularities present in the input language and the implicit learning mechanisms that are at work [2]. Phonetic learning in the first year of life has been characterized as a transition from language-general to language-specific speech perception resulting from experience with the native input. Attunement to the native sound properties leads to perceptual narrowing processes, described as a focus on language-specific segmental properties that sets the foundation of a contrastive, phonological system [3-4]. However, a review of the studies addressing discrimination of consonant contrasts in infancy reveals inconsistencies relative to the so-called canonical developmental pattern [5], especially affecting some nasal, liquid and fricative segments. The discrimination trajectory deviates from the universal model that posits an initial broad sensitivity for many different segmental contrasts, followed by an enhanced discrimination capacity for the native ones and a loss of sensitivity for the non-native ones. Alternative trajectories have been described, often revealing a pattern of a slower, protracted development, with gradual gains resulting from experience after an initial stage of poor or inconsistent discrimination [6]. The present research has placed the focus on infants' capacity to discriminate a native voiceless fricative place of articulation contrast, [s] - [f], to inform about its developmental trajectory and, additionally, to assess the possible contribution of audiovisual cues in the discrimination process of this consonant contrast. The role of visual speech, that is, seeing the articulation and paying attention to distinctive visual speech cues might enhance discrimination, as shown in research with young infants, tested with both native [7] and non-native consonant contrasts [8].

2. BACKGROUND RESEARCH

The study of fricative place of articulation discrimination in infancy has offered inconsistent data. Research has covered different contrasts, relative to different places of articulation, some of them more confusable than others. The location of the broadband aperiodic energy (higher vs. lower frequencies) can make discrimination relatively easy (e.g. [f] - [s]), or more confusable when the acoustic difference is reduced (e.g. [f] - [θ]). The former pair seems to be discriminated by 6-month-old infants [9], while discrimination of the latter contrast was not obtained till 12-14 months of age, still limited to

certain phonetic contexts [10]. However, a different study found positive evidence in 6-month-olds [11]. The alveolar - post-alveolar distinction, involving spectral and temporal differences was also found difficult to discriminate in 6- to 14 month-olds, contrary to data from an earlier study [12]. More recently, research exploring the effect of exposure on the discrimination of the alveopalatal - retroflex fricatives revealed restricted learning in 4- to 6-month-olds [13].

Overall, these studies offer a rather heterogeneous pattern of results, likely to be affected by acoustic as well as contextual factors, with some fricative pairs easier or less prone to confusion than others. Even though methodological approaches are slightly different, all have in common the use of auditory-only paradigms to test infants' discrimination or learning. Unfortunately, audiovisual (AV) paradigms have scarcely been used in this research domain, while they can offer relevant information to better understand and clarify the inconsistent pattern of results just described. Some place of articulation contrasts (such as the target contrast in this research, the labiodental - alveolar, [f] - [s] distinction) are produced with clearly visible articulatory cues, which can favour discrimination. It has already been shown, by studies exploring the distribution of attention to talking faces, that by around 6 months of age infants' attention to a talker's mouth gradually increases and the detection of the perceptual equivalence between auditory and visual properties of speech is likely to be connected to speech processing and phonetic learning [14]. Although in a different study, data suggesting a more restrictive attention shift to the mouth in 6 to 9 month-old infants has been reported limited to contexts in which auditory and visual information were in conflict [15], the possible contribution of AV cues in phonetic discrimination studies on challenging contrasts does need to be further assessed.

3. AIMS OF THE STUDY

Preliminary research in our Lab failed to find evidence of discrimination for a voiceless fricative place of articulation contrast in 6-month-olds, tested on the [f] - [s] native contrast, in spite of a clear difference in acoustic salience (larger distance in the center of broadband noise in the spectrum). Distributional properties of these segments in the native input favoring [s] might preclude an early differentiation or lead to an asymmetrical pattern of discrimination. Common temporal features of these fricatives might also constrain differentiation, while visual cues, if available, might have a facilitation effect. The present research aims at exploring the discrimination of this contrast using different conditions, controlling for the auditory characteristics of the material, and also including an AV version. To better understand the perceptual trajectory for this contrast, a group of one-year-olds was also been tested.

4. METHODOLOGICAL APPROACH

The experimental approach adopted in this research used versions of the familiarization-preference procedure [16-17], except for the AV modality experiment that required an habituation phase. Test trials were restricted to four (two Same and two Switch). Differential (longer) attention time to Switch trials in the test phase was interpreted as positive evidence of discrimination. Stimuli were CVC monosyllabic non-words, with the target fricative as syllable onset, followed by the vowel [i] and [t] as a coda. Variability, a common feature of the material in the these experiments, was implemented by using recordings from several speakers and a multi-token presentation.

5. DISCRIMINATION IN SIX-MONTH-OLD INFANTS

5.1. Experiments 1a and 1b: Auditory-only modality

5.1.1. Participants

A sample of N=18 (7 boys, M age in days = 186.1, range = 108-223) full-term 6-month-old infants with no history of hearing problems according to parents' report participated in Experiment 1a. An independent sample of N=18 (10 boys, M age in days 192.4, range 148-203) full-term 6-month-olds with similar characteristics participated in Experiment 1b. All participants came from Spanish-speaking or Catalan-speaking homes. The fricative contrast under study is present in both languages and has a similar distribution. The target stimuli were non-words in both these languages.

5.1.2. Material

For Experiment 1a, eighteen [sit] and [fit] tokens were selected from the recordings of six female speakers that were requested to produce the stimuli in a *motherese* style. In each category, twelve tokens were used for the familiarization phase and six for the test. Variability among tokens in each category was high, but significant differences were only found for the expected acoustic dimensions that characterize the target fricatives in this study (see Table 1).

For Experiment 1b, a selection of 16 tokens for each category (from those used in Experiment 1a, same six speakers) was used and they were classified in terms of speech register based on their proximity to Adult- or Infant-directed speech typical properties (ADS vs. IDS). Tokens to be used in the IDS condition differed from those in the ADS condition only in overall duration, vowel duration and F2 frequency at vowel transition.

	[sit]	[fit]	*p*-value
Duration (s)	1.06	1.11	n.s.
Noise duration (s)	0.27	0.29	n.s.
Vowel F0 (Hz)	256.7	280.9	n.s.
CoG (Hz)	6006.7	3634.5	< 0.001
F2 vowel trans. (Hz)	2609.9	2539.1	0.03

Table 1: Mean values of acoustic measures obtained for tokens in each category of fricatives used in Experiment 1a.

5.1.3. Procedure

In both a and b Experiments infants were tested with a version of the familiarization-preference procedure that consisted in a 2-minute infant-contingent familiarization phase with tokens of the same stimulus type (either [sit] or [fit]) presented on alternating trials from two lateral loudspeakers and followed by four test trials, quasi-randomly presented. Experiment 1b also involved two different conditions (ADS / IDS material) for each familiarization condition. Participants were randomly assigned to each testing condition.

5.1.4. Results

Experiment 1a: A one-way ANOVA ensured that both familiarization conditions ([sit] or [fit]) did not differ in number of familiarization trials needed to accumulate the 2 minutes' attention in this phase [$F(1, 16) = 1.95$, $p = .182$, petasq = .11]. Each participant's mean looking time (LT) to Same and Switch test trials was computed. A mixed ANOVA with Trial (Same, Switch) as a within-subjects factor and Familiarization Stimulus ([fit],[sit]) as a between-subjects factor. No significant results were obtained (all $p >$. 1; see Figure 2, left).

Experiment 1b: A mixed ANOVA with Trial (Same, Switch) as a within-subjects factor, and Stimulus type (ADS, IDS) and Familiarization Stimulus ([fit],[sit]) as between-subjects factors was run. No significant effects were found (all $p >$. 1).

5.2. Experiment 2: audiovisual modality

5.2.1. Participants

N=18 healthy, full-term and normally hearing infants (9 boys; M age=191.55, range=180-205). They came from the same language environments as participants in Experiment 1.

5.2.2. Material

Video stimuli were recorded for this experiment. Four different female speakers were recorded uttering several [fit] and [sit] tokens (see Figure 1) in a natural voice and using different intonation patterns as in IDS. It was first confirmed that adults were able to discriminate the fricative segments from a silent presentation.

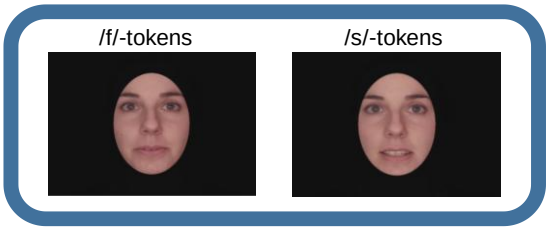

Figure 1: Two examples of frames from the video recordings with speakers producing [f] and [s] tokens for Experiment 2.

For each category ([fit] and [sit]) two blocks of video tokens were built to be used in the habituation phase trials. Each block was made of six different video tokens from two different speakers and lasted approximately 15 seconds. Another 12 different tokens from the same speakers were selected for the test phase, 6 for each category ([fit] and [sit]).The faces of the speakers were centered and faded in and out of a black background in each presentation.

5.2.3 Procedure

Infants were tested with a habituation-preference paradigm [18]. The high perceptual saliency of the audiovisual talking faces precluded the use of a 2-minutes' familiarization phase as in the previous experiments. The paradigm included a habituation phase using two different blocks of either [sit] or [fit] AV recordings, presented in a random order, until participants reached the habituation criterion (a decrease of 40% of attention time from the baseline). The test phase included 4 test trials, two Same and two Switch trials in a counterbalanced order. Testing took place in a soundproof room. The experimenter monitored infants' behavior from an adjacent control room, using Habit 1.0 software [19].

5.2.4. Results

A one-way ANOVA first ensured similar habituation times in both conditions (M habituation trials = 15.11; [fit] 15.33, [sit] 14.88; F(1, 16) = 0.04, p = .846, petasq < .01]. LT Scores were analyzed with a mixed ANOVA with Trial (Same, Switch) as a within-subjects factor and Habituation Stimulus ([fit], [sit]) as a between-subjects factor. The ANOVA yielded no significant results (see Figure 2), except for a trend in Trial [F(1, 16) = 3.95, p = .064, petasq = .20].

6. DISCRIMINATION IN ONE -YEAR-OLD INFANTS

6.1. Experiment 3: auditory-only modality

5.2.1. Participants

N=18 healthy, full-term 12-month-old infants (11 boys, M age = 370.94, range = 345-393), with no history of hearing problems. The language background was similar to the one described in the previous experiments.

5.2.2. Material and Procedure

Stimuli, procedure and apparatus were identical to Experiment 1a.

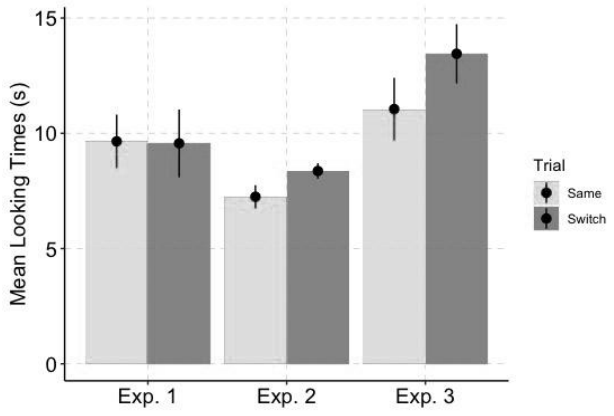

Figure 2: Mean attention time to Same and Switch trials for 6-month-old and 12-month-old infants tested in the auditory-only condition (Experiments 1 and 3), and for 6-month-old infants in the AV condition (Experiment 2).

5.2.4. Results

A one-way ANOVA on number of trials in each familiarization condition revealed that those familiarized with [fit] needed more trials to reach criterion [F(1, 16) = 4.88, p = .042, petasq = .23]. A mixed ANOVA with LT Scores and Trial (Same, Switch) as a within-subjects factor, and Familiarization Stimulus ([fit], [sit]) as a between-subjects factor revealed a significant main effect of Trial [F(1, 16) = 6.49, p = .022, petasq = .29], with longer attention to the Switch. Although the interaction of Trial with Stimulus Familiarization did not reach significance [F(1, 16) = 2.63, p = .124, petasq = .14], due to differences in the Familiarization phase infants' attention in the two groups were separately analyzed. Paired t-tests revealed that only infants familiarized with [fit] detected the change to [sit] [t(8) = 4.05, p = .004, d = 1.35], but not in the other direction (t <1).

7. CONCLUSIONS

The pattern of results in these experiments suggests a non-canonical developmental trajectory for a place of articulation fricative contrast present in the native language of the participants. Only by age 12 months first evidence of discrimination could be gathered, together with an asymmetrical pattern favouring one particular direction in change detection. Crucially, neither AV cues, nor exaggerated (IDS) presentation in the auditory-only modality were sufficient to increase the likelihood of contrast detection. This is especially relevant considering the variability in our testing material, based on different tokens produced by multiple speakers and designed to test phonetic categorization. Variability resulting from different voices and different stimuli used in our reserach paradigm clearly constrains discrimination by age 6 months, a result the diverges from data in [9], although the aim of that study and the paradigm used were different. The lack of a facilitation effect in the AV experiment, offering visible cues to support differentiation, reveals some limitation in the use of the redundant AV cues by 6 months, at least in this task. Further research should explore this issue with eye-tracking technology so as to identify how attention to the visual information is deployed during the task and which changes can be observed when older groups of infant participants are tested. Additional experiments using a different vowel to form the syllabic stimuli are also needed so as to confirm or modulate the present results.

8. ACKNOWLEDGEMENTS

The authors wish to thank the participating infants and their families. Research funded by Spanish MICINN projects (PGC 2018-097487-B-100, PID 2021-128159NB-100) to LB and FP. The authors also thank M. Ramon-Casas, J. Sánchez-Galán, M. Konjevod, N. Sanz-Gauntlett, and W. Elvira for their support with experiment preparation, phonetic analyses, data collection and coding.

9. REFERENCES

[1] Panneton, R., Newman, R. 2012. Development of speech perception. In: Werner, L.A. (ed.) *Human auditory development*. Springer, 197-222.

[2] Werker, J. F., Curtin, S. 2005. PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, *1*(2), 197-234.

[3] Kuhl, P. K. 2004. Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831-843.

[4] Werker, J. F., Yeung, H. H., Yoshida, K. A. 2012. How do infants become experts at native-speech perception?. *Current Directions in Psychological Science*, *21*(4), 221-226.

[5] Narayan, C. R. 2019. An acoustic perspective on 45 years of infant speech perception, Part 1: Consonants. *Language and Linguistics Compass*, *13*(10), e12352.

[6] Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental science*, *13*(3), 407-420.

[7] Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850-855.

[8] Danielson, D. K., Bruderer, A. G., Kandhadai, P., Vatikiotis-Bateson, E., & Werker, J. F. (2017). The organization and reorganization of audiovisual speech perception in the first year of life. *Cognitive Development*, *42*, 37-48.

[9] Beach, E. F., Kitamura, C. 2011. Modified spectral tilt affects older, but not younger, infants' native-language fricative discrimination. *Journal of Speech, Language and Hearing Research*, 54, 658-667.

[10] Eilers, R. E., Wilson, W. R., Moore, J. M. 1977. Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, *20*(4), 766-780.

[11] Tyler, M. D., Best, C. T., Goldstein, L. M., Antoniou, M. 2014. Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental psychobiology*, *56*(2), 210-227.

[12] Nittrouer, S. 2001. Challenging the notion of innate phonetic boundaries. *The Journal of the Acoustical Society of America*, *110*(3), 1598-1605.

[13] Cristià, A., McGuire, G. L., Seidl, A., Francis, A. L. 2011. Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of phonetics*, *39*(3), 388-402.

[14] Lewkowicz, D. J., Hansen-Tift, A. M. 2012. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431- 1436.

[15] Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., & Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology*, *10*(5), 611-624.

[16] Jusczyk, P. W., Aslin, R. N. 1995. Infants′ detection of the sound patterns of words in fluent speech. *Cognitive psychology*, *29*(1), 1-23.

[17] Bosch, L., Sebastián-Gallés, N. 2003. Language experience and the perception of a voicing contrast in fricatives: Infant and adult data. In *Proceedings of the 15th international conference of phonetic sciences* (pp. 1987-1990). UAB/Casual Prods.

[18] Fennell, C. T. 2012. Habituation procedures. In: E. Hoff (ed.) *Research methods in child language: A practical guide*, Blackwell, 1-16.

[19] Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). Habit 2000: A new program for testing infant perception and cognition.